IJCSI

International Journal of Computer Science Issues

Volume 10, Issue 5, No 2, September 2013
ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

www.IJCSI.org

IJCSI proceedings are currently indexed by:

IJCSI Editorial Board 2013

EDITORIAL

In this fifth edition of 2013, we bring forward issues from various dynamic computer science fields ranging from system performance, computer vision, artificial intelligence, software engineering, multimedia, pattern recognition, information retrieval, databases, security and networking among others.

Considering the growing interest of academics worldwide to publish in IJCSI, we invite universities and institutions to partner with us to further encourage open-access publications.

As always we thank all our reviewers for providing constructive comments on papers sent to them for review. This helps enormously in improving the quality of papers published in this issue.

Google Scholar reported a large amount of cited papers published in IJCSI. We will continue to encourage the readers, authors and reviewers and the computer science scientific community and interested authors to continue citing papers published by the journal.

Apart from availability of the full-texts from the journal website, all published papers are deposited in open-access repositories to make access easier and ensure continuous availability of its proceedings free of charge for all researchers.

We are pleased to present IJCSI Volume 10, Issue 5, No 2, September 2013 (IJCSI Vol. 10, Issue 5, No 2). The acceptance rate for this issue is 30.4%.

IJCSI Editorial Board
September 2013 Issue
ISSN (online): 1694-0784
ISSN (print): 1694-0814

www.IJCSI.org

IJCSI Reviewers Committee 2013

Mr. Sunil Kashibarao Nayak, Bahirji Smarak Mahavidyalaya, Basmathnagar Dist-Hingoli., India

Prof. Nikhil gondaliya, G H Patel College of Engg. & Technology, India

Mr. Nisheeth Joshi, Apaji Institute, Banasthali University, India

Mr. Nizar, National Ingineering School of Monastir, Tunisia

Prof. R. Jagadeesh Kannan, RMK Engineering College, India

Prof. Rakesh.L, Vijetha Institute of Technology, Bangalore, India

Mr B. M. Patil, Indian Institute of Technology, Roorkee, Uttarakhand, India

Dr. Intisar A. M. Al Sayed, Associate prof./College of Science and IT/Al Isra University, Jordan

Mr. Thipendra Pal Singh, Sharda University, K.P. III, Greater Noida, Uttar Pradesh, India

Mrs. Rajalakshmi, JIITU, India

Mr. Shrikant Ardhapurkar, Indian Institute of Information Techonology, India

Ms. Hemalatha R, Osmania University, India

Mr. Hadi Saboohi, University of Malaya - Faculty of Computer Science and Information Technology, Malaysia

Mr. Sunil Kumar Grandhi, Maris Stella College, India

Prof. Shishir K. Shandilya, NRI Institute of Science & Technology, INDIA

Dr. Umesh Kumar Singh, Vikram University, Ujjain, India

Prof. Prasun Ghosal, Bengal Engineering and Science University, India

Dr. Nagarajan Velmurugan, SMVEC/Pondicherry University, India

Dr. R. Baskaran, Anna University, India

Dr. Wichian Sittiprapaporn, Mahasarakham University College of Music, Thailand

Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia

Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology, India

Mrs. Inderpreet Kaur, PTU, Jalandhar, India

Mr. Palaniyappan, K7 Virus Research Laboratory, India

Mr. Guanbo Zheng, University of Houston, main campus, USA

Mr. Arun Kumar Tripathi, Krishna Institute of Engg. and Tech-Ghaziabad, Affilated to UPTU, India

Mr. Iqbaldeep Kaur, PTU / RBIEBT, India

Mr. Amit Choudhary, Maharaja Surajmal Institute, New Delhi, India

Mrs. Vasudha Bahl, Maharaja Agrasen Institute of Technology, Delhi, India

Dr. Ashish Avasthi, Uttar Pradesh Technical University, India

Ms. A. Lavanya, Manipal University, Karnataka, India

Dr. Jalal Laassiri, MIA-Laboratory, Faculty of Sciences Rabat, Morocco

Mr. Ganesan, Sri Venkateswara college of Engineering and Technology, Thiruvallur, India

Mr. V.Ramakrishnan, Sri Venkateswara college of Engineering and Technology, Thiruvallur, India

Prof. Vuda Sreenivasarao, St. Mary's college of Engg & Tech, India

Prof. Ashutosh Kumar Dubey, Assistant Professor, India

Dr. R.Ramesh, Anna University, India

Mr. Ali Khadair HMood, University of Malaya, Malaysia

Dr. Vimal Mishra, U.P. Technical Education, India

Mr. Ranjit Singh, Apeejay Institute of Management, Jalandhar, India

Mrs. D.Suganyadevi, SNR SONS College (Autonomous), India

Mr. Prasad S.Halgaonkar, MIT, Pune University, India

Mr. Vijay Kumar, College of Engg. and Technology, IFTM, Moradabad(U.P), India

Mr. Mehran Parchebafieh, Douran, Iran

Mr. Anand Sharma, MITS, Lakshmangarh, Sikar (Rajasthan), India

Mr. Amit Kumar, Jaypee University of Engineering and Technology, India

Prof. B.L.Shivakumar, SNR Sons College, Coimbatore, India

Mr. Mohammed Imran, JMI, India

Dr. R Bremananth, School of EEE, Information Engineering (Div.), Nanyang Technological University, Singapore

Prof. Vasavi Bande, Computer Science and Engneering, Hyderabad Institute of Technology and Management, India

Dr. S.R.Balasundaram, National Institute of Technology, India

Dr. Prasart Nuangchalerm, Mahasarakham University, Thailand

Dr. M Ayoub Khan, C-DAC, Ministry of Communications & IT., India

Dr. Jagdish Lal Raheja, Central Electronics Engineering Research Institute, India

Mr G. Appasami, Dept. of CSE, Dr. Pauls Engineering College, Anna University - Chennai, India

Mr Vimal Mishra, U.P. Technical Education, Allahabad, India

Mr. Amin Daneshmand Malayeri, Young Researchers Club, Islamic AZAD University, Malayer Branch, Iran

Dr. Arti Arya, PES School of Engineering, Bangalore (under VTU, Belgaum, Karnataka), India

Mr. Pawan Jindal, J.U.E.T. Guna, M.P., India

Dr. Soumen Mukherjee, RCC Institute of Information Technology, India

Dr. Hamid Mcheick, University of Qubec at Chicoutimi, Canada

Dr. Mokhled AlTarawneh, PhD computer engineering/ Faulty of engineerin/ mutah university, jordan

Prof. Santhosh.P.Mathew, Saintgits College of Engineering, Kottayam, India

Ms. Suman Lata, Rayat Bahara institue of engg. & Nanotechnology,Hoshiarpur, India

Dr. Shaikh Abdul Hannan, Vivekanand College, Aurangabad, India

Prof. PN Kumar, Amrita Vishwa Vidyapeetham, India

Dr. P. K. Suri, Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, India

Dr. Syed Akhter Hossain, Daffodil International University, Bangladesh

Mr. Sunil, Vignan College, India

Mr. Ajit Singh, TIT&S Bhiwani, Haryana, India

Mr. Nasim Qaisar, Federal Urdu Univetrsity of Arts , Science and Technology, Pakistan

Ms. Rshma, Maharishi Markandeshwar University, India

Mr. Gaurav Kumar Leekha, M.M.University, Solan (Himachal Pardesh), India

Mr. Ordinor Tucker, Ministry of Finance Jamaica, Jamaica

Mr. Mohit Jain, Maharaja Surajmal Institute of Technology (Affiliated to Guru Gobind Singh Indraprastha University, New Delhi), India

Dr. Shaveta Rani, GZS College of Engineering & Technology, India

Dr. Paramjeet Singh, GZS College of Engineering & Technology, India

Dr. G R Sinha, SSCET, India

Mr. Chetan Sharma, TechMahindra India Ltd., India

Dr. Nabil Mohammed Ali Munassar, University of Science and Technology, Yemen

Prof. T Venkat Narayana Rao, Department of CSE, Hyderabad Institute of Technology and Management , India

Prof. Vasavi Bande, HITAM, Engineering College, India

Prof. S.P.Setty, Andhra University, India

Dr. C. Kiran Mai, J.N.T.University,Hyderabad/VNR Vignana Jyothi Institute of Engineering & Technology/, India

Ms. Bindiya Ahuja, Manav Rachna International University, India

Mrs. Deepa Bura, Manav Rachna International University, India

Mr. Vikas Gupta, CDLM Government Engineering College, Panniwala Mota, India

Dr Juan José Martínez Castillo, University of Yacambu, Venezuela

Mr Kunwar S. Vaisla, Department of Computer Science & Engineering, BCT Kumaon Engineering College, India

Mr. Abhishek Shukla, RKGIT, India

Prof. Manpreet Singh, M. M. Engg. College, M. M. University, Haryana, India

Mr. Syed Imran, University College Cork, Ireland

Dr. Intisar Al Said, Associate Prof/Al Isra University, Jordan

Dr. Namfon Assawamekin, University of the Thai Chamber of Commerce, Thailand

Dr. Shiv KUmar, Technocrat Institute of Technology-Bhopal (M.P.), India

Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran

Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran

Dr. Mohamed Ali Mahjoub, University of Monastir, Tunisia

Mr. Adis Medic, Infosys ltd, Bosnia and Herzegovina

Mr Swarup Roy, Department of Information Technology, North Eastern Hill University, Umshing, Shillong 793022, Meghalaya, India

Prof. Jakimi, Faculty of Science and technology my ismail University, Morocco

Dr. R. Manicka Chezian, N G M College, Pollachi - 642 001, Tamilnadu, India

Dr. P.Dananjayan, Pondicherry Engineering College, India

Mr. Manik Sharma, Sewa Devi SD College Tarn Taran, India

Mr. Suresh Kallam, East China University of Technology, Nanchang, China

Dr. Mohammed Ali Hussain, Sai Madhavi Institute of Science & Technology, Rajahmundry, India

Mr. Vikas Gupta, Adesh Instutute of Engineering & Technology, India

Dr. Anuraag Awasthi, JV Womens University, Jaipur, India

Dr. Mathura Prasad Thapliyal, Department of Computer Science, HNB Garhwal University (Centr al University), Srinagar (Garhwal), India

Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia, Malaysia

Mr. Adnan Qureshi, University of Jinan, Shandong, P.R.China, P.R.China

Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India

Mr. Mueen Uddin, Universiti Teknologi Malaysia, Malaysia

Mr. Manoj Gupta, Apex Institute of Engineering & Technology,Jaipur (Affiliated to Rajasthan Technical University,Rajasthan), Indian

Mr. S. Albert Alexander, Kongu Engineering College, India

Dr. Shaidah Jusoh, Zarqa Private University, Jordan

Dr. Dushmanta Mallick, KMBB College of Engineering and Technology, India

Mr. Santhosh Krishna B.V, Hindustan University, India

Dr. Tariq Ahamad Ahanger, Kausar College Of Computer Sciences, India

Dr. Chi Lin, Dalian University of Technology, China

Prof. VIJENDRA BABU.D, ECE Department, Aarupadai Veedu Institute of Technology, Vinayaka Missions University, India

TABLE OF CONTENTS

On Performance of Logical-Clustering of Flow-Sensors

Rahim Rahmani, Hasibur Rahman and Theo Kanter

Department of Computer and Systems Sciences, Stockholm University
Kista, Stockholm SE-164 40, Sweden

Abstract

In state-of-the-art Pervasive Computing, it is envisioned that unlimited access to information will be facilitated for anyone and anything. Wireless sensor networks will play a pivotal role in the stated vision. This reflects the phenomena where any situation can be sensed and analyzed anywhere. It makes heterogeneous context ubiquitous. Clustering context is one of the techniques to manage ubiquitous context information efficiently to maximize its potential. Logical-clustering is useful to share real-time context where sensors are physically distributed but logically clustered. This paper investigates the network performance of logical-clustering based on ns-3 simulations. In particular reliability, scalability, and reachability in terms of delay, jitter, and packet loss for the logically clustered network have been investigated. The performance study shows that jitter demonstrates 40 % and 44 % fluctuation for 200 % increase in the node per cluster and 100 % increase in the cluster size respectively. Packet loss exhibits only 18 % increase for 83 % increase in the packet flow-rate.

Keywords: *Pervasive Computing, Wireless sensor networks, ubiquitous, context, distributed, logical-clustering, ns-3.*

1. Introduction

The Wireless Sensor Networks (WSNs) is an integral part of today's pervasive computing and expected to play a pivotal role in the future Networked Society. The primary use of sensors is to collect data from physical objects. It is foreseen that any situation can be sensed and analyzed anywhere which leads to more and more sensors deployment in today's Internet infrastructure and sensors are made available to the services through the distributed acquisition and dissemination of sensor data assembled from physical objects. Services can access this heterogeneous context information anywhere. The use of sensors is increasing rapidly. Billions of sensors will be used in the foreseeable future [2]. This will play a vital role in making context information accessible for anyone and anything in the future Networked Society [1]. These enormous numbers of sensors deployment in the Internet of Things (IoT) will allow gathering information from people, places, and objects i.e. from distributed sensor networks. Spontaneous human participation which is known as crowdsourcing is also envisioned [6]. This implies that rapid real-time data will be generated by crowd about the circumstances surround [7]. These will produce heterogeneous context information. Moreover, a single sensor might produce different data. For example, sensors carried by human on their smart devices might produce different data in different time. This necessitates proper management of heterogeneous contexts obtained from sensors. Data management should be reliable, and the high volume of data should be scaled appropriately in order to use efficiently and meaningfully. Clustering the context i.e. data is one of the proficient applications. Furthermore, it will be advantageous to cluster sensor data based on context similarity. In one of our previous papers, logical-clustering of flow-sensors has been presented [2]. Logical-clustering implies that sensors might reside remotely physically but clustered logically based on context similarity. Previous most work on sensors clustering concentrated on physical location nearness for energy and routing management, and to increase system scalability and robustness. Context in sensors clustering has been discussed too, but in all cases definition of context is specific. Moreover, their solution is limited to neighboring sensors. However, the concept of logical-clustering will allow resources (data, services) to be shared among different physically distributed sensors in distributed sensor networks. Sensors can share resources through distributed collaboration which was lacking in the existing management of context information. Once the clustering is done then each cluster is identified through a *context-ID* which is defined based on context similarity and published on the internet. Any interested sensor, may be located remotely, can subscribe to the context-ID.

OpenFlow based sensors are known as flow-sensors [3]. It has been proven that flow-sensors perform better than typical sensors [3]. However, it will be infeasible for a single OpenFlow controller to manage the increased number of sensors. In order to manage huge amount of sensors, more than one OpenFlow controller is desirable. HyperFlow addresses the issue and offers multiple controllers which are physically distributed but logically centralized [2]. The controllers are synchronized and can be resilient for network slicing. An important factor that was missing in the existing OpenFlow specification is interconnection between different OpenFlow networks, HyperFlow solves this problem by using the publish/subscribe mechanism. In HyperFlow, each controller can make decision locally which minimizes the response time. Controllers exchange messages to notify about any network-wide changes. These logically synchronized controllers are called *logical-sink* [2].

Network performance is one of the most researched issues in the field of wireless sensor networks. Network management becomes an important consideration as the number of sensor nodes increases. In future, network will encounter thousand times traffic volumes compared to today's traffic volumes. Latency, reliability, scalability, and data reachability are few of the challenges that future network would encounter [1]. Therefore, it is essential to design network carefully so that network does not incur performance degradation. In our previous paper, the feasibility and technical presentation of logical-clustering have been discussed [2]. In addition, computational efficiency of logical-clustering has also been shown in [2]. In this paper, the focus will be on investigating few of the significant network performance metrics of logical-clustering. The network has been designed in ns-3 (network simulator). A performance study has been made in terms of delay, jitter and packet loss to verify the reliability, scalability and reachability of the designed network. Hence, the main focus of this paper will be to:

- Design a WSN of logical-clustering of flow-sensors in ns-3
- Verify the reliability of the designed network in terms of packet delay and jitter
- Verify packet reachability
- Examine scalability of the network for increased number of nodes and groups
- Provide use cases of logical-clustering

The remainder of the paper is organized as follows: section 2 presents the related work. Section 3 discusses the motivation behind the work. Section 4 outlines the system model considered for the proposal. Next, section 5 describes the model checking of the proposal. Simulation results are analyzed in section 6 which is followed by section 7 that illustrates few of the possible use cases of the proposed concept. Finally section 8 concludes the work and a guideline for future work is presented.

2. Related Work

There have been many researches about clustering in the WSNs. Most of the previous researches have been on preserving energy and prolonging the battery for the resource-constrained sensor nodes. For example, LEACH [4] is the first clustering technique for achieving network longevity and energy dissemination reduction. Padmanabhan and Kamalakkannan in [5] further modified LEACH to prolong the network stability. Kumar et al. in [27] also examined different LEACH techniques in a view to prolong network lifespan. Literatures in [8 – 10, 24] discuss clustering of sensors but for the sake of data-accumulation. Clustering of sensors helps in reducing energy consumption, stabilizing network, efficient routing etc. S. Bandyopadhyay et al. in [8] analyzed hierarchical clustering and discussed that energy consumption is decreased if clustering level of hierarchy is increased. Abbasi, Younis and Lotfinezhad, Liang in [9-10] mention that inter-cluster communication is only limited to cluster-heads which results in communication bandwidth saving and in reducing message exchanges between sensors. Hyun and Hyuk in [28] discussed that efficient cluster-head selection prolongs the network life span and saves energy. D. Ma et al. in [25] proposed a clustering protocol with dual cluster-head concept to further improve network life time and more data accumulation to the base station.

Lombriser et al. in [11] presented distributed processing of context for dynamic WSNs. Their proposed E-SENSE computes context information from sensor networks. Sensors are clustered based on context-activity but limited only to neighboring sensors. It does not solve large-scale sensor network issue. Franco in [6] envisioned the idea of sensing, actuating and computing of anything anywhere for the future pervasive computing. He further outlined that spontaneous human participation i.e. crowdsourcing is vital for distributed collaboration to enrich urban networks. G. Barbier et al. in [7] presented maximizing the data obtained through crowdsourcing. They portrayed that crowdsourcing is faster and beneficial. With crowdsourcing, any event can be detected and analyzed. Event in the urban areas are fast changing. Moreover, some events are recursive and some are non-recursive [12]. Scalability, reachability and reliability of the obtained data from urban events through crowdsourcing become a challenge. Guo and Han in [13] discussed the reliability issue in data collection for WSN. They discussed the essence of reliable data collection for mobile nodes. The importance of latency in reliability for mobile WSNs has been discussed by Y. Rao et al. in [26]. They proposed a clustering based routing protocol for reliable data packet delivery in real-time. Ericsson in [1] further outlined the significance of reliability, latency, delay, maximum service (data) delivery i.e. reachability etc. for the future Networked Society.

Luca and Gian in [14] introduced logical-neighborhood of sensor nodes which replaced physical neighborhood concept. This idea more or less resembles our proposal. However, their solution is a programming language abstraction where nodes are said to be in the logical neighborhood if certain attributes are satisfied. A programmer defines the nodes' attributes and the data segment that can be part of a neighborhood. Therefore, it does not explicitly solve the real-time context sharing issue which is the prime objective of our proposal. In this paper, our focus is to examine the network performance of the logical-clustering of flow-sensors.

3. Motivation

Traditionally, sensors are used to obtain data from physical objects. Sensors also collaborate to achieve common goals. With the technological advancement, sensing devices have become more intelligent and affordable. Hence, the applicability of sensors is always rising, and it is believed that billions of sensors will be deployed in the future. Moreover, sensors are fundamental in the Internet of Things (IoT) deployment for any kind of urban event detection. These are used for different purposes and to obtain heterogeneous data from distributed sensor networks. Sensors deployment can be both deterministic (fixed) and random (mobile). Therefore, real-time context sharing will be a big challenge to existing and later distributed WSNs applications. Earlier solutions do not provide proper management of context information; hence current context information management does not support real-time sharing of context and do not scale well for heterogeneous interoperability. This necessitates proper management of the obtained data i.e. context information from sensors in order to use in an efficient and useful way. Most researches thus far concentrated on decreasing energy consumption so that sensors longevity is ensured. Several researchers have worked on clustering sensor nodes too, but again largely for sensor nodes stability. There have been some proposals for sensors data-management, but their proposals restrict to a certain area for adjacent sensors. It is also important that context generated by the sensors should be used meaningfully to take its full advantages. Real-time context sharing will be beneficial when clustered based on context similarity. The idea of clustering the sensors logically based on context similarity would allow resources (data, services) to be shared. Furthermore, the idea will provide topological sensor networks with scalability, reliability and high reachability in terms of delay, latency and packet loss.

4. System Model

Some of the definitions that have been used for modeling the system are presented below.

Sensor-ID: Sensors should have unique IDs. A *sensor-ID* can be obtained in different ways, e.g. the ID can be chosen randomly or can be obtained by hashing the sensor IP or MAC address [15].

Flow-ID: *Flow-ID* is the logical identification of the flow. According to [16], a flow could be defined based on capabilities of a particular implementation. The flow-ID is the flow packets from a particular sensor to the sink. As long as the sensor is interested in the same flow packets, the flow-ID remains same. But if sensor changes the flow of packets, the flow-ID is also changed. OpenFlow flow-tables consist of match-fields (i.e. packet header), action sets and statistics. The packet header defines the flow and action defines the flow-ID.

Context-ID: The *context-ID* is the identifier of a cluster. This can be compared to the idea that of a hashtag. As hashtag groups the similar messages, context-ID has the same objective. Context-ID is a mean of clustering similar data. The context-ID is published to the internet through the logical-sink and any interested entity i.e. sensor can subscribe to the context-ID.

Context flow-table: OpenFlow specification implies that match fields can be defined according to the research requirement [16]. A new flow-table for flow-sensor has been defined which includes flow-sensor's sensor-ID, flow-ID and context-ID. This flow-table is named *context flow-table*.

4.1 Network

A two-tier H-DHT system model has been considered. Controlling the ever-increasing number of sensors would be infeasible for single logically centralized controller (current OpenFlow standard), and in order to scale well for enormous number of sensors, the idea of HyperFlow (HF) has been exploited. This implies that multiple numbers of controllers (sinks) in the network has been used. The sinks are physically distributed but logically synchronized, hence this idea has been defined as *logical-sink* [2]. Another advantage of utilizing logical-sink is that each sink can do processing locally. And then other sinks get notified of the local changes and thereby synchronized. The network is divided into two-tier hierarchy (fig. 1). In the top-level overlay, CHORD concept is applied. And in the bottom-level hierarchy, the flow-sensors are clustered in single-connection manner. Flow-sensors communicate with the logical-sink. Sink that is part of a cluster virtually acts as a flow-sensor with very high-computational capabilities. This eliminates the burden of choosing or electing a cluster-head. This virtual flow-sensor can be thought as the cluster-head (one for each cluster). These virtual flow-sensors i.e. cluster-heads are organized in the top-tier overlay as CHORD. In fig. 1, there are three clusters that communicate with the logical-sink. And, for each cluster there is a virtual flow-sensor. A flow-sensor does not need to concern about the physical sink the communication takes place as all the physical sinks are synchronized and aware of any change inside the network.

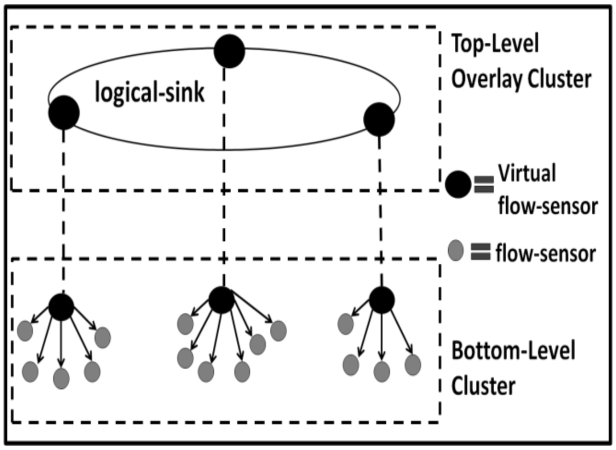


Fig. 1. The two-tier Network

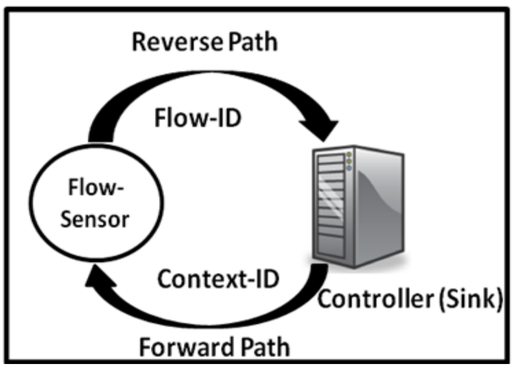


Fig. 2. Communication

4.2 Communication

The communication is shown in fig. 2 and is of three kinds: logicalsink-to-sensor, sensor-to-logicalsink and sink-to-sink. Logicalsink-to-sensor communication takes place in the forward path. This communication is straight forward in a sense that sink has better communication capabilities and can communicate with flow-sensors directly. For any exception, the communication can still take place through distributed collaboration. In the reverse path, sensors communicate with logical-sink via overlay hop [15]. Those sensors that are not part of a particular cluster collaborate with other sensors so that sensors can reach nearby logical-sink. Sink-to-sink (inside a HF network) communication can further be divided into two: physical and virtual sink-to-sink. The physical communication among sinks follows the same procedure as in HF. And, the virtual communication implies the communication between virtual flow-sensors and a CHORD top-level overlay is formed by the virtual flow-sensors. Hence, this communication follows the idea of CHORD.

4.3 Implementation

Both fixed and mobile flow-sensors have been assumed. Flow-sensors traffic are controlled and managed by logical-sink. The flow-sensor usually has flow-tables in the hardware layer [3]. Each flow-table contains flow-entries and an action for each flow-entry which decides flow routing. Each flow-entry has match-fields that define the flow, instructions correspond the way packets should be routed, and statistics takes care of packet updating. Packets from flow-sensors are matched in each flow-entry, instruction set defines the flow-ID if already not available, and statistics updates the packets. Statistics checks if the current packet matches the old packets, otherwise a new flow-ID is defined for any mismatch. Flow packets are then forwarded to the nearby physical sink in the reverse path. The flow packets include the flow-ID. The logical-sink maps flow-ID and returns the corresponding context-ID in the forward path, a sensor-ID is also returned to the flow-sensor if already not assigned. The sensor-ID is unique and unchanged for a flow-sensor. In case the context-ID is not available with the contacted physical sink, this sink contacts other physical sinks and the corresponding context-ID is returned. Search will follow the CHORD look-up mechanism. Viewed this way, the context-ID search will also follow the similar procedure. The logical-sink modifies the context flow-table with the context-ID along with sensor-ID and flow-ID. Logical-sink also updates the group table with the context-ID. By the mean time, other sinks get notified about all the changes in each sink and get updated thereby. In case the received flow-ID does not match any existing context-ID, then logical-sink defines a new context-ID. This context-ID is then published to other HF networks. When any sensor is interested in the context-ID in other network, then sensors subscribe to the context-ID. The algorithm for above is as follows:

- Flow-sensor match-fields define the flow and the action defines the flow-ID
- Flow-ID is sent to the nearby physical sink S1
- S1 resolves flow-ID and returns corresponding context-ID
- S1 returns the sensor-ID if already not assigned
- S1 forwards the request from flow-sensor to other physical sinks (S2, S3… Sn) if no match found for the request in S1
- If no context-ID found in the logical-sinks then a new context-ID is defined and published to other networks

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

- Logical-sink returns the context-ID to the requested flow-sensor
- Regular and context flow-tables are updated by the logical-sink
- Statistics check for new and old packet mismatch, new flow-ID is defined in case of any mismatch

4.4 Example Scenario

Fig. 3 shows an example of MATLAB implementation. There are 4 H-DHT HF net-works with 50 sensors. Some are fixed (16) and some are mobile (34). The sensors have been clustered based on context-similarity. Different cluster is represented by different color. As seen that sensors might be resided in different networks but they are logically clustered and belong to same context-ID. Each HF network has been facilitated by four sinks ('+' signs). Their positions are fixed and act as single logical-sink for single HF network. It is assumed that sinks are placed carefully so that all the flow-sensors are covered. This explains how logical clustering of sensors can be achieved.

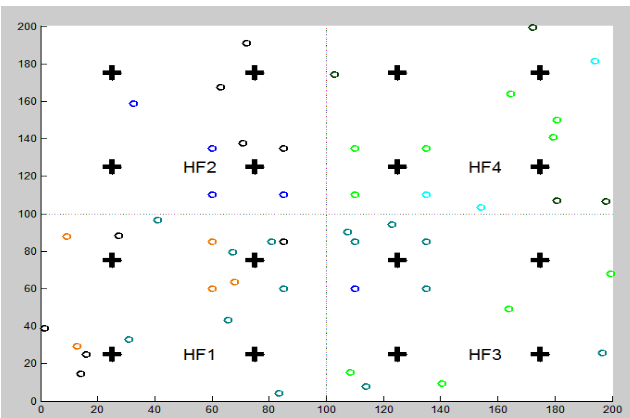


Fig. 3. An example scenario

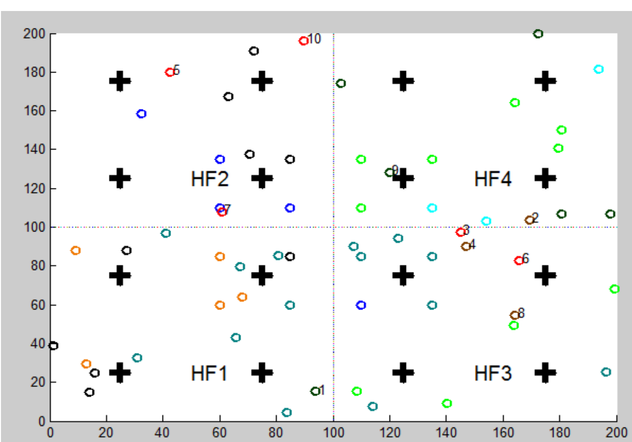


Fig. 4. Sensors joining

4.5 Sensors Joining

Fig. 4 shows 10 sensors (depicted by 1 to 10) joining. It can be seen that sensors (1, 9) have joined an existing cluster; while sensors (2, 4 and 8) and rest of the sensors have formed two new clusters respectively. These can be distinguished by different colors. When new sensors join the network, they send their flow-IDs to the nearby sinks. Context-IDs are shared by all the logical-sinks, and all the logical-sinks have the knowledge of existing context-IDs. Therefore, when sensors send their flow-IDs, then logical-sink checks the existing context-IDs. If match found, then new sensors are said to have subscribed to the existing context-ID. Otherwise, logical-sink defines new context-ID based on the received flow-IDs and context similarity. And, the sensors form new clusters.

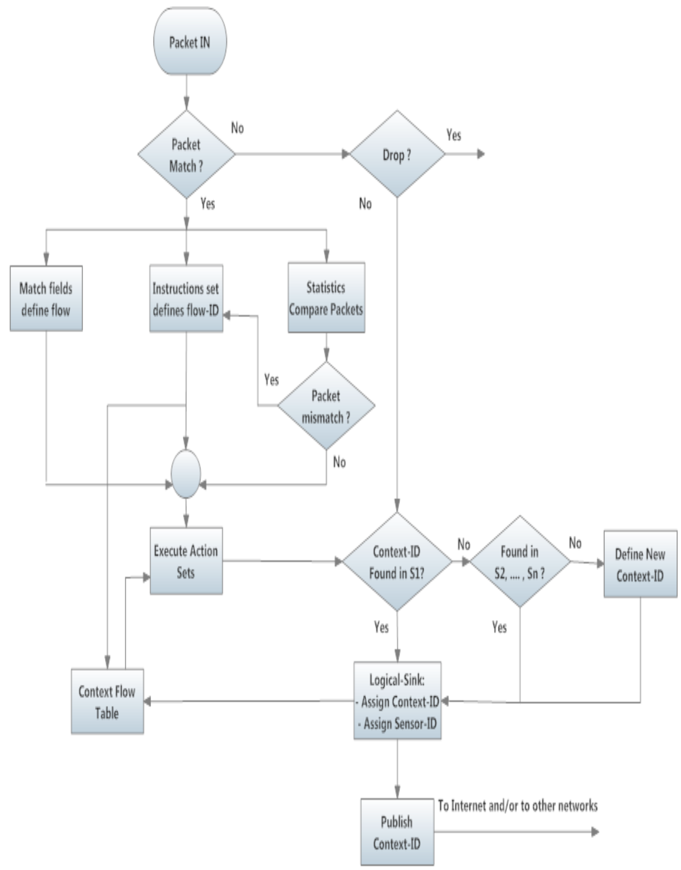


Fig. 5. Flow chart

5. Model Checking of the Concept

The combination of PROMELA and SPIN has been used for simulation and verification of system model in [17-19]. It provides versatility and is very useful for model checking. The combination has been used extensively for modeling and verifying communication protocols [3]. The proposed model of this paper has been examined using the PROMELA and SPIN combination. First, fig. 5 shows the flow chart of the proposed model. The explanation of the flow chart has been described already (see 4.3).

5.1 Context-ID Match Algorithm

The following algorithm defines the mechanism for communication between sensor nodes and sinks. The first process (proctype node) defines sensor nodes flow send and receive method, and the second process (proctype sink) defines the mechanism for logical sink.

```
/*Algorithm for context-ID definition
or matching*/

bool flow_id, sensor_id, context_id;
proctype node(chan in, out) {
#define node_add  /*define address of
the sensor node*/
int pkt; /*packet*/
bool chk;
xs src_node; /*send channel of source
node*/
xr sink_add; /*receive channel of
sink*/
in?input_port,dst_add; /*Channel sends
input port number and destination
address*/
if
:: (src_node == node_add && pkt! =Null)
-> out!pkt; goto pkt_match; /*if
address is authenticated and packet is
not empty, send packet to check for
packet matching*/
fi;
pkt_match: in?pkt
if
:: (chk = true) -> goto
pkt_send2flowtable; /*if packet is for
matching, send to flow table*/
:: (chk = false) -> goto pkt_drop;
/*check if packet is to be dropped*/
fi;
pkt_drop: in?pkt
if
:: (chk = true) -> skip; /*Packet is
dropped*/
:: (chk = false) -> pkt_send2sink;
/*Packet is forwarded to the nearby
physical sink*/
fi;
pkt_send2flowtable: in?pkt
if
:: (input_port == 1) ->
write(match_fields); /*If Packet is not
empty, update the match fields*/
:: (input_port == 2)
write(instructions_set); /*Update
instructions set and define the flow-
ID*/
:: (input_port == 3) write(stat);
/*Update Statistics- store packet
information*/
:: goto pkt_send2sink; /*Packet is
ready to be sent to the nearby sink*/
else -> skip; /*Drop Packet, if empty*/
fi;
pkt_send2sink: in?pkt
read(sink_add); /*Get the address of
the nearby sink*/
read(match_fields); /*Check the match
fields for flow*/
read(instructions_set); /*Check for
flow-ID*/
read(stat); /*Check for any packet
mismatch*/
if
:: (flow_id = true) -> goto
context_flowtable; /*If flow_id is
found then insert to the context flow
table*/
fi;
context_flowtable:
read(instructions_set);/*Update the
context flow table's flow-ID field*/
end; /*End the process*/
}
proctype sink(chan in,out) {
#define dst_add /*Define the current
sink address*/
if(dst_add == sink_add && pkt! =Null) -
> goto flow_match; /*If the sink
address is authenticated and packet is
not empty, then check for flow
matching*/
fi;
flow_match: in?pkt
read(match_fields); /*Check the match
fields for flow*/
read(instructions_set); /*Check for
flow-ID*/
read(stat); /*Check for any packet
mismatch*/
if
```

```
:: (flow_id = true) -> out!context_id
/*If flow-ID matches any existing
context, send the context-ID*/
:: (sensor_id = false) -
>write(sensor_id) /*If no sensor-ID is
assigned, assign the sensor-ID*/
:: out!sensor_id; /*Send the sensor-
ID*/
:: goto context_flowtable; /*Go to the
context flow table to update the table
fields*/
:: goto publish; /*Go to publish if
context-ID is ready to be published*/
::else goto sink_n; /*If flow does not
match any context-ID in the current
sink, go to other sinks*/
fi;
sink_n: in?flow_id
if
:: (flow_id = true) -> out!context_id
/*If flow-ID matches any existing
context, send the context-ID*/
:: goto context_flowtable; /*Go to the
context flow table to update the table
fields*/
:: goto publish; /*Go to publish if
context-ID ready to be published*/
::else write(context_id); /*If no
context-ID found for the flow, define a
new context-ID*/
fi;
context_flowtable:
in?flow_id,sensor_id,context_id
write(stat); /*Update the statistics
with IDs*/
publish: in?context_id
if
:: (context_id = false) ->
write(context_id); /*If context-ID is
not yet published, publish the ID*/
fi;
end; /*End the process*/
}
init { /*Initialize the processes*/
  chan send = [2] of {int, bool};
/*Send channel would carry two
different type of messages*/
  chan rcv = [2] of {int, bool};
/*Receive channel would carry two
different type of messages*/

  run node(send,rcv); /*run the node
process*/
  run sink(send,rcv); /*run the sink
process*/
}
```

6. Simulation Results

Table 1: Simulation parameters

Parameter	*Value*
Number of Networks	3
Number of Nodes	60
Number of Groups	3*
Nodes per Group	9*
Packet Flow Rate (per second)	8*
Packet Size	512* bytes
Routing	Static
Propagation Path Loss Model	Fixed RSS Loss Model
Delay Model	Constant Speed Propagation Delay Model
Error Model	ns-3 YANS Error Model
Sensors Mobility Model	Random Walk 2d Mobility Model
Receiver Noise Factor	10.25 dB
Received Signal Strength (RSS)	-95 dBm
Total Number of Transmitted Packets	2000
Physical Model	IEEE 802.11b
Data Rate	1 Mbps
* = varies in different simulations	

A network has been designed and simulated in the ns-3 simulator. Simulation parameters are tabulated in table 1. The focus of this paper was not to verify the physical layer behaviors, hence the sensor node reachability, interferences, received signal strength (RSS), energy consumption, and signal-to-noise ratio impacts have not been explored. These are beyond the scope of this work. The focus largely lies on the behavior of the system with regard to real-time context sharing. Therefore, the simulation has been carried on constant values of RSS, receiver noise factor, etc.

6.1 Simulated Network

Fig. 6 shows the network that was designed and simulated in the ns-3. Although our proposal makes use of multiple distributed and synchronized OpenFlow controllers (logical-sink), but ns-3 as of now does not allow external controller for OpenFlow [20]. Hence, we stick to the current ns-3 implementation. As for H-DHT for sensor nodes and context-IDs management, this is also left for the upcoming paper as no working model of H-DHT is available right now in ns-3 [21]. In the designed network, there are three wireless sensor networks as seen in fig. 6. Sensor nodes in network 1 are fixed while sensor nodes in both network 2 and 3 are mobile (randomly moving). Each network has one gateway and gateways are connected by the OpenFlow controller. Each network has 20 sensor nodes. Other parameter values can be found in table no. 1.

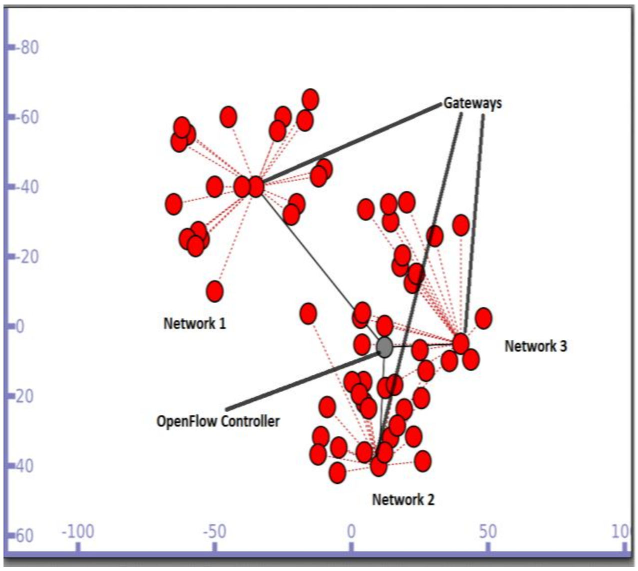


Fig. 6. Simulated Network

6.2 Performance Measurement

In this section, performance measurement for various scenarios has been presented. For the evaluation, reliability, scalability and reachability metrics have been chosen and the proposed approach has been highlighted with respect to these metrics. The importance of reliability and scalability has been suggested by the earlier researches. As for reachability, we believe that packet reachability would be an important performance metrics in the real-time context sharing e.g. in urban event detections.

6.2.1 Effect of Varying Flow-Rate

Firstly, the performance has been measured for different flow rate i.e. number of packet per second (p/s). The number of node per group and the group size has been kept unchanged for this particular evaluation. There are total 3 groups for this scenario and each group has 9 nodes. The packet size for this scenario has been kept to be 512 bytes.

Mean Delay Performance

Fig. 7 shows mean delay performance for the simulated network of each group for different flow of packet. Packet flow varies between 6 and 11 p/s. X-axis shows the packet flow rate and y-axis shows the mean delay of each group. It can be seen from the figure that at the start, i.e. for packet flow of 6 p/s, each group more or less demonstrates similar results with respect to delay. All groups qualitatively demonstrate similar performance for packet flow rate up to 10 p/s. While the packet flow is increased to 11 p/s, all groups show increase in the delay for 11 p/s. It can be seen that delay is increased with the increase in the packet flow rate, however, it does not incur high increase up to 10 p/s. Group 1, 2 and 3 mean delay increase by 0.3172s, 0.2629s and 0.2166s respectively for 11 p/s i.e. 83 % increase in the packet flow rate.

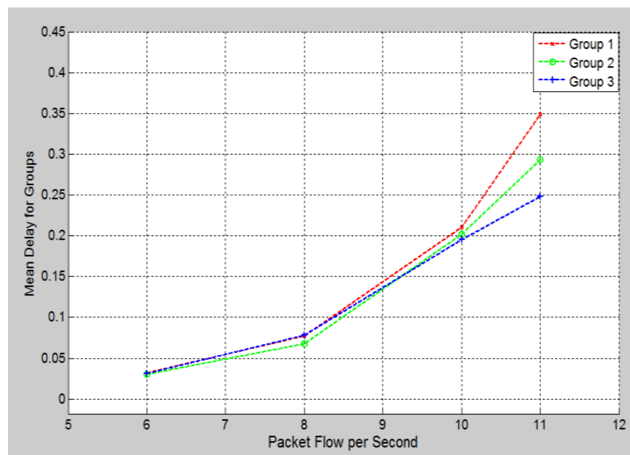


Fig. 7. Mean Delay Performance

Mean Jitter Performance

Mean jitter performance for each group for different flow of packet can be seen in fig. 8. Mean jitter performance demonstrates similar pattern like mean delay. At the beginning, all groups show similar jitter performance. Like mean delay of the packet, jitter does not encounter a performance degradation when packet flow is increased. Jitter for 11 p/s increases by only 0.0347s, 0.0327s and 0.0313s for group 1, 2 and 3 respectively compared to 6 p/s. The increase is very minimal. Therefore, from fig. 7 & 8 it can be concluded that packet delivery is reliable with minimum delay and jitter.

Packet delivery with minimum delay and jitter is very significant issue in crowdsourcing paradigm and for any urban event detection. When crowd generate data, it should be delivered swiftly. Service requesters would want to access data in the shortest possible time. Moreover, reliability of packet delivery in real-time context sharing largely depends on how quickly service is delivered. Another important characteristic scalability can be seen from fig. 10 &11. As for the packet flow rate, packet size plays an important role defining flow rate. Fig. 16 shows the impact of changing packet size.

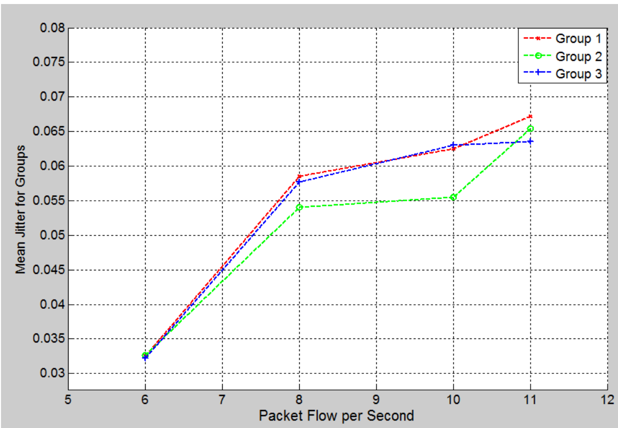


Fig. 8. Mean jitter performance

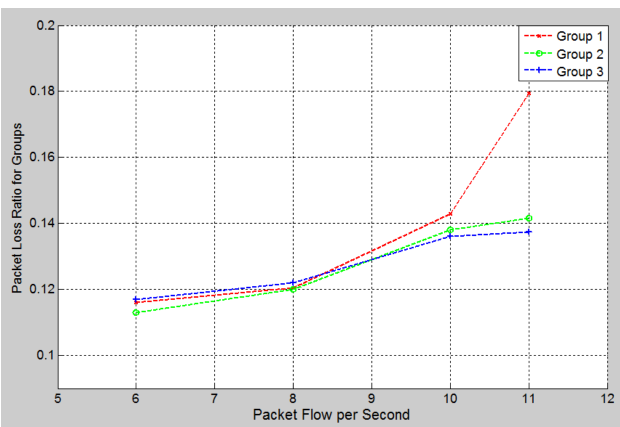


Fig. 9. Packet Loss Ratio

Packet Loss Ratio

Reachability of packet principally depends on number of packet loss. Fig. 9 shows packet loss ratio of each group. Although each group demonstrates similar pattern in packet loss for packet flow rate up to 10 p/s. Group 1 shows a rise in the packet loss for flow rate 11 p/s. Packet loss ratio increases with the increase in the flow rate. For the packet size of 512 bytes, packet loss ratio does not incur a high fluctuation for flow rate up to 10 p/s. This assures high reachability of packet. Group 1, 2 and 3 incurs packet loss ratio increase by 0.0635, 0.0285 and 0.0205 respectively compared to flow rate of 6 p/s. In terms of percentage, the increase is 55 %, 25% and 18% compared to flow rate of 6 p/s. This increase of packet loss is for 83 % increase in the flow rate. From this, it is clear that our proposal assures rich packet reachability. In the real-time urban event detection, this high reachability of packet would be very beneficial. This will ensure sharing rich amount of urban event detections.

6.2.2 Effect of Increasing Nodes per Group and Group Size

In the previous section, reliability and reachability have been discussed for variant flow of the packets. Although in the urban event detection packet flow rate would always be fluctuating, however, at the same time the participants in data acquisition i.e. sensor nodes in this case would also vary. This implies that different context would be generated which leads to different clustering of contexts i.e. group of data. Scalability becomes a significant consideration with respect to increasing number of nodes and groups for real-time context management. Here, effect of increasing nodes per group and group size is discussed.

Mean Delay Performance

Fig. 10 shows the mean delay performance for variant number of nodes per group. As seen earlier that performance degrades from 11 p/s, and packet flow rate has been kept constant at 10 p/s for this evaluation. As expected, delay increases with the increase in the node per group. If the node per group is doubled then mean delay increases by 17 %, 18 % and 15 % respectively for group 1, 2 and 3. And, if the node per group is tripled i.e. increase by 200% then group 1, 2 and 3 incurs mean delay increase by 25 %, 20 % and 22 % respectively. This clearly shows that the proposed concept scales well for increased number of node per group. Figure 11 shows impact of increasing the number of groups. The figure illustrates only results for group 1 and node per group has been kept steady (9 node per group). It has been evaluated for different packet flow rate. As seen from figure 11, mean delay increases nominally with increase in size of the groups. For instance, for the packet flow rate of 5 p/s, group size of 6 incurs 6 % delay increase compared to group size of 3. For the packet flow rate of 8 p/s, the delay increased to 19 %. It is seen that only 13% delay fluctuated when flow rate is increased by 60 % and group size is doubled. However, it is observed that for group size of 6 with flow rate 9 p/s, mean delay decreased. This is due to the fact that packet loss for this scenario is higher due to probable wireless interferences and random nodes' movement. It can be concluded that the proposed concept provides scalability for delay in terms of increased node per group and increased group size.

Mean Jitter Performance

Mean jitter performance for scalability is shown in next two figures. Fig. 12 shows the effect of changing node per group on jitter. As was the case with the mean delay, jitter also understandably increases with raise in the node per group. If node per group is doubled (100 % increase) then jitter increases by 33 %, 35 % and 33% for group 1, 2 and 3 respectively compared to 6 nodes per group. And if the node per group is increased by 200% then group 1, 2 and 3 encounter jitter increase by 41 %, 40 % and 35 % respectively compared to 6 nodes per group. This clearly shows the proposed concept scales well in terms of jitter too. Figure 13 further shows mean jitter performance while keeping node per group constant (9 node per group), and varying the size of the group and flow rate. The figure evidently demonstrates that jitter fluctuates nominally for the aforementioned scenario. For the flow rate of 8 p/s, jitter demonstrates only 44 % fluctuation for 100 % increase in the group number. Due to packet loss it is observed that jitter decreases for flow rate of 9 p/s. This packet loss depends on the flow rate and packet size. Figure 16 clarifies effect of packet size variation. This low jitter fluctuation will particularly provide advantage in crowdsourcing paradigm, when there would be different number of clusters of context generated by crowd. Different clusters of context imply different types of urban events detection. Therefore, our proposal can cope in terms of reliability with reliable real-time context sharing for changeable number of nodes and clusters.

Packet Loss Ratio

Along with the reliable context delivery it is also imperative that context delivery ratio is high and scalable at the same time. Figure 14 shows the packet loss ratio performance for variant number of node per group. As seen from the figure 14, when node per group is increased by 100% then group 1, 2 and 3 respectively have packet loss increased by 22 %, 22 % and 41 %. And, if the node per group is increased by 200% i.e. to the full capacity of the designed network, packet loss ratio increases by 71 %, 48 % and 102 % for group 1, 2 and 3 respectively. Group 1 and 3 exhibited higher packet loss compared to group 2. Also seen from the figure, for packet flow rate of 9 p/s, group 1 has a leap in the packet loss for group size of 5 and 6. This higher ratio is due to random movement of the nodes and wireless interferences. Compensating these effects is beyond the scope of this work. It can further be seen from figure 15 that group size of 6 exhibits packet loss ratio degradation only by 26 % and 33 % compared to group size of 3 (100 % increase in the group size) for flow rate of 5 p/s and 8 p/s respectively. This confirms that the proposed idea is scalable for packet reachability as well.

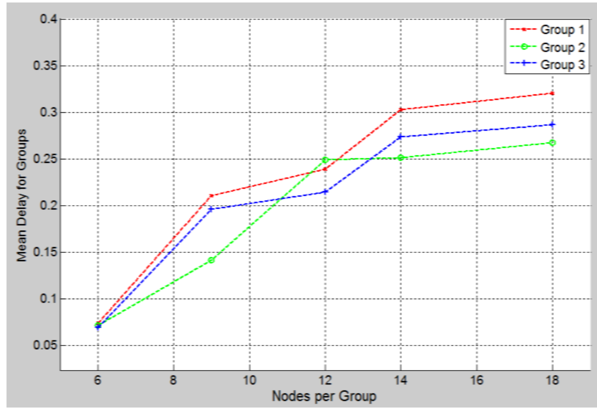


Fig. 10. Mean delay performance for different number of nodes per group

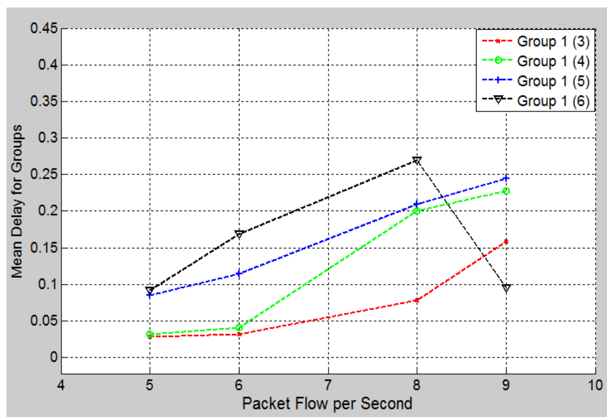


Fig. 11. Mean delay for variant size of groups

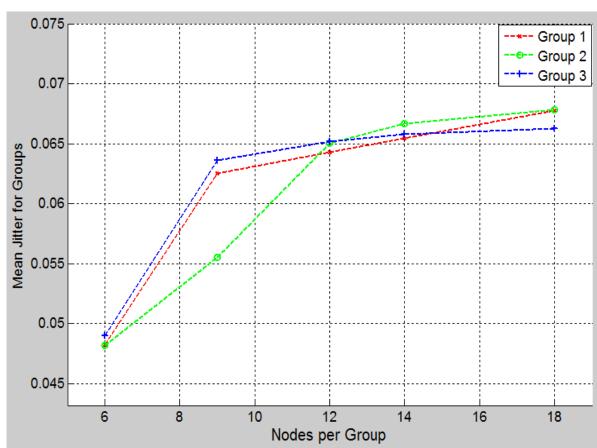


Fig. 12. Mean jitter for alternate number of node per group

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

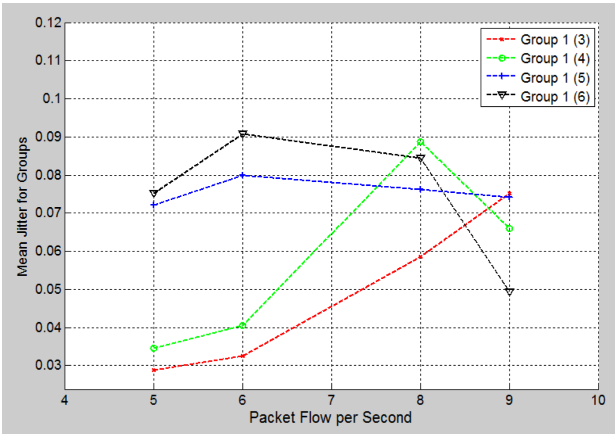


Fig. 13. Mean jitter for variant size of groups

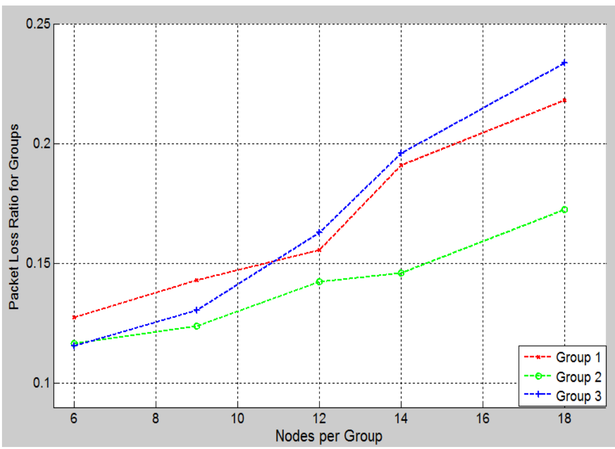


Fig. 14. Packet loss ratio for different number of node per group

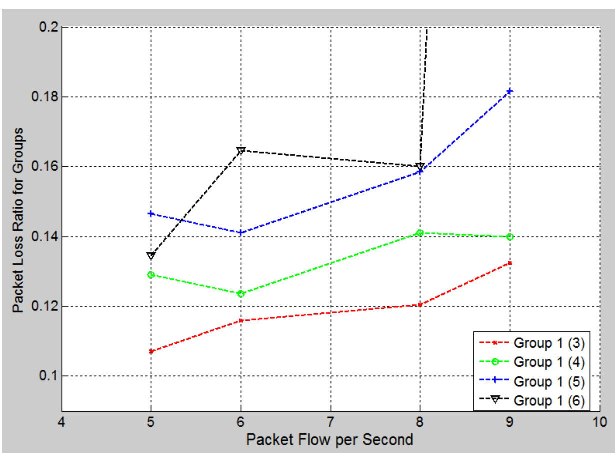


Fig. 15. Packet loss ratio for variant number of groups

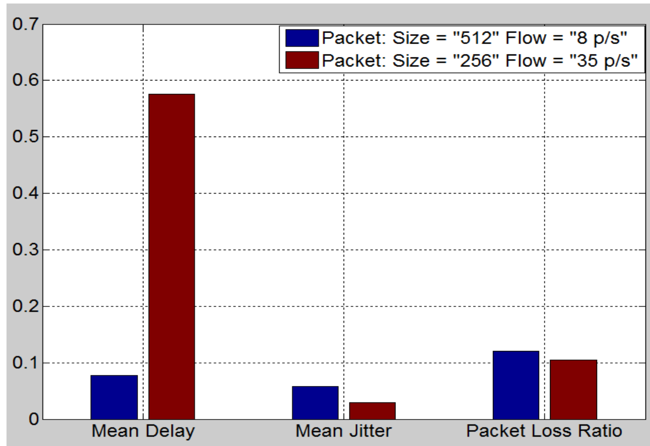


Fig. 16. Packet size impact on packet flow rate

6.2.3 Different Packet Size

The above results have been evaluated for a particular packet size of 512 bytes. It is clear that for this packet size, performance metrics shows better results if packet flow rate is below or equal to 10 p/s. Now fig. 16 shows how packet size affects the packet flow rate. It can be seen that if the packet size is halved then packet flow rate increases by 338 %. However, mean delay also increases by 642 %. As for mean jitter and packet loss ratio, these metrics decreased by 98 % and 15 % respectively. From this it can be concluded that in the crowdsourcing paradigm or in the urban event detection, if the generated data i.e. context is small then packet flow rate will be high. However, this might result in high delay but jitter and packet loss ratio would be lower. Therefore, with high packet flow rate, the idea can scale well for jitter and packet loss ratio but delay performance might degrade.

7. Use Cases of the Concept

The proposed approach would be useful for heterogeneous interoperability of physical objects, thereby heterogeneous contexts. In our opinion, this logical-clustering will be advantageous to many sensor network applications; for example, medical science, agriculture system, security surveillance, disaster management etc. Two probable scenarios are portrayed below.

7.1 Animal Tracking

The use of WSN for animal tracking is gaining tremendous attention recently [22]. The animal tracking can be further divided into two: wildlife and farming monitoring. Our proposed approach can be applied in both of them. One probable application, for example, farmers can optimize their business by means of WSN for feeding and growing conditions of the animals [22]. This will provide benefit for monitoring meat, milk production and to observe how good animal racing results. Hence, animal tracking would be easier by applying our proposed logical-clustering approach. This implies that animals that produce similar desired context would be clustered together. In a large farm, it is often difficult to manage the animals efficiently; it would require incredible manpower to monitor all the animals. Therefore, animals' location and conditions can be monitored by clustering. The farmers can find the groups of animal ready for meat and milk production through the context-ID. This will reduce human labor to find out the animals for the above mentioned purposes.

7.2 Medical Healthcare

The approach can be applied in the medical healthcare too. One possible application scenario can be that medical researchers can conduct a research in real-time on a recently spread disease from remote places and provide prompt solutions simultaneously. Normally in medical healthcare, patients are outfitted with wireless wearable sensors [23]. If there is any outbreak of a disease, then people can be outfitted with wearable sensors. Medical team can observe the severity of the disease by means of clustering different symptoms and different level of patients. For example, patients might not have same level of severity and they would need different level of attentions. Hence, medical assistance can be provided faster and efficiently by clustering patients based on the different context of severity. Patients that show similar symptoms would be clustered together and would have same context-ID. This will eliminate burden of individual attention for a patient. Thus medical services can be maximized. Moreover, if the situation gets worse and out of control then medical personnel can seek help from other specialists in the respective field from remote places. In the traditional way, this can be done by gathering data from people and then forwarding to others. However with our proposed logical-clustering, medical researchers from distant places can instantly access the data by subscribing to the context-ID.

However, our proposal is not limited to these scenarios rather this shows two of the many possible solutions our proposal can offer.

8. Conclusions

Real-time context sharing would be an important challenge in state-of-the-art ubiquitous computing. The enormous data that are expected to be generated by the billions of sensors would require efficient management. These huge heterogeneous data would need to be processed reliably, and reachability should also be assured to take its full advantage. Location agnostic clustering of flow-sensors i.e. logical clustering is one of the possible solutions for efficient context management. In this paper, performance of logical-clustering in terms of delay, jitter and packet loss ratio has been examined and backed up by ns-3 simulations. These parameters have been evaluated for different scenarios such as: variant packet flow rate, different number of node per group and different group size.

The results suggest that the proposal is reliable and scalable. For a 200 % increase in the node per cluster, delay increases by around 20 %. For the same increased node per cluster, latency demonstrates around 40 % increase. Delay and latency exhibit 13 % and 44 % fluctuation respectively when the cluster size is doubled. This clearly illustrates efficiency of reliability and scalability of the proposed concept. Furthermore, the proposed approach shows rich packet reachability. Only 18 % increase in the packet loss for a flow rate increase of 83 %, packet loss increases by merely 44 % for 200 % increase in the node per cluster. Moreover, packet loss demonstrates no more than 33 % increase for 100 % increase in the cluster size for high flow rate.

The proposed approach in particular can perform more efficiently for smaller packet size as suggested by the result. Flow rate increased by 338 % when packet size is halved. Latency and packet loss ratio further decreased by 98 % and 15 % respectively. Two tangible use cases have also been portrayed. Therefore, our proposed idea will be of great interest for the future Networked Society where instantaneous and reliable accesses to context are two of the big challenges. Our approach can adopt quickly and share real-time data reliably to the service requestors. The vision of detecting any urban event via crowdsourcing paradigm will be made easier through the adoption of our proposal.

However, the approach can perform better than the results obtained in this paper through real implementation of logical-sink and H-DHTs in ns-3. The logical-sink would outperform the current packet reachability; reliability and response time would also be minimal. Future work includes designing the system with logical-sink and inclusion of H-DHTs. An investigation into routing protocol for the logical-clustering would also be explored.

References

[1] [online] 5G Radio Access, Research and Vision: http://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf [Accessed: 15-July-2013]

[2] R. Rahmani, H. Rahman, and T. Kanter: Context-Based Logical Clustering of Flow-Sensors - Exploiting HyperFlow and Hierarchical DHTs, In Proceeding(s) of 4th International Conference on Next Generation Information Technology, 2013 ICNIT, June 2013.

[3] A. Mahmud., R. Rahmani: Exploitation of OpenFlow in Wireless Sensor Networks, IEEE ICCSNT, volume 1, pages 594-600, 2011.

[4] W. R. Heinzelman, J. Kulik, and H. Balakrishnan: Adaptive Protocols for Information Dissemination in Wireless Sensor Networks, Proc. ACM MobiCom '99, Seattle, WA, 1999, pp. 174–185.

[5] K. Padmanabhan, Dr. P. Kamalakkannan, " Energy Efficient Adaptive Protocol for Clustered Wireless Sensor Networks", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011.

[6] Franco Zambonelli: Pervasive urban crowdsourcing: Visions and Challenges, 2011 IEEE (PERCOM Workshops), Pages 578-583, March 2011.

[7] G. Barbier, R. Zafarani, H. Gao, G. Fung, H. Liu: Maximizing benefits from crowdsourced data, Computational and Mathematical Organization Theory, September 2012, Volume 18, Issue 3, pp 257-279.

[8] S. Bandyopadhyay, E. J. Coyle: An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks, IEEE (INFOCOM), Volume 3, Pages 1713-1723, April 2003.

[9] A. Abbasi, M. Younis: A survey on clustering algorithms for wireless sensor networks, Journal of Computer Communications, 2007, 30, pp. 2826–2841.

[10] M. Lotfinezhad and B. Liang: Energy efficient clustering in sensor networks with mobile agents, in Proceedings of the IEEE (WCNC '05), pp. 1872–1877, usa, March 2005.

[11] C. Lombriser, M. Marin-Perianu, R. Marin-Perianu, D. Roggen, P. Havinga, G. Troster: Organizing Context Information Processing in Dynamic Wireless Sensor Networks,"Proc. ISSNIP,pp. 67-72,December 2007.

[12] T. Thomas, E.C. van Berkum: Detection of incidents and events in urban networks, IET Intelligent Transport Systems, September 2008.

[13] Lin Guo and Qi Han: Reliable Data Collection from Mobile Users with High Data Rates in Wireless Sensor Networks, 2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Pages 1-6, June 2012.

[14] L. Mottola and G. P. Picco: Logical Neighborhoods: A Programming Abstraction for Wireless Sensor Networks, Distributed Computing in Sensor Systems, Volume 4026, 2006, pp 150-168.

[15] G. Fersi, W. Louati, M.B. Jemaa: Distributed Hash table-based routing and data manage-ment in wireless sensor networks: a survey, Wireless Networks, Volume 19, Issue 2, Pages 219-236, February 2013.

[16] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, J. Turner: OpenFlow: Enabling Innovation in Campus Networks, ACM SIGCOMM Computer Communication. Rev., Volume 38 issue 2, pages 69-74, April 2008.

[17] Gerard J. Holzmann: The Model Checker SPIN, IEEE Transaction on soft ware Engi-neering ,23(5): 1-17, May 1997.

[18] Gerard 1. Holzmann, SPIN Online Reference. Bell Labs, http://cm.bell-labs.com/cm/cs/what/spin/Man lindex.html)August 1997.

[19] Rob Gerth: Concise Promela Refence. Technical report, Eindhoven University, (htt p://cm.bell-labs.com/cm/cs/what/spiniManiQuick.html) June1997.

[20][online] http://www.nsnam.org/docs/release/3.13/models/html/openflow-switch.html [Accessed: June-13-2013]

[21][online] http://www.nsnam.org/wiki/index.php/Current_Development [Accessed: June-21-2013]

[22] [online] http://www.tfe.umu.se/english/research/embedded-systems-lab/projects/animal-sensor-networks/ [Accessed: July-26-2013]

[23] V. Shnayder, BR. Chen, K. Lorincz,Thaddeus R. F. Fulford-Jones, and M. Welsh, : Sensor Networks for Medical Care,Technical Report TR-08-05, Division of Engineering and Applied Sciences, Harvard University, 2005.

[24] B. Zarei, M. Zeynali and V.M. Nezhad, "Novel Cluster Based Routing Protocol in Wireless Sensor Networks". IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 1, July 2010.

[25] D. Ma, J. Ma, B. Huang, P. Xu: Energy-aware Clustering Protocol with Dual Cluster Heads using Niching Particle Swarm Optimization for Wireless Sensor Networks, International Journal of Advancements in Computing Technology(IJACT) Volume5,Number5,March 2013.

[26] Y. Rao, C. Yuan, Z. Jiang, L. Fu, J. Zhu: Latency and Reliability-aware Geographic Routing for Mobile Wireless Sensor Networks, Advances in information Sciences and Service Sciences(AISS), Volume5, Number8, April 2013.

[27] V. Kumar, S. Jain and S. Tiwari, "Energy Efficient Clustering Algorithms in Wireless Sensor Networks: A Survey", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011.

[28] O.S. Hyun, L.S. Hyuk: An Energy-Efficient Cluster-Head Selection Algorithm for Wireless Sensor Networks, International Journal of Digital Content Technology and its Applications (JDCTA), Volume7, Number2, January 2013.

The Cerebellum: New Computational Model that Reveals its Primary Function to Calculate Multibody Dynamics Conform to Lagrange-Euler Formulation

Lavdim Kurtaj[1], Ilir Limani[2], Vjosa Shatri[3] and Avni Skeja[4]

[1] **Department of Automation, Faculty of Electrical and Computer Engineering, University of Prishtina**
Prishtina, 10000, Kosovo

[2] **Department of Automation, Faculty of Electrical and Computer Engineering, University of Prishtina**
Prishtina, 10000, Kosovo

[3] **Department of Fundamental Engineering Subjects, Faculty of Electrical and Computer Engineering, University of Prishtina**
Prishtina, 10000, Kosovo

[4] **Department of Automation, Faculty of Electrical and Computer Engineering, University of Prishtina**
Prishtina, 10000, Kosovo

Abstract

Cerebellum is part of the brain that occupies only 10% of the brain volume, but it contains about 80% of total number of brain neurons. New cerebellar function model is developed that sets cerebellar circuits in context of multibody dynamics model computations, as important step in controlling balance and movement coordination, functions performed by two oldest parts of the cerebellum. Model gives new functional interpretation for granule cells-Golgi cell circuit, including distinct function for upper and lower Golgi cell dendritc trees, and resolves issue of sharing Granule cells between Purkinje cells. Sets new function for basket cells, and for stellate cells according to position in molecular layer. New model enables easily and direct integration of sensory information from vestibular system and cutaneous mechanoreceptors, for balance, movement and interaction with environments. Model gives explanation of Purkinje cells convergence on deep-cerebellar nuclei.

Keywords: *New Cerebellar function model, CMAC, Cerebellar elementary processing unit, Golgi cell, Basket cell, Stellate cell.*

1. Introduction

The cerebellum is part of vertebrate brain, and it is only part that spans to both sides continually without interruption. It plays important role in lower levels functions of motor control in relation to sensory information, such as coordination of complicated multi-joint bodies, balance, motor planning. According to connections with other parts of the brain, it is suggested for cerebellum to be involved in many other functional levels, up to cognition and emotional level. What makes debatable broad and diverse functional involvement of cerebellum is its uniform structure, in neuronal elements present, in local connectivity between them, and in circuits with other parts of the brain. This has lead authors of [1] to propose the general-purpose coprocessor function for the cerebellum. Based on connections between neuronal elements cerebellum can be viewed as a hierarchically organized system. Lowest level contains array of elementary processing units, each with group of neurons and single Purkinje cell, all arranged in highly ordered fashion [2]. Starting from functions of oldest parts of the cerebellum, and importance of robot dynamics model for high performance control algorithms, new cerebellar function model is developed that relates cerebellar circuits with multibody multijoint dynamics model computations.

2. Multibody System Dynamics

Meaning of term multibody system is very broad. In following, this term will imply to a set of bodies, called links, assembled together in a specific way with a set of movement constraining elements, called joints [3]. Interaction between bodies can happen only in joints. Joint can be only a concept that describes relative position between bodies, or specific construction where bodies are rigidly connected. Former case of joint infers that constraint motion is achieved with special construction of link endings that will be assembled together. In later case, all bodies that are rigidly connected together are treated as single link. Minimum number of parameters required to define relative movement of links connected with a joint denotes joint degrees of freedom. Sum of all joint degrees of freedom gives multibody system degrees of freedom. Each joint may have actuator for intentional altering of joint position (active joint), otherwise joint is passive and its state may change from external interactions acting on corresponding links. Links may be rigid or flexible. In general all links will express some degree of flexibility (elasticity), but for normal operating conditions deformation are bellow some maximal acceptable value. Theory of multibody system can be used to analyze all animals with skeletal structure (vertebrata), including humans, were links are represented by bones and joints are usually called articulations [4]. Muscles are used for joints actuation. A typical artificial multibody system, that sometimes serves as synonym for them, are robots [5], [6], sometimes with structure inspired by nature.

Robotics is interdisciplinary field and many disciplines contribute to different aspects of its development, naming some: mechanics, electronics, mechatronics, control engineering, computer science, biomechanics, etc. Main topics of robotics are kinematics, dynamics, trajectory planning, sensing, motion control, and intelligence.

2.1 Robot kinematics

Robot consists of links connected with joints. Robot kinematics deals with problem of finding analytical model that describes robot motion, with respect to fixed coordinate frame over time, without regard to forces and/or moments that cause the motion. Forward kinematics will give position $\boldsymbol{p}$ and orientation $\boldsymbol{\psi}$ of some frame attached to the robot, usually end-effector frame, as a function of joint variables $\boldsymbol{q}$

$$\begin{bmatrix} \boldsymbol{p} \\ \psi \end{bmatrix} = f_{FK}(\boldsymbol{q}). \qquad (1)$$

Opposite problem of finding joint variables $\boldsymbol{q}$ for given end-effector position $\boldsymbol{p}$ and orientation $\boldsymbol{\psi}$ is solved with inverse kinematics. Solution of forward kinematics is unique, whereas solution of inverse kinematics is not, necessitating sometime to add additional variables [7] and sometime to use numerical methods [8].

2.2 Robot dynamics

General form of dynamics equation of motion in matrix form (inverse dynamics) for robots with rigid links, derived following Lagrange-Euler formulation (L-E) is

$$\tau = D(q)\ddot{q} + C(q,\dot{q}) + G(q) + F_f(\dot{q}) + J^T(q) f_e, \qquad (2)$$

and gives torques/forces τ needed to drive joints for desired trajectory given by joint variables q, $\dot{q}$, and $\ddot{q}$, for joint position, velocity, and acceleration.

$D(q)$ denotes inertia matrix, $C(q,\dot{q})$ is the vector of Coriolis and centripetal terms, $G(q)$ is the vector of gravity terms. Friction is represented with $F_f(\dot{q})$ and any friction model may be used. Usually it includes dynamic friction and other unstructured friction effects. Possible interaction of end-effector with environment is expressed with last term, where f_e is vector of external forces and torques, and $J(q)$ is robot Jacobian. Eq. (2) is used for different forms of control strategies, e.g., feedforward, computed torque, feedback linearization [9], [10].
Forward dynamics equation can be derived from (2) by solving it for acceleration

$$\ddot{q} = D^{-1}(q)\left(\tau - C(q,\dot{q}) - G(q) - F_f(\dot{q}) - J^T(q) f_e\right), \qquad (3)$$

and is used for simulation of robot movement.

2.3 Processing units for robot dynamics computation

Advanced control strategies rely on dynamics model of the robot to achieve and maintain desired system dynamic response in accordance with specified performance criteria. To be able to track desired trajectories as close as possible, computation of appropriate torques/forces to drive joint actuators of the robot has to be done in real time. Delay caused by computer system to calculate control algorithm has negative influence on system performance. To shorten this delay when control computer system is based on single processor, computation is based on Newton-Euler formulation (N-E) for dynamics model, as a set of forward and backward recursive equations. Computational complexity of N-E is of order O(n), where n is the number of degrees of freedom (DOF) of the robot, as compared to $O(n^4)$ for L-E. Two main additional approaches are recursive Lagrangian formulation [11] with O(n) complexity, and generalized d'Alambert equations of motion [12] with complexity of order $O(n^3)$.

Classification of computational complexity is based in number of multiplications and additions for dynamics model calculation (n_{dm}). When calculations are done in single processor computers with given execution time per instruction (T_{ins}), total number of instructions n_{dm} will determine time delay introduced to control system by computation of dynamics model to $T_{sp}=n_{dm}\cdot T_{ins}$. For computer architectures with many processors, recursive computations will limit minimum delay proportional to the number of recursion steps (n_{rs}), i.e. $T_{mpr}=n_{rs}\cdot T_{ins}$, even under assumption that calculation of single recursion step is one "super" instruction and executes within T_{ins}. This can not be made smaller because of data dependencies between recursion steps. For N-E case there are n forward and n backward steps and $n_{rs}=2\cdot n$, where n is number of degrees of freedom for the robot.

If for some multibody system, with n=10 and with many very slow processors with $T_{ins}=100\mu s$, control system must update control values every $T_c=10ms$, question is: what the control computer architecture should be and which set of "super" instructions should it posses? Simple solution that pops up, when it is required to do complex, time consuming calculations with slow processors is to store precalculated values in a table. "Super" instruction would be the one that reads input data, calculates address from input data, reads value from the table, and writes result to the output. If input data are directly used as address, processor would be nothing but a memory of sufficient capacity. Calculation time with this "computer" system would be $1\cdot T_{ins}$. Despite its simplicity, this approach has as a drawback enormous size of memory for most applications. Input data space dimension for robot with n DOF is 3n (n joint positions, n velocities, and n accelerations). If each dimension of input data space is quantized with b_k values, memory size would be $b_1\cdot b_2 \cdots b_{3n}$, indicating exponential growth. If all b_k are equal memory size would be b_k^{3n}, and for $b_k=16$ (that is considered to be very coarse quantization) it would be 16^{30}, i.e. 120-bit address and 1024^{12} byte memory space, if single byte values are stored. Decreasing memory size of this unstructured approach is sought by exploiting structure of the problem to be solved.

Dynamics equation of motion for single joint is

$$\begin{aligned}
\tau_k &= d_{k1}(q)\ddot{q}_1 + d_{k2}(q)\ddot{q}_2 + \cdots + d_{kn}(q)\ddot{q}_n + \\
&\begin{bmatrix} h_{k11}(q)\dot{q}_1 + h_{k21}(q)\dot{q}_2 + \cdots + h_{kn1}(q)\dot{q}_n \\ h_{k12}(q)\dot{q}_1 + h_{k22}(q)\dot{q}_2 + \cdots + h_{kn2}(q)\dot{q}_n \\ \vdots \\ h_{k1n}(q)\dot{q}_1 + h_{k2n}(q)\dot{q}_2 + \cdots + h_{knn}(q)\dot{q}_n \end{bmatrix}^T \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \vdots \\ \dot{q}_n \end{bmatrix} + \\
&G_{k1}(q)g_{x_0} + G_{k2}(q)g_{y_0} + G_{k3}(q)g_{z_0} + \\
&F_{kd}\dot{q}_k + F_{ks}(\dot{q}_k) + \\
&J_{1k}(q)F_{ex_0} + J_{2k}(q)F_{ey_0} + J_{3k}(q)F_{ez_0} + \\
&J_{4k}(q)M_{ex_0} + J_{5k}(q)M_{ey_0} + J_{6k}(q)M_{ez_0}
\end{aligned} \tag{4}$$

where gravity term for joint k, $G_k(q)$, has been written in form so calculations with individual components of gravitational constant vector $\boldsymbol{g}$ are visible explicitly. This form has two advantages; first, it creates gravity terms that are similar in form with most of others, and second, it opens possibility to compensate directly when multibody reference changes orientation relative to gravitational field, that is present almost all the time for example in humans and mobile robots.

To calculate dynamics model for single joint, as seen from Eq. (4), we need only a small set of prototype functions. Main function prototype of form

$$f_p(q,x) = N(q)x\,, \tag{5}$$

will suffice for calculation of inertial torques, $d_{kj}(q)\ddot{q}_j$, gravitational torques, $G_{kj}(q)g_j$, and torques exerted from interaction with environment, $J_{jk}(q)F_{ej}$ and $J_{jk}(q)M_{ej}$. Calculation of Coriolis and centrifugal torques $h_{kij}(q)\dot{q}_i\dot{q}_j$ can be done by using first prototype function

$$\tau_{Cc} = f_p(q,x)y\,, \tag{6}$$

or they can be calculated by defining second prototype function

$$f_{p2}(q,x,y) = N(q)xy\,. \tag{7}$$

Calculating previous terms of dynamics equation with second prototype can be done by giving constant value equal to 1 for second variable, y=1, or by providing some controlled switching mechanism for variable y. Calculation of dynamic and static friction torques, $F_{kd}\dot{q}_k$ and $F_{ks}(\dot{q}_k)$, is possible with either prototype function, but, having in mind their simplicity compared to other terms, it seams like waste of computational resources. This could be solved with some add-on to selected prototype function. Extended second prototype function would become

$$f_{p2s+}(q,x,s,y,a,b) = N(q)xy(s) + \mu_d a + \mu_s(b)\,, \tag{8}$$

were we have included logical variable s as switch for variable y, with $y(s)$=1 for one value of s and $y(s)=y$ for the other. Two last terms of Eq. (8) can be used to calculate friction torques of Eq. (4).

Computer architecture for calculation of dynamics model that exploits structure of the model would have first layer with one processing unit (PU) for each term in Eq. (4) for each DOF, excluding friction terms, with single "super" instruction for doing calculation according to Eq. (8) within T_{ins} time, and second layer with one PU for each DOF for summing corresponding results together, also within T_{ins} time. Total calculation time would be $2\cdot T_{ins}$, independent of DOF for the system. First layer of architecture resembles Single Instruction Multiple Data

(SIMD) class of parallel computer architectures according to Flynn's taxonomy [13], [14]. Number of PU in first layer for n=10 is n·(n+n·n+3+3+6)=1220, and second layer has 10 PU. With new structure, dimension of input space for nonlinear function is n (as opposed to 3n for unstructured approach), and if value is taken from table, table memory size for single PU would be 16^{10}, i.e. 40-bit memory address and 1024^4 byte (at least, now we can name it: 1 Tera byte) memory space. Despite enormous reduction, this value is still high. Different approaches are followed to reduce this value, and approximation of nonlinear functions with neural networks is one of them.

3. Cerebellum and CMAC

There is a small number of neuron types present in cerebellum, common to all vertebrates, arranged in highly ordered fashion. Among them are largest neuron in the brain, (giant) Purkinje cell, and most numerous type of neuron, small granule cell [2]. They are organized in three layers. Middle layer, so called Purkinje cell layer, contains only Purkinje cells packed in one-cell thick layer. Inner layer (below Purkinje layer), granule cell layer, inhabits granule cells, Golgi cells, and some other interneurons (some only at specific regions) [15]. Outer layer of cerebellar cortex is named molecular layer. It consists of granule cells axons with characteristic T-shaped form crossing at right angle with plane of flat-shaped Purkinje cells dendrite trees. Present in this layer are also basket and stellate cells. Granule cells are the only excitatory cells in the cerebellum, all the others are inhibitory. Axons of Purkinje cell are sole outputs of the cerebellar cortex, whereas inputs are through group of mossy fibers and through climbing fibers. There are lot of evidences about neuronal connections inside cerebellum, and higher functional organizations, into microzones and even higher to zones, when considering connections with other parts of the brain [16]. Simple sketch of connections between cerebellar neurons is given in Fig. 1. There are two main feedforward routes from input to output. In first main route afferent mossy fibers project to granule cells, granule cells project to Purkinje cell, Purkinje cell sends its efferent axon out of the cerebellar cortex. Second main route is from climbing fiber to Purkinje cell, and again from Purkinje cell out of the cerebellar cortex. Connection climbing fiber-Purkinje cell is rather specific in that each Purkinje cell receives strongest synaptic contact in the brain, and that exclusively from single climbing fiber. Additional routs are through basket and stellate cells inhibitory interneurons. Golgi cell forms local feedback loop with granule cells, and an additional feedforward route from mossy fibers to Purkinje cell output fiber. Connectivity like this is present in all vertebrates, almost invariant, with difference only in quantities [17]. There are also additional similar structures in vertebrates and invertebrates [18]. Other difference is anatomical and has to do with appearance of deep folds that resulted with increase of cerebellar cortex, with purpose of more compact packing. If unfolded human cerebellum would have approximate dimensions of 1 m long and 50 mm wide, with as high as $15x10^6$ Purkinje cells [19]. Number of Purkinje cells in rat is $3.38x10^5$ [20] and in frog around 8300 [21].

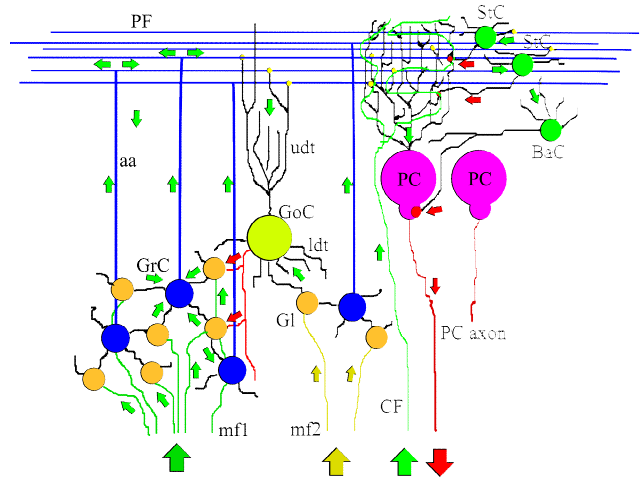


Fig. 1 Cerebellar neurons, connections, and signal routes. **PC**: Purkinje cell; **GC**: Granule cell; **GoC**: Golgi cell; **StC**: Stellate cell; **BaC**: Basket cell; **Gl**: glomeruli; **aa**: granule cell ascending axon; **PF**: parallel fibers; **adt**: Golgi cell ascending dendritic tree; **ddt**: Golgi cell descending dendritic tree; **mf1**, **mf2**: mossy fibers; **CF**: climbing fiber; **PC axon**: Purkinje cell axon.

Cerebellum is collection of numerous elementary processing units (ePU), thousands to tens of millions, which may, more or less, share processed input data, and influence other ePU with limited number of collaterals from output fiber. To understand function of the cerebellum, as a first step would be to understand function of single ePU by neglecting possible mutual couplings, as they would belong to higher level organization of several ePU. This approach is followed by renowned theories of Marr [22] and Albus [23], and formalized later by Albus with Cerebellar Model Articulation Controller (CMAC) [24] as type of artificial neural network (ANN) that can be used as robotic controller. Model retains two main feedforward routs, with attribution of learning mechanism to climbing fiber-Purkinje cell route. Purkinje cell functions as a perceptron with adjustable weights that model synapses between this cell and granule cell axons (parallel fibers), as only site were learning takes place mediated by climbing fibers assumed to carry error signal. Model does not include explicitly molecular layer inhibitory neurons (basket and stellate cells), but their

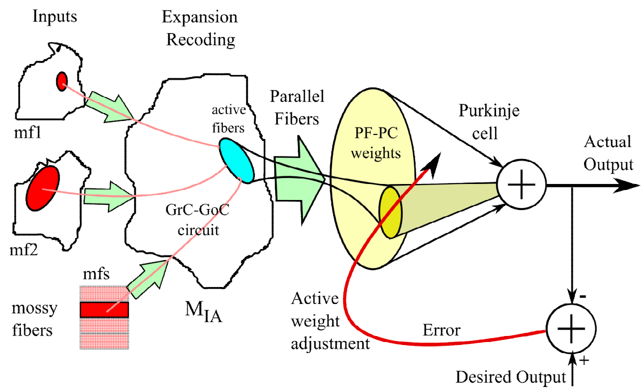


Fig. 2 CMAC structure.

functionality according to theory is included in possibility of having negative weights. Fig. (2) shows basic CMAC structure. Granule cell layer processing is represented with mapping M_{IA} from input space (mossy fibers) to association space (inputs of the perceptron). This type of ANN is characterized with fast learning, attributed to local generalization property, which in turn is dependent on mapping M_{IA}. By selecting proper mappings CMAC can turn to simple perceptron that wont be able to solve even XOR problem, to storage table, to network with properties equivalent to radial basis functions (RBF) ANNs, it can turn to be equivalent with fuzzy ANN. or to some quite different ANN. In case of multidimensional input spaces, standard operation used to process receptive fields is multiplication operation [25] [26] (in standard CMAC with rectangular shaped receptive fields AND operation is used [27]). With this broad range of possibilities, success of the network (perceptron) is highly dependent on proper selection of mapping for selected problem. So, network can learn, but quality and quantity of learning relies on opportunities that are given for the network to learn from associations presented by mapping M_{IA}. Since this mapping represents processing in the granular layer, turns attention to process of creating multidimensional receptive fields, and if multiplication really is present in biological neurons [28], [29] [30]. If it is present, is it implemented by a single neuron or by a group of neurons [31], [32]? In other side, coordination of multi-joint systems may have great benefit from it, as dimension of input space can be greatly reduced, because of multiplicative nature of speed and acceleration in dynamics model, Eq. (4). Does multiplication between two different inputs happen in cerebellum? Where possible sites are where multiplication in the cerebellum could take place, and how could they be related to dynamics model computation, as important issue in controlling balance (assisted by evolutionary oldest part of the cerebellum, archicerebellum), and movement coordination (second oldest part of the cerebellum, spinocerebellum or known also as paleocerebellum)?

4. New Computational Model of Cerebellar Function

Multi-joint articulated robots dynamics is characterized with many interactions between joints. High quality control requires carefully designed controller, and robot dynamics model is usually part of this controller [33], [34]. With exception of simple cases, finding this model for more complicated structures is a challenge. Having in mind that oldest parts of the cerebellum are related to the problems that in essence contain dynamic interactions, better understanding of cerebellar function will help us to solve robotic problems of ever increasing complexity, which still in number of sensors and actuators is far behind from humans and animals that serve as inspiration [35]. This is new anatomy-based model [36] that gives new functional description for cerebellar circuits, revealing relation to dynamics calculation. Model includes:

- new functional interpretation for granule cell-Golgi cell circuit, and distinction between upper and lower dendrite trees,
- explanation of scenarios for sharing Golgi cells between Purkinje cells,
- new function for basket cells,
- new function for stellate cells, with distinct feature for upper and middle ones,
- definition of coding for different signals at parallel fibers and mossy fibers for consistency with new model,
- direct integration of information from vestibular system,
- direct integration of multilevel interaction with environment,
- explanation of convergence on deep-cerebellar nuclei.

4.1 Cerebellar elementary processing unit

Cerebellum will be treated as computer architecture with several levels of organization. Lowest level contains thousands to millions of cerebellar elementary processing units (CePU), with only one output form axon of single Purkinje cell (PC). Functional diagram of CePU is shown in Fig. 3. Most numerous inputs to the PC come from T-shaped axons of granule cells (GrC), know as parallel fibers (PF), and are excitatory by nature. Other inputs to PC are from molecular layer inhibitory interneurons, stellate cells (StC) and Basket cells (BaC). StCs provide input through PC dendrite tree, whereas group of BaCs supply inhibitory synapses to the basket-like structure at initial segment of PC soma. PF provide inputs to the StC and BaC, and serve as data path that is shared between many PCs. Each PC will receive data from as many as 200000 (in humans) from about 10 million PF that will pass by.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

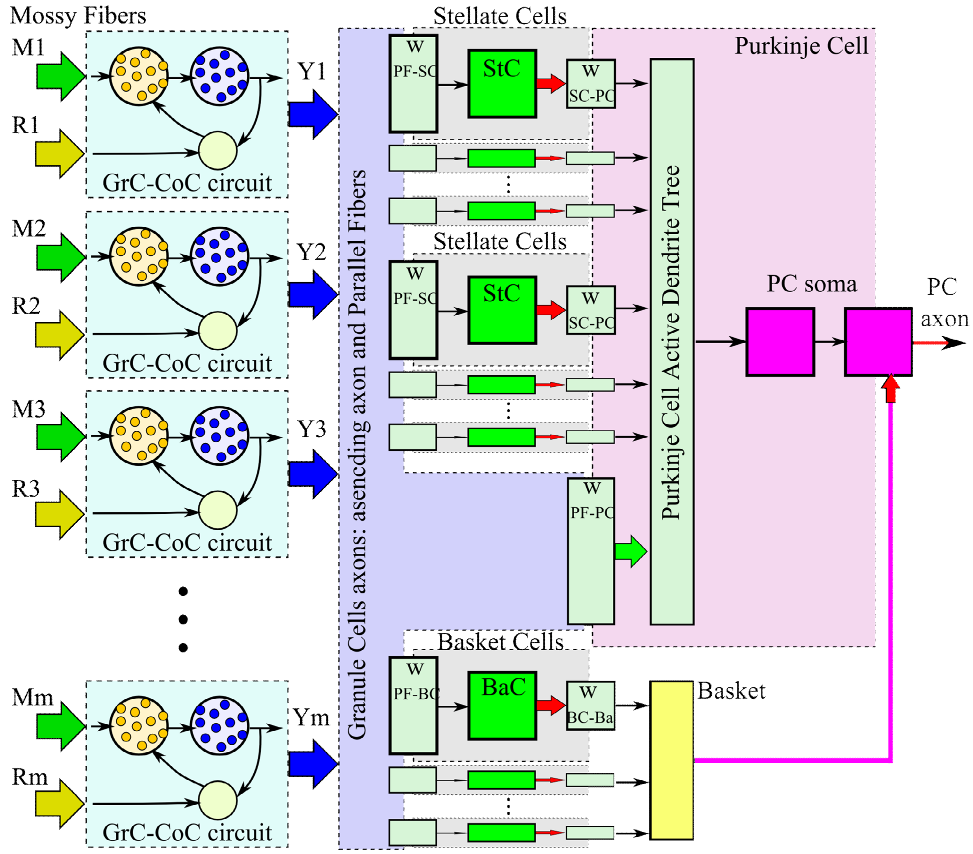


Fig. 3 Functional diagram of **C**erebellar **e**lementary **P**rocessing **U**nit, CePU.

Data at PF is result of preprocessing done by GrCs and Golgi Cells (GC) over cerebellar input data from mossy fibers (MF).

4.2 Granule cells-Golgi cell circuit

Granule cell receive inputs from a number (usually 4-5) of distinct MF and generates output to a single PF. A group of GrC will map information from m_g-dimensional MF input space to a sparse p_g-dimensional PF output space. Index g is used to indicate existence of many different groups. Marr [22] referred to output of GrC as codon representation of an input, while Albus [23] named mapping process as expansion recoding and output as association of input. In concepts of ANN output would correspond to some sort of multidimensional basis function. Sparsity is considered important feature of new representation and will influence learning speed and capacity.

In an ideal situation, when mossy fibers provide information in form of normalized basis functions, and GrC would function as multiplier of its inputs, output would be normalized higher dimensional basis function. This ideal approach is usually followed when using ANN. GoC are used to deal with real situations and keep mean output activity on controlled level. GoC influences output over two control mechanisms, feedback and feedforward, with former acting through upper part off dendritic tree residing on molecular layer, and synapsing with GrC axons (PF and ascending axon [37]). Feedforward uses lower dendritic tree and synapses with MF. According to Marr [22] output activity is controlled by influencing GrC threshold by that tree that is more powerfully stimulated [22]. Albus suggested that automatic gain control [23] is used to maintain constant output activity independent of input activity. Distinction between two trees is on speed of acting, with feedforward path being faster acting.

Our opinion is that part of cerebellum for handling dynamic interactions (balance control, movement coordination, and joint decoupling) between joints of the multijoint system works in mode of rate coded version of continuous control system. Functional relation between inputs and outputs will be analyzed from functional diagram of GrC-GoC circuit, Fig. 4 (see also Fig. 7 from [23]). GrC receive inputs L_{jk} from structures called mossy fiber rosettes (cerebellar glomeruli); place where GoC

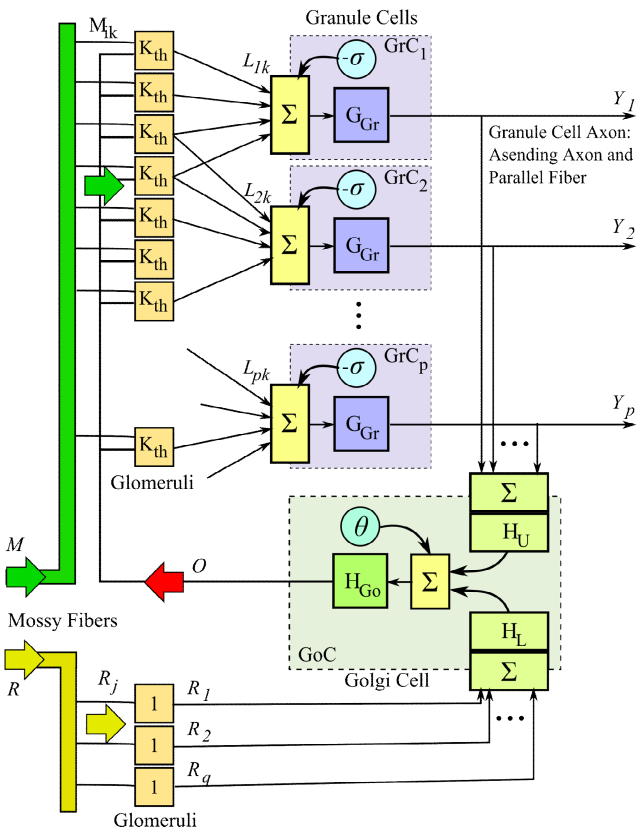


Fig. 4 Granule cells-Golgi cell functional diagram for preprocessing input data from mossy fibers.

axon terminates too and influences mossy fiber signals M. Output from single GrC will be

$$Y_i^+ = \left(\sum_{k=1}^{4} M_{ik} - \sigma_i - K_{th} O \right) \cdot G_{Gr} \tag{9}$$

if control is done by changing GC threshold, and it will be

$$Y_i^+ = \left(g(O) \cdot \sum_{k=1}^{4} M_{ik} - \sigma_i \right) \cdot G_{Gr} \tag{10}$$

with

$$g(O) = 1 - K_g O \tag{11}$$

if control is done through automatic gain control. K_{th} and K_g are glomeruli specific constants. Function of GrC is assumed simple summation of four incoming signals from dendrites followed by G_{Gr} part, which may include neuron dynamic and nonlinear static characteristics (except firing threshold σ_i already included). Plus sign over Y indicates that value can be only positive or zero, i.e. neurons by nature are excitatory or inhibitory and their functionality can not be reversed. GoC output is

$$O^+ = \left(H_U \cdot \sum_{i=1}^{p} Y_i + H_L \cdot \sum_{j=1}^{q} R_j + \theta \right) \cdot H_{Go} \cdot \tag{12}$$

In contrast to [22] it is assumed that action potentials from both dendritic trees are simply summed. Possible behavioral differences are taken into account with functions H_U and H_L. p and q denote number of synapses (assumed fixed) made by upper and lower dendritic trees, respectively. Spontaneous firing is represented with θ, while dynamic and static (without spontaneous firing) characteristics of GoC are represented with G_{Go}. Plus sign over O, again, has same meaning, whereas minus sign preceding it will indicate inhibitory nature of GoC. Static characteristics of GrC and GoC are given in [38], where it can be seen that GoC is characterized with spontaneous firing activity, while GrC will start firing after crossing some threshold, modeled with σ_i and θ, respectively. From Eq. (9), (11), (12) and after summing over m active outputs Y_i we can find

$$\sum_{i=1}^{m} Y_i^+ = \frac{G_{Gr} \cdot \sum_{i=1}^{m} \left(\sum_{k=1}^{4} M_{ik} - \sigma_i - K_{th} H_{Go} \theta \right)}{1 + G_{Gr} \cdot K_g H_U H_{Go} \cdot m} - \frac{G_{Gr} \cdot K_g H_U H_{Go} \cdot m \cdot \frac{H_L}{H_U} \sum_{j=1}^{q} R_j}{1 + G_{Gr} \cdot K_g H_U H_{Go} \cdot m}. \tag{12}$$

Aim is to keep m in a small fraction of p, around 1%. Independence of output activity from input activity M_{ik} cannot be achieved, contrary to conclusion of Albus [23], because only positive term in Eq. (12) is function of inputs, and by removing that (ex. by making closed loop gain $G_{Gr} \cdot K_g H_U H_{Go} \cdot m$ very large) no output will be active. But despite that, output activity can be kept constant if inputs M_{ik} are in form of basis functions, which would result in constant first term of Eq. (12). Sparsity then is controlled with widths of basis functions. All this could have been done with GrC only, without complicating things with cerebellar glomeruli and GoC. Our conclusion is that usefulness of this structure lies on interpretation of second term of Eq. (12). Since, except input information, all parameters are related to constructive parameters, this equation can be written as

$$Y = k_1 M - k_2 - k_3 R. \tag{13}$$

For given M it is equation of a line with negative slope. This slope will effectively turn to positive one at final action: inhibiting inhibitor neuron. Our opinion is that main functionality behind preprocessing with GoC and cerebellar glomeruli is realization of linear dependence of group of outputs Y_i from an additional input R, i.e. simple multiplication operation, and preparation phase for approximating function prototypes given by Eq. (5)-(8). Consequence of this functional interpretation is that coding to be used for information on R_j inputs is rate coding with no GoC axon termination on corresponding glomeruli. Outputs Y_i will provide higher dimensional basis functions in M space modulated in amplitude with additional variable R. Same outputs can be used between many PC for final function approximation, ex. when calculating dynamics model terms of form $d_{km}(q)\ddot{q}_m$, same set of Y_i outputs (PF) can be used for all joints, where M would correspond to joint positions vector q, and R to acceleration $\ddot{q}_m$. This removes the doubt put by Marr [22] about sharing GoC between PC. One prediction can be made from this interpretation: number of PC that will contact this set of PF should be related to dimension of joint space where dynamic interactions may occur. To calculate whole dynamics equation, for each term in Eq. (4) at least one GoC will be needed. Recoding higher dimensional spaces can be done with several GoC over corresponding subset of dimensions, and number of GrC dendrites could match the dimensionality of subspace. These subsets may have realistic meaning, like shoulder position, head position, etc.

4.3 Basket cells and Stellate cells

Role of BaC and StC according to Marr [22] was in setting up the PC threshold, with distinguishing role of most superficial StC as on preventing false initial response by the PC. Albus in his theory [23] assigned them function of providing negative weights for PC that serves a function as a perceptron, a feature that would enable PC high learning capabilities. Many models and theories of cerebellar function do not include them at all [39], or make no difference between them and refer them together as basket/stellate cells [39]. [40] concluded that StC and BaC are similar types of interneurons, with functional difference to synaptic strength, with that of BaC (somatic) being 7-fold higher than that of StC (dendritic). Different spatial, temporal, and physiological effects and consequences for stellate (dendritic), and basket (somatic)-type inhibition was attributed by [39].

When cerebellum is set in context of balance and movement coordination, we hypothesize different function for each of them in duty of computing dynamic interactions. Basket cells are part of the circuit for calculating Coriolis and centrifugal terms of dynamics equation. Stellate cells will mainly handle calculations of friction terms. A set of GrC with one GoC and one PC will be able to approximate calculation of product between one multidimensional nonlinear function and one linearly dependent variable, Eq. (5). This set can be used to calculate terms of inertial torques of Eq. (4)

$$\begin{aligned} d_{km}(q)\ddot{q}_m &= f_p^k(q,\ddot{q}_m) = N^k(q)\ddot{q}_m \\ &= \sum_{s=1}^{p} Y_s^{k+} \cdot w_{PCs}^{km} \\ &= \sum_{s=1}^{p} B_s(q)\ddot{q}_m \cdot w_{PCs}^{km} \end{aligned} \tag{14}$$

To calculate Coriolis and centripetal terms an additional multiplication is needed, Eq. (6). One BaC with its axon synapsing on PC soma will approximate this, under assumption that this interaction performs multiplication between these two signals, and ij^{th} term for joint k can be written as

$$\begin{aligned} h_{kij}(q)\dot{q}_i \cdot \dot{q}_j &= f_p^k(q,\dot{q}_i) \cdot \dot{q}_j \\ &= \left(\sum_{s=1}^{p} Y_s^{k+} \cdot w_{PCs}^{kij}\right) \cdot \dot{q}_j \\ &= \left(\sum_{s=s_q}^{s_q+m} B_s(q)\dot{q}_i \cdot w_{PCs}^{kij}\right) \cdot \dot{q}_j \end{aligned} \tag{15}$$

with $B_s(q)$ being m+1 nonzero bases functions for current joint vector q, and $w^{kij}{}_{PCs}$ are weights of PF-PC synapses. This approach would need n^2 PC and BaC for single joint. Additionally, PF with joint speed data $\dot{q}$ is needed. To find more suitable solution we will start from a group of Coriolis and centrifugal terms

$$h_{ki} = h_{ki1}(q)\dot{q}_i\dot{q}_1 + h_{ki2}(q)\dot{q}_i\dot{q}_2 + \cdots + h_{kin}(q)\dot{q}_i\dot{q}_n. \tag{16}$$

By applying Eq. (15) for given position q_0 and square basis functions of width 1, Eq. (16) takes this form

$$\begin{aligned} h_{ki} &= B_s(q_0)\dot{q}_i w_{PCs}^{ki1} \cdot \dot{q}_1 + B_s(q_0)\dot{q}_i w_{PCs}^{ki2} \cdot \dot{q}_2 + \cdots + B_s(q_0)\dot{q}_i w_{PCs}^{kin} \cdot \dot{q}_n \\ &= B_s(q_0)\dot{q}_i w_{PCs}^{kii}\left(w_{Xs}^{ki1} \cdot \dot{q}_1 + w_{Xs}^{ki2} \cdot \dot{q}_2 + \cdots + w_{Xs}^{kin} \cdot \dot{q}_n\right) \\ &= B_s(q_0)\dot{q}_i w_{PCs}^{kii}\left(B_s(q_0)\dot{q}_1 w_{BCs}^{ki1} + B_s(q_0)\dot{q}_2 w_{BCs}^{ki2} + \cdots + B_s(q_0)\dot{q}_n w_{BCs}^{kin}\right) \end{aligned} \tag{17}$$

Last form of Eq. (17) can be interpreted as sum of n signals from BaC controls output from one PC. Weights $w^{kij}{}_{BCs}$ represent synaptic connections PF-BaC. This form as input for BaC uses already present outputs from GrC. Now we need only n PC for single joint, but n^2 BaC are needed. To simplify further, BaC that uses basis functions modulated with j^{th} joint speed will be used for all PC that need it, resulting with total Coriolis and centrifugal term for joint k as

$$\begin{aligned} h_k = {} & B_s(q_0)\dot{q}_1 w_{PCs}^{k1}\left(B_s(q_0)\dot{q}_1 w_{BCs}^{k1} + B_s(q_0)\dot{q}_2 w_{BCs}^{k2} + \cdots + B_s(q_0)\dot{q}_n w_{BCs}^{kn}\right) + \\ & B_s(q_0)\dot{q}_2 w_{PCs}^{k2}\left(B_s(q_0)\dot{q}_1 w_{BCs}^{k1} + B_s(q_0)\dot{q}_2 w_{BCs}^{k2} + \cdots + B_s(q_0)\dot{q}_n w_{BCs}^{kn}\right) + \\ & \vdots \\ & B_s(q_0)\dot{q}_k w_{PCs}^{kk}\left(\cdots + B_s(q_0)\dot{q}_{k-1} w_{BCs}^{k(k-1)} + 0 + B_s(q_0)\dot{q}_{k+1} w_{BCs}^{k(k+1)} + \cdots\right) + \\ & \vdots \\ & B_s(q_0)\dot{q}_n w_{PCs}^{kn}\left(B_s(q_0)\dot{q}_1 w_{BCs}^{k1} + B_s(q_0)\dot{q}_2 w_{BCs}^{k2} + \cdots + B_s(q_0)\dot{q}_n w_{BCs}^{kn}\right) \end{aligned} \tag{18}$$

Value zero at kk^{th} term indicates anatomical observations that BaC do not inhibit the PC immediately adjacent [23], that is consistent also with physical interpretation that h_{kkk}=0 [7]. Interpretation is valid too for wider basis functions, which will justify much less dense dendritic tree of BaC. Justification is based on the fact that all outputs from a group of GrC modulated with given joint speed carry joint speed information, and it will suffice to sample it sparsely, ex. 2 per basis width m, i.e. if m=20 number of inputs to BaC would be p/10, meaning ten times less than inputs to GoC, and not necessarily related to number of PC synapses. Even further simplification can be in using fixed PF-BaC synapses. Consequence of this would be that BaC make synapses with ascending axon of corresponding GrC outputs, which are essentially hard wired [41] connections, but there are disagreements about functionality between GrC synapses on ascending axon and PF fiber [42], [43]. Additional possibly adjustable places are BaC-PC synapses, which depending on the configuration may be redundant to GrC-BaC synapses, or the only one adjustable. Other extreme, in contrast to mammals and birds, is total lack of BaC, like in most fish, amphibians, and most reptiles [39]. This could be interpreted as lack of given type of dynamic interactions, insignificant contribution, or different mechanism of approximation.

There are two friction terms in dynamics equation of motion, dynamic and static. Calculation of dynamic friction is readily achieved with StC that would receive inputs from sparsely sampled joint speed modulated basis functions, like in BaC, where output of the StC would correspond to recovered joint speed information. StC synapse with PC dendrite would represent dynamic friction coefficient F_d

$$\tau_{kfd} = F_{kd}\dot{q}_k = \left(\sum_{s=rnd_1(p)}^{rnd_d(p)} B_s(q)\dot{q}_k \cdot w_{SCs}^k\right) \cdot w_{SP}^k . \tag{19}$$

Weights w^k_{SCs} are for PF-StC synapses that may be fixed one, whereas adjustable weight is the one between StC and PC w^k_{SP}. Friction effect is local and in ideal case there is one friction term per joint. Since in living organisms many joint are with many DOF driven by many paralleled actuators (muscles), each actuator that drives given joint will need one StC. Those StC at most superficial regions of molecular layer would probably be the ones related to joint with 1-DOF, and deeper one for joints with more DOF, where resultant speed would govern dynamic friction torques.

Calculation of static friction term would need GrC with space encoded joint speed. This task can be accomplished with smaller GoC and small number of GrC. If outputs of these GrC space coded joint speed information, static friction can be calculated as

$$\tau_{kfs} = F_{ks}(\dot{q}_k) = \left(\sum_{s=1}^{v} B_s(\dot{q}_k) \cdot w_{SCs}^k\right) \cdot w_{SP}^k . \tag{20}$$

In this case weights w^k_{SCs} at PF-StC synapses are adjustable, whereas between StC and PC w^k_{SP} is redundant, or can be used as global adjustment parameter.

4.4 Functional integration of sensory information

All types of information generated from different that are related to balance, movement, and interaction with environment are easily integrated with this new functional model of the cerebellum. Processed information from vestibular systems (saccule and utricle), combined with information from joints, will result with orientation of body relative to gravitational field. This information can be used in computation of gravity terms of dynamics model in a similar way as was done with inertial torques, Eq. (4) and Eq. (14), with only difference that instead of joint accelerations modulatory functions would be gravity projections.

In a completely same way, and same set of computation prototypes (a set of GrC with one GoC and one PC) can be used to integrate cutaneous mechanoreceptors information for movement and interaction with environment. External forces, torques, and places of action will be found by processing sensor information. Their dynamic effects will be computed with expressions similar to Eq. (14) with modulatory function being six components of interaction force/torque for each segment of multijoint system. Last six terms of Eq. (4) represent this group of interactions, but it shows only one set from many possible, usually robot end-effector force/torque interaction with environment.

4.5 Higher levels of organization

Second level of organization between CePU is done with PC collaterals [23], [44]. Third level is constituted from microzones [1], [45], each with a group of around 1000 PC arranged in a narrow longitudinal strip that crosses PF into right angle. PC axons terminate to a group less than 50 deep cerebellar nuclei (DCN) [46]. According to new model, microzone would correspond to computational unit that computes total joint torques, Eq. (4), by collecting together all results from individual CePU. This model can help to resolve dilemma about reason that hundreds of PCs converge to single DCN [47]. Climbing fibers (CF) that innervate PCs of microzone (each CF about 10 PC) come from a group of olivary neurons that tend to be coupled by gap junctions. Also axones of BaC are much longer in longitudinal direction, and stay mainly inside a microzone. It is thought that microzones represent effective cerebellar functional units, and even higher level organization represents functional modules (stripes, zones, and multizonal microcomplexes) [1]. These organizations are marked also with molecular markers, as an additional fact beside anatomical and physiological facts [16].

5. Conclusions

New computation model of cerebellar function tries to relate cerebellar neuronal circuits with problem of multijoint dynamics model computations. Multiplicative inclusion of joint speeds, accelerations, gravitational acceleration, and forces/torques of interaction with environment, will make great reduction of dimensionality of problem space to be learned and extend generalization over wide range of multiplicative variables. Model gives functional explanation of connections between main neuron types. To be consistent with model, position information should be space-coded with bases functions of selected width. Same data is additionally amplitude modulated (rate code) with an additional information source. Position information on mossy fibers should be in form of basis functions, whereas other variables that are used multiplicatively, such as joint speed, joint acceleration, etc., use simple rate code. Compared to traditional CMAC, new model may serve as guideline for preprocessing and possibilities of inclusion of pure multiplication in model, that is biologically plausible, for solving problems in robotics with higher level of generalization.

Issues to be resolved in context of new model are functions of Purkinje cell collaterals, and for collaterals of climbing fibers. Additional information for seeking more probable explanation is branching of climbing fiber to 10 Purkinje cells: is it related to redundancy, or 10-fold increase in resolution, or maybe load sharing between paralleled actuators (muscle fibers), or group is related to timing issues like interpolation.

References

[1] Egidio D'Angel and Stefano Casali, "Seeking a unified framework for cerebellar function and dysfunction: from circuit operations to cognition", Frontiers in Neural Circuits, Januray 2013, Vol. 6, Article 116.
[2] Dan H. Sanes, Thomas A. Reh, William A. Harris, Development of the Nervous System, Second Edition, Elsevier Academic Pess, 2006.
[3] Abhinandan Jain, Robot and Multibody Dynamics: Analysis and Algorithms, Springer, 2012
[4] Aydın Tözeren, Human Body Dynamics: Classical Mechanics and Human Movement, Springer-Verlag, 2000.
[5] Bruno Siciliano, Lorenzo Sciavicco, Luigi Villani, Giuseppe Oriolo, Robotics: Modelling, Planning and Control, Springer, 2011
[6] Siciliano, Bruno; Khatib, Oussama (Eds.), Springer Handbook of Robotics, Springer, 2008.
[7] Fu, Gonzalez, Lee, Robotics: Control, Sensing, Vision, and Intelligence, McGraw-Hill, Inc., 1987.
[8] Peter Corke, Robotics, Vision and Control: Fundamental Algorithms in MATLAB, Springer-Verlag, 2011.
[9] Zhihua Qu, Darren M. Dawson, Robust Tracking Control of Robot Manipulators, IEEE Press, 1996.
[10] Rafael Kelly, Victor Santibáñez Davila, Julio Antonio Loría Perez, Control of Robot Manipulators in Joint Space, Springer, 2005.
[11] John M. Hollerbach, "A Recursive Formulation of Manipulator Dynamics", Massachusettes Institute of Technology, Artificial Intelligence Laboratory, A.I. Memo No. 533, June, 1979, Revised June, 1980.
[12] C. S. G. Lee, B. H. Lee, R. Nigam, "An Efficient Formulation of Robot Arm Dynamics for Control Analysis and Manipulator Design", The University of Michigan, Center for Robotics and Integrated Manufacturing, April 1982.
[13] Michael J. Flynn, Computer Architecture: Pipelined and Parallel Processor Design, Jones and Bartlett Publishers, Inc., 1995.
[14] Flynn's taxonomy, (last accessed 2013.08.30), http://en.wikipedia.org/wiki/Flynn's_taxonomy.
[15] Diño MR, Schuerger RJ, Liu Y, Slater NT, Mugnaini E., "Unipolar brush cell: a potential feedforward excitatory interneuron of the cerebellum", Neuroscience, 2000, 98(4): 625-36.
[16] Karl Herrup and Barbara Kuemerle, "The Compartmentalization of the Cerebellum", Annu. Rev. Neurosci. 1997.
[17] Ann B. Butler, William Hodos, Comparative Vertebrate Neuroanatomy: Evolution and Adaptation, Second Edition, John Wiley & Sons, Inc., 2005.
[18] Curtis C. Bell, "Evolution of Cerebellum-Like Structures, Brain, Behavior and Evolution", 2002, 59:312–326.
[19] Gordon M. Shepherd (Ed.), The Synaptic Organization of the Brain, Fifth Edition, Oxford University Press, Inc., 2004.
[20] R. J. Harvey, Dr. R. M. A. Napper, "Quantitative study of granule and Purkinje cells in the cerebellar cortex of the rat", Journal of Comparative Neurology, 8 August 1988, Volume 274, Issue 2, pages 151–157.
[21] A. Pellionisz, R. Llinas, "A Computer Model of the Crebellar Cortex of the Frog", Neuroscience, 1977, Vol. 2, pp. 19-35, Pergamon Press.
[22] David Marr, "A Theory of Cerebellar Cortex", J. Physiol., 1969, 202, pp. 437-470.

[23] James S. Albus, "Theory of Cerebellar Function", Mathematical Biosciences, February 1971, Volume 10, Numbers 1/2, pgs. 25-61.

[24] James S. Albus, "New Approach to Manipulator Control: The Cerebellar Model Articulation Controller (CMAC)", Transactions of the ASME Journal of Dynamic Systems, Measurement, and Control, September 1975, pgs. 220 - 227.

[25] Stephen H. Lane, David A. Handelman, and Jack J. Gelfand, "Theory and Development of Higher-Order CMAC Neural Networks", IEEE Control Systems, April 1992.

[26] Wen Yu, Floiberto Ortiz Rodriguez, and Marco A. Moreno-Armendariz, "Hierarchical Fuzzy CMAC for Nonlinear System Modeling", IEEE Transaction on Fuzzy Systems, October, 2008, Vol. 16, No. 5.

[27] W. Thomas Miller, III, Filson H. Glanz, L. Gordon Kraft, III, "CMAC: An Associative Neural Network Alternative to Backpropagation", Proceedings of the IEEE , Oct. 1990, Vol 78, No. 10, pp. 1561-1567.

[28] José Luis Peňa and Masakazu Konishi, "Robustnes of Multiplicative Processes in Auditory Spatial Tuning", The Journal of Neuroscience, October 6, 2004 - 24(40) 8907 - 8910.

[29] Fabrizio Gabbiani, Holger G. Krapp, Nicholas Hatsopoulos, Chun-Hui Mo, Christof Koch, Gilles Laurent, "Multiplication and stimulus invariance in a looming-sensitive neuron", Journal of Physiology - Paris 98 (2004) 19-34.

[30] Panagiotis Nezis and Mark C W van Rossum, "Accurate multiplication with noisy spiking neurons", Journal of Neural Engineering, 2011, 034005 (5pp).

[31] Christof Koch and Idan Segev, "The role of single neurons in information processing, nature neuroscience supplement", November 2000, Vol. 3.

[32] Koch, C. and Poggio, T. "Multiplying with synapses and neurons". In: McKenna, T., Davis J.,and Zornetzer F.S. (Eds.) Single Neuron Computation, 1992, p. 315-345.

[33] Zhang Wenhui, and Zhu Yinfa, "Control of Free-floating Space Robotic Manipulators base on Neural Network", IJCSI International Journal of Computer Science Issues, November 2012, Vol. 9, Issue 6, No 2.

[34] Yongqiao Wei, Jingdong Zhang, Li Hou, Fenglan Jia, Qinglin Chang, "Backstepping Adaptive Fuzzy Control for two-link robot manipulator", IJCSI International Journal of Computer Science Issues, January 2013, Vol. 10, Issue 1, No 2.

[35] Maki K. Habib (Ed.), Bioinspiration and Robotics: Walking and Climbing Robots, I-Tech Education and Publishing, 2007.

[36] Egidio D'Angelo, "The cerebellar network: revisiting dhe critical issues", The Journal of Physiology (J Physiol), 2011, 589 14, pp 3421-3422.

[37] Cesana E, Pietrajtis K, Bidoret C, Isope P, D'Angelo E, Dieudonné S, Forti L., "Granule Cell Ascending Axon Excitatory Synapses onto Golgi Cells Implement a Potent Feedback Circuit in the Cerebellar Granular Layer", J Neurosci., 2013 Jul 24, 33(30):12430-12446.

[38] Fabio M Simões de Souza, Erik De Schutter, "Robustness effect of gap junctions between Golgi cells on cerebellar cortex oscillations", Neural Systems & Circuits 2011, 1:7.

[39] James M. Bower, "Model-founded explorations of the roles of molecular layer inhibition in regulating Purkinje cell responses in cerebellar cortex: more trouble for the beam hypothesis", Frontiers in Cellular Neuroscience, August 2010, Volume 4, Article 27.

[40] Jin Bao, "The role of short-term synaptic plasticity in neuronal microcircuit", PhD thesis, Biology Faculty, Georg-August-University Göttingen, Germany, 2010.

[41] Robert E. Sims and Nicholas A. Hartell, "Differences in Transmission Properties and Susceptibility to Long-Term Depression Reveal Functional Specialization of Ascending Axon and Parallel Fiber Synapses to Purkinje Cells", The Journal of Neuroscience, March 23, 2005, 25(12):3246 – 3257.

[42] Girija Gundappa-Sulur, Erik de Schutter, and James M. Bower, "Ascending Granule Cell Axon: An Important Component of Cerebellar Cortical Circuitry", The Journal of Comparative Neurology, 2013, 408:580–596.

[43] Paul Dean, John Porrill, Carl-Fredrik Ekerot & Henrik Jörntell, "The cerebellar microcircuit as an adaptive filter: experimental and computational evidence", Nature Reviews Neuroscience, 2010, 11, 30-43.

[44] O'Donoghue DL, Bishop GA., "A quantitative analysis of the distribution of Purkinje cell axonal collaterals in different zones of the cat's cerebellum: an intracellular HRP study", Exp Brain Res., 1990, 80(1):63-71.

[45] Masao Ito, The Cerebellum: Brain for an Implicit Self, Pearson Education, Inc., 2012.

[46] Richard Apps and Martin Garwicz, "Anatomical and physiological foundations of cerebellar information processing", Nat Rev Neurosci., 2005 Apr, 6(4):297-311.

[47] Gilad Jacobson, "Spatio-temporal aspects of information processing in the cerebellar cortex", PhD thesis, Hebrew University of Jerusalem, January 2009.

Human Aspect in security of M-Commerce services in ICTD: A Siyakhula Living Lab Case Study

Marufu Anesu M.C., Sibanda Khulumani and Scott Mfundo S.

Department of Computer Science
University of Fort Hare, Private Bag X1314
Main Campus, Alice, 5700, South Africa
Tel: +27 40 602 2746, Fax: +27 86 248 9404

Abstract
This paper is a build up to a bigger work in progress which essentially looks at undertaking an extensive evaluation of security threats on M-Commerce platform usage in Information Communication technology for development (ICTD) contexts. This initiative will subsequently draw up a framework to inform deployment of such services and platforms in rural marginalized communities. As is commonly acknowledged that security requirements cannot be addressed by technical means alone, a significant aspect of protection comes down to the attitudes, awareness, behaviour and capabilities of the people involved. Likewise the aim of this work was to evaluate rural user's social habits and day-to-day interaction with mobile devices, in order to ascertain the human aspect as a source of threat in mobile device use. M-Commerce information and systems security seeks to provide: Confidentiality, Integrity, Availability, Non repudiation Authentication. We propose to use these security properties as metrics to investigate how security vulnerabilities were introduced by human to device interaction in our field of study. An ethnographic field study was performed under a Living Lab experience in order to understand the related problems and security issues on mobile phone and ICT usage in a marginalized community. Qualitative measures such as contextual inquiry, participant observation, focus and individual interviews were used during the field data collection. However, only preliminary results from field studies are described in this paper. This work however does not provide a concrete solution on how to secure M-Commerce systems but highlights some socio-technical suggestions that can be used in order to attain that goal. Product designers, service providers and all value chain providers must consider the discoveries noted in this paper in order to deliver successful ICT and mobile-based services to users in these areas.
Keywords: *M-Commerce Security; Mobile Device Security; Rural Marginalised Areas (MRA's); Living Lab, Siyakhula Living Lab; ICT, Threats and Vulnerabilities*

1. Introduction

Mobile transacting or Mobile Commerce (M-Commerce) has found its way to previously inaccessible parts of the world as a major Information and Communication Technologies (ICT) tool for development due to widespread introduction of mobile phones even in remote areas[1]. An information and communication technology for development (ICT4D) is a general term referring to the application of ICTs within the fields of socioeconomic development, international development and human rights [2]. A considerable number of mobile phones are now owned and used by people dwelling in resource constraint rural communities of emerging markets. This could be attributed to the idea that mobile phones provide a less expensive means to narrow the digital divide and because they only require basic literacy. This therefore makes them accessible to a large segment of the population across the world [3]. It is worth noting therefore that mobile phones have become essential tools for communication and information exchange in the last two decades [4].

Many people rely on their mobile phones in their personal lives as well as their businesses. As such, most mobile phone users exchange very sensitive and private information using their mobile phones. Likewise, it may be argued that although mobile use in M-Commerce platforms contribute to providing a cheaper and easier method of transacting for rural users, it is not without its challenges which among them security is of paramount importance. In [5] security was noted by the authors to be very critical to the success of M-Commerce and further research should be conducted in this area, in order to determine methods of making mobile transactions more secure. Further works from [6], [7], [8], [9], [10] also agree to the idea that security is a worthwhile endeavour to pursue in M-Commerce field. It is from such works and understanding that our work has sprung.

As is commonly acknowledged, security requirements cannot be addressed by technical means alone. A significant aspect of protection comes down to the attitudes, awareness, behaviour and capabilities of the people involved [11][12][13]. Indeed, people can potentially represent a key asset in achieving secure mobile use, but at present, factors such as lack of awareness and understanding, combined with unreasonable demands from security technologies, can dramatically impede their ability to do so. Ensuring appropriate attention and support for the needs of users should therefore be seen as a vital element of a successful security strategy. It is therefore the intention of this work to ascertain the human aspect as a source of threat in M-Commerce use through evaluation of their social habits and behavioural patterns in mobile phone use in order to suggest new solutions to support secure M-Commerce by rural users.

As described by [14] and [15], M-Commerce information and systems security seeks to provide: authentication, integrity, authorisation, availability and non-repudiation. We propose to use these security properties as metrics to investigate how security vulnerability is introduced by human to device interaction in our study field.

The rest of the paper is organized as follows: Section 2 briefly describes the related work; Section 3 outlines the description of our Study setup; the study site, study participants and the process of data collection used, the research methods, and analysis used. Section 4 presents the themes under investigation in alignment with the results. Finally, our paper ends with presenting study conclusion and future work.

2. Related Work

Some work is evident in the area of M-Commerce, cyber security awareness and mobile device use in the marginalized areas. Some of the work is actually done under LL environment which relates to the choice of using a Living Lab adopted by this research as well. Work by [16] described the development and deployment of an E-Commerce platform in Dwesa, the same area through which initial data collection was carried out for this research. Their work involved the implementation of an E-Commerce platform which can make a contribution to rural development and poverty alleviation in the area. However their work looked at an E-Commerce platform which was designed as a web application, to which security considerations were limited to desktop usage and not mobile usage. Even so the work is a valuable footstool for this work as it maps out some security considerations to be considered in creating an M-Commerce structure. More so, our work goes on further to take a direct assessment on the human aspect in security.

In [17] authors focused on promoting cyber security awareness towards the newly realised broadband capability and knowledge transfer within rural communities by means of a voluntary based training program. The cyber security awareness program modules in their work were divided into physical security, malware and malware countermeasures, safe surfing and social aspects of cyber security. However though this work can have great impacts on rural user awareness on cyber threats, less was mentioned in the area of human-to-mobile device interaction as a source of threat. Henceforth this work takes an angle towards exposing first the threats from human to device interaction before documenting the findings into a manual or awareness campaign. The paper roots its methodology approach in the Critical Social Theory (CST) adopted from related work by [18]. The author aims to contribute to on-going ICT for development (ICT4D) discourses by representing an African voice for international ICT policy interpretation and implementation.

In [19] authors tried to tackle the security issues from both technical and human perspectives; how computer attacks are performed, including how to gain illicit access, the types of attacks, as well as the potential damage that they can cause. They also uncovered sociological and psychological traits of attackers including their community, taxonomy, motives and work ethics. The work supports that humans play important roles in computer security. The paper however focused more on the attackers' side, a clear deviation from our work, which focuses more on the end user's (victims) of M-Commerce systems.

Whenever the human aspect is involved in security issues, culture has to be a part of it. A need to understand the cultural issues affecting security in large, distributed and heterogeneous systems, in this case M-commerce systems is essential. Authors in [20] present a model of security culture for e-Science, grounded both in the security literature and in empirical data from an e-Science project. From this model, we present five concepts, which have differing effects on security culture. Each concept is discussed in terms of how the literature treats it, and how it impacts security culture in practice. This discussion highlights differences and similarities between the two domains. It is from this work that a large part of our discussions are rooted.

3. Study Setup

For a better understanding of the current use of mobile phones and the security implications of user's interaction with their devices, we organized an ethnography field study in one rural community of South Africa under SLL. Through field studies, we tried to identify the different security vulnerabilities introduced by the human aspect. Data to be collected in the scope of highlighting the human aspect in compromising mobile devices (thereby M-Commerce systems) security included: 1) the day to day use of the handheld devices which may introduce some vulnerabilities; from general internet use, transacting, sharing of the device with other users, taking the device for repairs and use of PIN (Personal Identification Number) or passwords; 2) the security measures users employ to protect valuable information on the device from being stolen, lost or misused; 3) social aspects that may undermine the security which include (not limited to) social engineering, identity theft, twin evil attacks and literacy issues.

3.1 Participants and Site

We selected the Siyakhula Living Lab (SLL) located in the Eastern Cape Province of South Africa for our research study. The SLL is a joint initiative of the Telkom Centres of Excellence in the Departments of Computer Science at University of Fort Hare and Rhodes University. The SLL is currently running in the area around the Dwesa-Cwebe Nature reserve within the Mbashe Municipality. The Mbashe Municipality is a deep rural area situated along the Wild Coast of the Eastern Cape province of South Africa. The (initial) five villages targeted in the SLL are adjacent to the Dwesa-Cwebe area, which comprises the nature reserve and adjacent communities which are extended over a land area of approximately 153 square kilometres [21]. The reason we chose to carry out this study under a Living Lab context is that interaction with the rural communities is made easier since the platform for interaction is already in place. The LLs; SLL in particular, also offer better ICT infrastructure which equips rural users with a sense of readiness to engage in IT related issues and provide availability of fundamental ICT distribution channels. The targeted communities in SLL, by sheer size and because of political dynamics, represent a strategic emergent ICT market [22] for such M-Commerce platform security study. The SLL also enables the researchers to have a direct experience of the marginalized rural reality. This brings a crucial understanding of the rurality context to our work, allowing for a study that is relevant and well positioned to meet the M-Commerce security needs of the communities.
Our study consists of three main stakeholders also referred as study participants i.e., students (age group between 16-25 and not working), community members (none teaching and mostly staying at home), teachers/educators (all age groups). The idea was to have participants from literacy training classes at each of the two centres in the SLL. There are basically two centres one at Ngwane and the other at Nqabara at which computer literacy training courses are being held. Each class constituted at least a student, teacher or community member from schools affiliated to the SLL. Figure 1 shows the SLL Network Map and the a distribution of the centres in the Living Lab. Participants who converged at Nqabara came from Badi, Mevana, Kunene, and Nquba, while those at Ngwane were from Mpume, Nondobo, Ntubeni, Ngoma and Mtokwane.

3.2 Research Methods

For our research we applied the observation technique to 1) build rapport with informants, 2) provide a better platform to later cross-check information and possible differences between what people do and what they say they do, 3) get

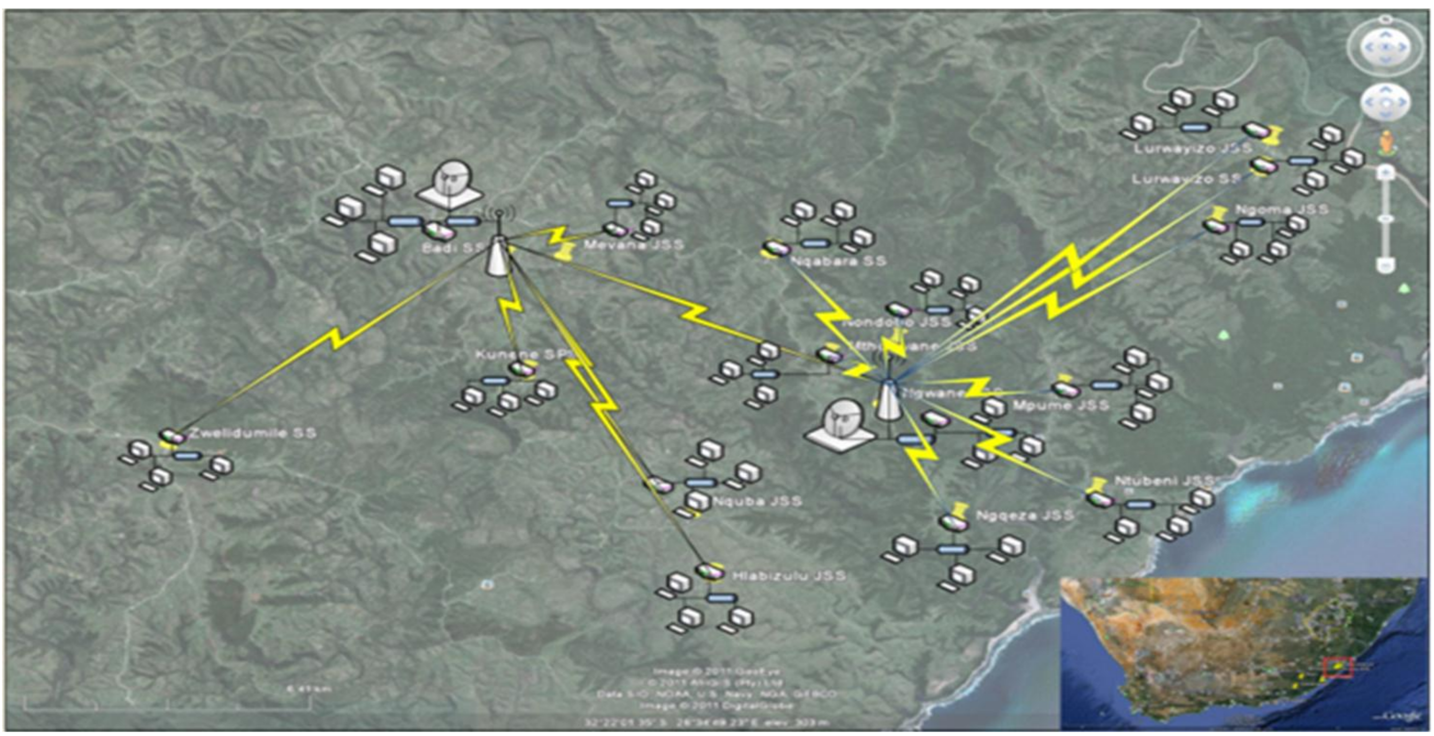


Figure 1: Siyakhula Living Lab Network

a better understanding of mobile device use in an ICTD context 4) to gain new insights or to discover things that people may not wish to reveal in interviews, or may be not asked about in surveys and may not have thought of mentioning, 4) gather data on how the users interacted with their handheld devices and how security vulnerabilities may be introduced. To facilitate the observation technique some field notes, videos and photos/pictures were taken. During the field study, field notes were written to capture and preserve indigenous meanings. To do so, we tried to recognize and limit reliance upon preconceptions about members' lives and activities. We tried to be responsive to what others were concerned about. But while field notes are about others, their concerns and doings gleaned through empathetic immersion, they necessarily reflect and convey the ethnographer's understanding of these concerns and doings [16].

After building rapport with the community members through the literacy training courses at both centres we used the observation technique to gather data on how the users interacted with their handheld devices and how security vulnerabilities may be introduced. We tried to obtain inferences from the way in which the cultural context shapes participants' perceptions about personal data, data use, mobile devices and institutional trust.

The types of mobile devices being used were also noted for the purposes of cross referencing with a baseline study that was carried out in [23]. This was done to determine if the type of device used might influence the type of security measures a user may use and also with the interfacing aspect.

The observation method paved way for the focus groups that were then held in the same context. Focus groups were used in this study because 1) they could yield a large amount of information over a relatively short period of time; 2) they were an effective tool for accessing a broad range of views on our specific topic, as opposed to achieving group consensus, and 3) could act as an enabling tool in developing drafts of interviews and/or questionnaires and 4) could be used effectively in conjunction with other qualitative methods (observations). More so, the group dynamics stimulate conversation and reactions. Within this research, use of focus groups is typically one method among many that we chose to create a complete picture of how lack of enhancing M-Commerce systems affects a community of people. Use of focus groups contributes to this broad understanding by providing well-grounded data on mobile phone use, social norms, the pervasiveness of these norms within the community, and people's opinions about their own values over the M-Commerce environment.

The fundamental data collected by this technique were the transcripts of the group discussions and the moderator's reflections and annotations. Together with the observations technique and the semi-structured questionnaires, the focus groups aided this preliminary research to ascertain human aspect as security loopholes, and even illuminated the results of other data obtained in a baseline study of the SLL on mobile device use. It is a situation described as triangulation, which is the use of two or more different methods, in a complementary way, for the same research subject [].

Ethnography study was carried out in a period of about 5 months. The collected data was analysed using grounded theory method [24]. Based on the study findings, appropriate adjustments were made to our original study. Some questions were modified and added in order to know more about the unclear facts.

During our ethnography field study, we practiced a pure qualitative research methodology where a set of qualitative methods were combined in order to gain rich user data. We practiced observation, focus and contextual inquiry for performing this triangulation. Table 1 lists the number of people who participated and their respective literacy training centres and their areas of residence.

Table 1: Participants in the focus groups at Ngwane and Nqabara in the SLL

Participants	Focus Group Centre	
	Ngwane	Nqabara
Students	3	1
Teachers	6	6
Community Members	2	3

In our study, we took certain precautionary steps to avoid any possible bias in our data collection process. For example, to reduce the gap between researchers and participants, engagement was done in the literacy training sessions prior to the commencement of the actual data collection. In writing field notes, we gave special attention to the indigenous meanings and concerns of the people studied. However, even after taking these precautionary steps, we still faced some challenges for example, most of the participants are fluent in isiXhosa and have difficulty in expressing themselves well in English, and hence translation was required from time-to-time as the facilitators were not fluent in isiXhosa. A colleague fluent in isiXhosa language assisted in the transcription of focus group recordings.

4. Study Results

The results section is divided into two main sub-sections: Observations and Focus groups, and results obtained from the respective tools of collection are presented respectively. Table 2 shows the mapping of the obtained themes to principles of security which were violated from user activity.

Table 2: Focus group themes mapping to elements of security

Systems Security	*Confidentiality*	*Integrity*	*Availability*	*Non repudiation*	*Authentication*
Theft/Loss of Device	X				X
Mobile Device Repairs	X	X	X		X
Device Sharing	X				X
Password and PIN Use	X	X			X
Internet and Safety					
Transacting via M-Commerce	X	X	X	X	X

4.1 Observations

4.1.1 Phone/device penetration

High end smartphones have managed to penetrate into this community, with special mention going to a Samsung galaxy S2 model at Ngwane. This widens the range of devices that would be under scrutiny for security vulnerabilities. The participants had at least one cellphone on them, which may actually indicate the idea of mobile devices becoming more affordable. Lower end devices with weak encryption capabilities were also noted but their prevalence was not as preconceived (were expected to be the dominant devices, which was not exactly the case). Enquiries on penetration of smart devices revealed that there is high readiness to embrace new technology. Teachers in the community are the pioneers of new technologies. In both centres high end smartphones where noted in with the highest being a Samsung galaxy S2.

4.1.2 Threat from user behaviour

In both centres placing a cellphone on the workbench was a norm and from the few conversations held, the environment feels safe to leave one's device on the work bench in a room with a lot of people. We observed Collectivist cultural norms -were sharing of mobile devices was common, and perceptions that mobile devices can be used to store personal data and that no-one else will make use of or retrieve that data. The use of password or PIN as mobile device locking facilities was not a common practice. Most devices with a Bluetooth facility were not fully secured due to user unawareness and/or ignorance.

4.2 Focus Groups

This section is presented into sections showing the different themes that were noted in the focus groups. Table 2 shows a mapping of the themes in the focus groups to the elements of security adopted in this work

4.2.1 Theft/ Loss of mobile devices

The first challenge is that mobile devices are by design small and portable which can result in loss or theft of the device. While their size is good for day to day usability and convenience it is not as good then when it comes to security. As pointed before modern mobile devices contain more and more sensitive information creating a need to protect the sensitive data on the phone even if it is lost or stolen [13]. Under this theme we were looking to find out a number of possible security issues that may expose user privacy through device loss or theft. Data collected included how often the devices were being lost or stolen, the measures the participants were employing to protect data on the devices or device loss/ theft. At both centres mobile device loss was a norm especially for those users who travel a lot with public transport. As was picked from the observations users behave in a way that seemed to support the reason of the device being lost or stolen. For instance during the literacy training classes most mobile devices were placed on the work benches, with little attention, which suggested that the environment may have been more trusted. This trust may be abused and used by perpetrators to obtain the devices and pursue their agenda. But as one member pointed out and we subscribe to; safety of one's mobile phone is up to the individual of that mobile phone. Some participants at Ngwane even suggested the use of purses and bags, but in the end they even adhered to the idea that even with that, there was no guaranteed way to keep devices from being lost or stolen as some may be taken from one with their consent by robbers. As a measure of protecting the data on the devices, use of PIN's and passwords remained synonymous to most members. One member in Nqabara stated the need to register a SIM card to a device on purchase of a new device as a security measure. But as was also noted, a good number of the participants never really had that option as they got the devices from a relative as a gift usually. This also is ceasing to be an option as some acts in place are now ensuring that a device is not locked to a specific network provider. Moreover, an aware user is a secure user. Regardless of the make or model of a device, keeping data secure comes down to how one uses and maintains the device.

4.2.2 Mobile device repairs

Under this theme, it was our intention to try to find out; who the participants and members of the community consulted when faced with challenges in operating their mobile devices, to which they take the devices to for repairs, and what level of trust they have for such places? Among the answers that we got from the participants at both centres, some consultations were evident from a majority of the participants. In response to the first question we found out that they consulted students (for teachers), one's own children, mobile phone retailers, the user manual, colleagues or family members. Of interest is the response we got from one participant at Ngwane: "I take it to anybody who can repair or help me…"At the same centre suggestions of calling the customer care of the service provider, and buying another one were also mentioned.

We learnt that in case of device malfunction these devices are taken to repairmen in the nearest towns or cities. Most repairmen are found in Willowvale, were contact with the majority of the rural users is established. When devices are taken for repairs there, they are most of the times left behind with the repairmen. The members in the focus groups acknowledged and stated that the process was not secure. One participant at Nqabara who had past encounters with these repairmen points out how untrustworthy such places are for mobile device repairs. Due to desperation and a lack of secure trustworthy alternatives, participants still take their devices to the mentioned places.

It was our observation henceforth that major security concerns were raised under this theme. It was noted, if bogus repairmen could get the custody of a device they would be able to search for personal plus important details on the device or even install applications (for example Soundminer [25]) that can listen for sensitive data on the phone and post it to the perpetrator. Data elicitation through some open source tools like MOBILedit! Forensic, EnCase, FTK, Cellebrite and Sleuthkit can be used. Not saying that this was observed to be the direct case but, if mobile device use is to be considered secure such issues are supposed to be addressed in the Security Plan on M-Commerce use by rural users.

4.2.3 Device Sharing

Under this theme we tried to ascertain if some security threats may be introduced by some user activities, in this case; sharing of the mobile device. The first impression we got from the floor in both sessions was that the members did not share their devices to which we later discovered that the majority of the participants did share. As with the observations and informal interviews the most occurring answer was that at least someone could answer their calls in cases where they are absent. This was surely the case as some echoed statements like "my kids can answer my phone." People who confirmed to sharing the mobile device noted sharing it with their kids, colleagues, siblings, spouses, and one's boyfriend/girlfriend. Among those who said they shared one even mentioned that she even allowed her siblings to take the device with them and then return it. At Ngwane however some even stated that they did not share (even in cases where she's helping someone else, the person can only use it in their presence). We also noted that most if not all participants who shared devices, did not employ the use of a password which seemed to be a security concern in cases were the trusted party decides to abuse the trust.

4.2.4 Use of Passwords and PINs

A major security weakness that occurs in mobile use is the lack of user awareness. The typical user is not aware that their actions have security implications. A typical user will not worry about securing their phone through a locking mechanism, or they will not ensure that transactions are secure before proceeding. This can lead to the spread of viruses or leaking of sensitive data. Under this theme we were trying to find out 1) how the participants felt about use of passwords and PINs as security, and 2) if they encountered any problems in storing or memorising the password. At the session at Nqabara most of the members attributed to memorising the password as a real problem. One actually stated she resolved not to use any password as a result. We noted a bit of confusion as some of the participants did not know whether their devices had a PIN or password facility or not. Recalling the PIN is an evident challenge and one user had this to say: "common numbers for example my date of birth as no one will obviously forget his/her date of birth".

Some however stated memorisation of passwords was not a big challenge as they try to eliminate the challenge of memorising by simply using date of birth, sibling's name, or even ID numbers. Besides being easy to recall these participants felt it was safe to use as well stating: "I used to use my date of birth number now my ID number…" and "I use my bank number as my security code, so that nobody knows the number…". These practices may leave the users greatly exposed as stated before, as most perpetrators of cyber-crimes are aware that users make use of their personal information for their password.

When we enquired on how the participants felt about replacing regular passwords with the use of biometrics, we noticed that there was a general impression that there are doubts in the new technologies about authentication. One, at Nqabara was quick to suggest: "if security can be done by a human being, then a human being can change that thing". More sentiments suggested lack of trust on the use of the thumbs as an authentication mechanism, pointing out exposure to identity theft in cases where databases with their information are breeched. At Ngwane we noticed the majority of the participants were a bit confused with the use of biometrics despite explanation of the concept to them. In the end most agreed to the use of biometrics, but this could have been to pressure to conform to other people's views.

Considering how social engineering is on the high as a password one can easily see the users lack awareness of the repercussions of their actions. A possible solution for overcoming the lack of user awareness is to have the wireless carriers offer security training or provide a security awareness pamphlet when a user purchases a mobile device. There will always be people who refuse to read a pamphlet or pay attention to security training but the more people that are aware of the risks the more likely the risks will be reduced. This reason leads to the suggestion of employing a culture of security approach discussed later on. Also, the mobile device manufacturers should be held responsible for making the devices a secure as possible by default. Currently many security features are either disabled by default or are set to the lowest security setting. If manufacturers set their device security defaults to a secure setting then users will always have some level of protection even if they decide not to further configure the phone.

4.2.5 Internet Use and Safety

At this level it was brought to the focus group facilitators' attention that a majority of the participants employed their mobile devices to access the internet and its services. The participants were therefore asked about their surfing habits (which services they accessed, through which sites, if any use of mobile apps was done to access these internet services and which applications they were familiar with). We also established that mobile transacting was a norm among some of the participants with activities including: money transfers (via MPESA); purchasing and transfer of airtime, making online mobile payments (of electricity bills) etc. In general surfing for news and information through mobile browsers, social networking (e.g. Facebook), access of email services and communication were part of the list of activities noted in both focus group sessions conducted. Cellphone banking was a main M-Commerce service that the people engaged in. We also noted that the use of mobile apps was another source of threat to user information as most users downloaded applications from unsafe sources over the internet. It was also noted that the internet is fast becoming part of the participants' lives. Likewise as was established by [21] that accessing internet using mobile phones comes with the same consequences as using a computer or a laptop; it is therefore necessary that users are advised to apply all basic safe surfing best practices on mobile phones as well. Users are further advised to install and frequently update the mobile phone anti-virus software and other security related software as well as the mobile phone patches.

4.2.6 M-Commerce Transacting and Trust Issues

This study being part of a bigger work in progress which essentially looks at undertaking an extensive evaluation of security threats on M-Commerce platform usage in ICTD contexts is was essential to: 1) find out how the participants and members in marginalized areas feel about paying for goods and services using a cellphone, and 2) to identify if there were trust issues on conducting such transactions using mobile devices.

In both sessions we got mixed sentiments in response to the first objective; some felt its better than queuing for services, as services are made accessible without hustles. Some pointed the idea of M-Commerce being actually cheaper and accessible than traditional commerce in some instances. They felt it really helps as they usually encounter transport problems and takes time than just doing it online. However we came to learn that the impression most rural users get is that these banking systems are not secure or safe. One participant at Ngwane explicitly suggested the real threat is at the bank or service provider's end, and not necessarily at the end user side.

Trust is a big issue and if M-Commerce systems are going to reach their fullest potential. Likewise some major breakthrough is required to make sure trust through improved security within the systems is established. Some even suggested that they can only transact using small amounts of money. In the case of mobile banking or internet banking, they suggest it is the bank personnel who are the real perpetrators of most bank scams and fraud.

The focus groups also reviewed a usual occurring trend of losing money unscrupulously through the banks and financial institutions where money "is just transferred" without user or customer consent. In addition, there was a great worry that their personal information is all over the internet (online) - in need of clarification by responsible parties into who has what access to their information and why. Special mention of some governmental departments was stated to be the likely source distributing the information.

It was also established that participants and other members of the community regularly receive numerous calls by different people or companies through whom they do not know where they got their numbers from. These anonymous callers usually would tell them their details and ask them to confirm and provide some information. We noted this down to be a typical social engineering scam.

Furthermore, from the user behaviours and focus groups we established that the users must be made aware of the repercussions of their actions when utilizing their mobile devices. They must understand the losses that they may incur if proper precautions are not followed. The best method would be providing guidance to the user when they purchase the device.

One method of providing guidance would be to provide a pamphlet or booklet along with the device that details all the security features on the phone and what settings should be utilized. The booklet should also detail what a user should be aware of when making transactions on their phone. For example, they should be shown how to verify that they are making secure connections and they are not being tricked by an attacker. This method of providing guidance may not be followed by all users. In order to ensure that all users receive some sort of guidance, the mobile device should also have warnings appear on screen when a user performs a potentially harmful action. For example, if the user attempts to disable the password feature then the phone should make them aware of the repercussions by displaying a warning on the screen that the user must agree to before continuing on. As demonstrated by social engineering, a gullible user could be an Achilles' heel of system security. For that reason, it is necessary to educate the marginalised community users of the different M-Commerce platforms so that they will follow security policy guidelines and do their best to help maintaining the security of their information and money. [13] emphasises the importance of information security awareness and training. He also recommends steps and procedures that one needs to take in order to avoid being a victim of a social engineering attack. This research therefore led to the development of some user guidelines and a training module for the SLL. More so, each participant of an M-Commerce platform is an important actor for ensuring security. Participants, as appropriate to their roles, should be aware of the relevant security risks and preventive measures, assume responsibility and take steps to enhance the security of information systems and networks. This can be achieved by shaping the marginalised communities towards a culture of security- defined by ICASA as a pattern of behaviours, beliefs, assumptions, attitudes and ways of doing things that promotes security. A culture of security is and must be a joint endeavour, from the government through to the M-Commerce service providers, down the value added chain to the ordinary user in a community such as Dwesa.

Promotion of a culture of security will require both leadership and extensive participation and should result in a heightened priority for security planning and management, as well as an understanding of the need for security among all participants. Security issues should be topics of concern and responsibility at all levels of government and business and for all participants. The Guidelines by Organisation for Economic Co-Operation and Development (OECD), and ICASA constitute a foundation for work towards a culture of security throughout society [26]. This will enable participants to factor security into the design and use of all information systems apart from M-Commerce. They propose that all participants adopt and promote a culture of security as a way of thinking about, assessing, and acting on the operations of information systems and networks.

It is important to note that, a culture of security is not an end in itself, but a pathway to achieve and maintain other objectives, such as proper use of information. The greatest benefit of a culture of security is the effect it has on other dynamic interconnections within an enterprise. It leads to greater internal and external trust, consistency of results, easier compliance with laws and regulations and greater value in the enterprise as whole [27].

In [13][28], the author discusses how to achieve a meaningful, intentional security culture. It provides information on the benefits of, and inhibitors to, a culture of security. It further discusses positive and negative reinforcement strategies and the steps to take to achieve the right balance in a security culture program

5. Conclusion and Future Work

In this study, we have presented the results of our ethnography study, organized in the SLL. The aim of this study was to evaluate the human aspect as a source of threat in mobile device use in marginalised communities. It was evident after this study that the security threats are real and need to be addressed. There is a growing demand for understanding the needs and expectations of users dwelling in resource constraint, rural regions of different emerging markets. Ensuring appropriate attention and support for the needs of users should be seen as a vital element of a successful security strategy. User guidelines and a SLL M-Commerce Security training manual are deliverables as output from this study. The major limitation of our existing study was great focus on one set of participants that was attending the literacy training sessions. There is a possibility that some more data or user experiences where left out in such an approach. This study showed some interesting results to be incorporated to the ICTD M-Commerce Security Framework in progress. The approach used in this work can be seen as a roadmap security evaluation plan for other ICTD services which include but not limited to, M-Health, E-Judiciary, and E-Government. We envisage the creation of a framework which can be used by to inform the value added chain on deployment of such services and platforms in rural marginalized communities.

6. Acknowledgments

This work is based on the research undertaken within the Telkom CoE in ICTD supported in part by Telkom SA, Tellabs, Saab Grintek Technologies, Easttel, Khula Holdings, THRIP, NRF SA and GRMDC. The opinions, findings and conclusions or recommendations expressed here are that of the authors and none of the above sponsors accepts any liability whatsoever in this regard

7. References

[1] A. Dhir, I. Moukadem, N. Jere, P. Kaur, S. Kujala, and A. Ylä-jääski, "Ethnographic Examination for Studying Information Sharing Practices in Rural South Africa," no. c, pp. 68–74, 2012.

[2] R. Heeks, "The ICT4D 2.0 Manifesto: Where Next for ICTs and International Development?," *Dev. Informatics Work. Pap. Ser.*, no. 42, pp. 1–35, 2009.

[3] A. Dhir, N. Jere, P. Kaur, M. Heiskala, I. A. Albidewi, and D. M. Alghazzawi, "Design Guidelines for Pervasive Computing : Implications of Technology Use in Africa," no. March, pp. 925–930, 2012.

[4] M. Ahmed, J. Penney, S. Ikki, A. Salami, T. Bath, M. A. Allah, and S. Mansour, "Threats to Mobile Phone Users ' Privacy," pp. 709–737, 2009.

[5] B. L. Charles, F. Monodee, and T. Nurek, "The Critical Success Factors For Mobile Commerce: An Empirical Research paper," University Of Cape town, 2000.

[6] M. Uk, "Challenges in Mobile Electronic Commerce," in *Proceedings of IeC 2000. 3rd Int. Conf. on Innovation through E-Commerce*, 2000, no. 1.

[7] S. I. Sheikh, "MASTER ' S THESIS Mobile Commerce," Luleå University of Technology, 2006.

[8] R. A. J. Gururajan, "A Discussion On Security Risks In Mobile Commerce," *E-bus. Rev.*, vol. 7, no. 2, pp. 1–14, 2006.

[9] A. Grami and B. H. Schell, "Future Trends in Mobile Commerce : Service Offerings , Technological Advances and Security Challenges."

[10] A. Bailes, J. Brown, and L. Coley, "M-Commerce Security : Issues , Trends , & Threats," *ISM 4320*, pp. 1–14, 2006.

[11] S. Lineberry, "The Human Element : The Weakest Link in Information Security," *J. Account.*, vol. 204, no. 5, pp. 44–49, 2007.

[12] I. Arce, *The weakest link revisited [information security]*, vol. 1, no. 2. 2003, pp. 72–76.

[13] K. D. Mitnick and W. L. Simon, *The art of deception: Controlling the Human Element of Security*. Wiley.com, 2001, pp. 1–105.

[14] M. Fuller, "M-Commerce and Security," 2005. [Online]. Available: http://goo.gl/0WZ9i. [Accessed: 20-Aug-2012].

[15] C. Lee, W. Kou, and W.-C. Hu, *Mobile Commerce Security and Payment Methods*. Idea Group Inc, 2005, pp. 292–316.

[16] L. Dalvit, "The Deployment of an e-Commerce Platform and Related Projects in a Rural Area in South Africa," vol. 1, no. 1, pp. 9–18, 2007.

[17] M. Grobler, Z. Dlamini, S. Ngobeni, and A. Labuschagne, "Towards a cyber security aware rural community," in *Proceedings of the 2011 Information Security for South Africa (ISSA) Conference.*, 2011, pp. 1–7.

[18] K. Krauss, "Towards Self-Emancipation in ICT for Development Research : Narratives about Respect , Traditional Leadership and Building Networks of Friendships in Rural South Africa Towards Self-Emancipation," vol. 4, no. 2, 2012.

[19] B. Arief and D. Besnard, "Technical and Human Issues in Computer-Based Systems Security."

[20] S. Faily, I. Fléchais, and S. Culture, "A Model of Security Culture for e-Science," 2003.

[21] R. Palmer, H. Timmermans, and D. Fay, "From Conflict to Negotiation Nature-based development on the South African Wild Coast.," HSRC Press, Pretoria, 2002.

[22] S. Gumbo, H. Thinyane, M. Thinyane, A. Terzoli, and S. Hansen, "Living Lab Methodology as an Approach to Innovation in ICT4D : The Siyakhula Living Lab Experience," in *IST-Africa 2012 Conference Proceedings*, 2012, pp. 1–9.

[23] C. Pade, R. Palmer, M. Kavhai, and S. Gumbo, "Siyakhula Living Lab -Baseline Study," no. April. pp. 1–88, 2009.

[24] B. Glaser and A. Strauss, "The discovery of grounded theory," *Strategies for Qualitative Research.* 1967.

[25] R. Schlegel, K. Zhang, X. Zhou, M. Intwala, A. Kapadia, and X. Wang, "Soundcomber : A Stealthy and Context-Aware Sound Trojan for Smartphones," in *Proceedings of the 18th Annual Network and Distributed System Security Symposium (NDSS)*, 2011, pp. 17–33.

[26] C. O. F. Security, V. Une, and C. D. E. La, "OECD Guidelines for the Security of Information Systems and Networks."

[27] E. N. C. Ena, B. L. I. Ng, U. P. P. Ort, and H. Factors, "The Business Model for Information Security (by ISACA)," 2010.

[28] S. J. Ross, J. Stewart-Rattray, B. Cameron, C. Dimitriadis, W. Goucher, N. Kromberg, O. Sveen, V. Poole, and R. Sethi, "Creating a Culture of Security." © 2 0 1 1 I S A C A, pp. 1–251, 2011.

M.A.C Marufu: is a researcher in the department of Computer Science at University of Fort Hare and is presently studying towards his Master of Science degree at the same institution. He attained his undergrad and honours degrees at the University of Zimbabwe. His research interests include Mobile security, Penetrative testing, M-Commerce and ICT4D

K. Sibanda: is a researcher and senior lecturer in the Department of Computer Science at the University of Fort Hare.

M.S Scott is the Head of Department, a researcher and lecturer in the Department of Computer Science at the University of Fort Hare.

The impact of indexes on data warehouse performance

El Amin Aoulad Abdelouarit[1], Mohamed El Merouani[2] and Abdellatif Medouri[3]

[1] Laboratory modeling and information theory
Abdelmalek Essaadi University, Tétouan, Morocco

[2] Laboratory modeling and information theory
Abdelmalek Essaadi University, Tétouan, Morocco

[3] Laboratory modeling and information theory
Abdelmalek Essaadi University, Tétouan, Morocco

Abstract

A data warehouse designer should consider the effectiveness of data query, this depends on the selection of relevant indexes and their combination with the materialized views, note that the index selection is a NP-complete problem, because the number of indexes is exponential in the total number of attributes in the database, So, it is necessary to provide, while the data warehouse design, the suitable type of index for this data warehouse.

This paper presents, in some steps, a comparative study between the index B-tree and Bitmap type, their advantages and disadvantages, with a real experiment based on two factors: size of index and clustering factor, this shows that the Bitmap index is more advantageous than the B-tree one.

Keywords: *Data Warehouse DBMS, indexes, business intelligence.*

1. Introduction

The data warehouse administrator takes several decisions regarding the administration tasks, such as databases logical and physical designs, management of storage space and performance tuning (performance tuning).

The most important task is the physical design of databases, including data organization and improving access to these data. To improve the access time, the administrator uses general index to quickly find the necessary information without a request to review all the data [1], [3], [5], [7].

Index selection is difficult because their number is exponential in the total number of attributes in the database. So the index plays an important role in the performance of databases, for that we focus on this aspect of the data warehouse, which it considers the focus of the designer when editing and query optimization selection.

The objective is to minimize the query execution time. And as queries in a data warehouse are based on the index, we will work on the problem of choosing the type of index when designing our warehouse data.

There are several types of indexes supported by databases such as Bitmap [4] B-tree [3], [6], [7], [8], Bitmap join [9], range-based bitmap index [10] etc... In this sense we have chosen two types of index relevant to this study, the index type: B-tree index and type Bitmap.

2. Bitmap Index

2.1 Definition

A bitmap index is a data structure defined in a DBMS used to optimize access to data in databases. It is a type of indexing is particularly interesting and effective in the context of selection queries. The index bitmap attribute is encoded in bits, where its low cost in terms of space occupied. [7] All possible attribute values are considered, the value is present or not in the table. Each of these values is an array of bits, called bitmap, which contains as many bits as n-tuples present in the table. Thus, this type of index is very effective when the attributes have a low number of distinct values. Each bit represents the value of an attribute for a given tuple. For each bit, there is an encoding presence / absence (1/0), which indicates that a tuple or not the present value characterized in bitmap.

To illustrate how a bitmap index works, we take an example EE-PP-O'Neil and O'Neil [2]. Table 1 illustrates a basic bitmap index into a table containing 9 records, where the index is created in the C column with integers ranging from 0 to 3, we say that the cardinality of the column C is 4, by what there are 4 distinct values [0, 1, 2, 3], where the index bitmap C Contains 4 bitmaps shown as B0, B1, B2 and B3 corresponding value represents. In this example, the first line where RowID = 0, column C is worth 2, consequently, B2 column bit value "1", while the other bitmaps are set to "0". Same for the next line, where C = 1 corresponds to the bitmap B1 is set to 1 and the rest to "0". This process is repeated for the remaining lines. [12].

Table 1: Basic Bitmap adopted by [9]

ROWID	***C***	***B0***	***B1***	***B2***	***B3***
0	2	0	0	1	0
1	1	0	1	0	0
2	3	0	0	0	1
3	0	1	0	0	0
4	3	0	0	0	1
5	1	0	1	0	0
6	0	1	0	0	0
7	0	1	0	0	0
8	2	0	0	1	0

2.1 Properties

Bitmap indexes have a very interesting property of responding to certain types of requests without returning the data themselves, thus optimizing the response time, disk storage. This is possible by counting operations (COUNT) and logical operators (AND, OR, etc.) that act "bit by bit" on bitmaps.

3. Bitmap Index

3.1 Definition

The index B-tree stores the index values and pointers to other index nodes using a recursive tree structure. [3], [6], [7], [8] The data are easily identified by traces pointers. The highest level of the index is called the root while the lowest level is called the leaf node or "leaf node". [7] All other levels between them are called branches or internal nodes. All roots and branches contain entries that point to the next level of the index. Leaf nodes consist of the index key and pointers pointing to the physical location of records. We present details of the index B-tree structure [7].
The B-tree structure is used by the database server to configure the index (Figure 1)

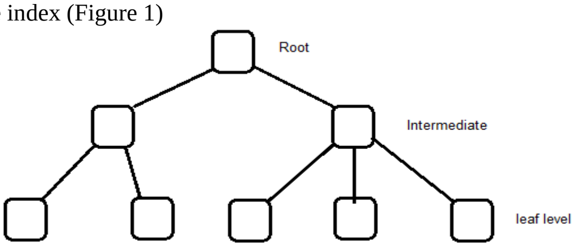


Fig 1: B-tree structure

Root or root is the highest level of the index points to the following levels of nodes branches.

Intermediate nodes or branches contain pointers to the following branches or to the leaf nodes level.

Node leaves or leaf nodes: the lowest level of the index points to other node leaves.

4. Hypothesis

The conventional wisdom is that bitmap indexes are most appropriate for columns having low distinct values - such as gender, marital status, and relationship. This assumption is not entirely accurate, however. In reality, a bitmap index is always advisable for systems in which the data is not updated frequently by many competing systems. In fact, as I will demonstrate here, a bitmap index on a column with unique values to 100% (candidate of the primary key column) is as effective as a B-tree index.

5. Analysis and results

5.1 Analysis

As known, the bitmap index is more efficient than the b-tree index by its low cardinality columns, we present in this experimentation the performance given by Bitmap index comparing with the B-tree.
We use two factors to of cost-based optimizer:

- The index size
- Clustering factor

Step 1:

In our Data warehouse schema, we created a table named “Employees” with 100000 records and with a column named employee_id with 100000 distinct values, and then we added a GRADE column with 4 distinct values only.

Step 2:

We create now a standard B-Tree index on the GRADE column using this SQL:

SQL> create index employees_grade_i on employees(grade);
Then we check the index size using this query:

SQL> select index_name, index_type, distinct_keys, blevel, leaf_blocks from dba_indexes where index_name=’EMPLOYEES_GRADE_I’;

And we got this result:

Table 1: Basic Bitmap adopted by [9]

INDEX_ NAME	***INDEX _TYPE***	***DISTINCT _KEYS***	***BLEVEL***	***LEAF_ BLOCKS***
employee s_grade_i	Normal	4	1	176

Step 3: Creating a bitmap index on the same column to compare the size (dropping the b-tree index created in first step)
SQL> create bitmap index employees_grade_bitmap_ii on employees(grade);

Step 4: In the bitmap index size checking, we use the same query:

SQL> select index_name, index_type, distinct_keys, blevel, leaf_blocks from dba_indexes where index_name='EMPLOYEES_GRADE_II';

Table 3: Bitmap index size checking result in GRADE column

INDEX_ NAME	***INDEX _TYPE***	***DISTINCT _KEYS***	***BLEVEL***	***LEAF_ BLOCKS***
employee s_grade_i i	Bitmap	4	1	10

Note that the index size is reduced from 176 to 10 (while going from B-tree to bitmap index)

Step 5: the bitmap index creation on employee_id column that contains 100000 distinct values:

SQL> create bitmap index employees_empid_bitmap_i on employees(employee_id).

By checking the index size using the same query, we have this table as result:

Table 4: Bitmap index size checking result in EMPLOYEE_ID column

INDEX_ NAME	***INDEX _TYPE***	***DISTINCT _KEYS***	***BLEVEL***	***LEAF_ BLOCKS***
employee s_empid_ bitmap_i	Bitmap	100000	1	348

And when trying with B-tree index we have this result:

Table 5: B-Tree index size checking result in EMPLOYEE_ID column

INDEX_ NAME	***INDEX _TYPE***	***DISTINCT _KEYS***	***BLEVEL***	***LEAF_ BLOCKS***
employee s_empidb tree_i	B-tree	100000	1	222

5.2 Results

For large distinct values B-tree index occupies less size, and for minimal distinct values, the bitmap index occupies less size.

Clustering factor: considered as the sum of rows orders in a table based on the index values.

- If this amount is near the number of blocks, then the table order is well done, and the index entries in a single leaf block are pointing to rows stored in the same data blocks.
- If the value is near the number of rows, then the table is randomly ordered, so is improbable that the index entries in a single leaf block are pointing to rows stored in the same data blocks.

Table 6: Clustering factor and blocks used for B-Tree index on GRADE column

INDEX_NAME	***CLUSTERING _FACTOR***	***BLOCKS***
employees_grade_i	1148	256

Table 7: Clustering factor and blocks used for Bitmap index on GRADE column

INDEX_NAME	***CLUSTERING _FACTOR***	***BLOCKS***
employees_grade_ii	20	16

6. Conclusion and future work

By using the B-tree index, the optimizer opted for a full table scan; this operation makes a higher clustering factor, whereas in the case of bitmap index that makes a low Clustering factor, he used to answer the query. You can deduct the performance by the number of I / O required fetching the result.
The message here is pretty clear. Both indices have a similar goal: to return results as fast as possible. But the choice of which one to use should depend only on the type of application, and not on the level of cardinal.
As future work, another study will be done on data warehouse schema comparison, especially it impact on data warehouse performance.

References

[1] S. Chaudhuri, U. Dayal,An Overview of Data Warehousing and OLAP Technology., ACM SIGMOD RECORD. 1997

[2] E. E-O'Neil and P. P-O'Neil, Bitmap index design choices and their performance impli-cations, Database Engineering and Applications Symposium. IDEAS 2007. 11th International, pp. 72-84.

[3] R. Kimball, L. Reeves, M. Ross, The Data Warehouse Toolkit. John Wiley Sons, NEW YORK, 2nd edition, 2002

[4] W. Inmon, Building the Data Warehouse., John Wiley Sons, fourth edition, 2005

[5] C. DELLAQUILA and E. LEFONS and F. TANGORRA, Design and Implementation of a National Data Warehouse. Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006 pp. 342-347

[6] D. Comer,Ubiquitous b-tree, ACM Comput. Surv. 11, 2, 1979, pp. 121-13

[7] R. Strohm, Oracle Database Concepts 1g,Oracle, Redwood City,CA 94065, 2007

[8] C. Dell aquila and E. Lefons and F. Tangorra, Analytic Use of Bitmap Indices. Proceedings of the 6th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, Greece, February 16-19, 2007 pp. 159

[9] P. O'Neil and G. Graefe, Multi-table joins through bitmapped join indices, ACM SIGMOD Record 24 number 3, Sep 1995 , pp. 8-11.

[10] K. Wu and P. Yu, Range-based bitmap Indexing for high cardinality attributes with skew, In COMPSAC 98: Proceedings of the 22nd International Computer Software and Applications Conference. IEEE Computer Society, Washington, DC, USA, 1998, pp. 61-67.

El Amin Aoulad Abdelouarit Is a Database administrator in Tanger Med Port, Morocco, PhD Student doing research in Data warehouse and Data mining and it application in Port Management.

Mohamed El Merouani is professor of mathematics, with interests in Probability, Statistics, Stochastic operational research and Data mining, professor of Statistics and Computer Sciences, Poly disciplinary Faculty of Tétouan, Abdelmalek Essaâdi University, Morocco.
University of Granada, Spain, Ph.D. in Mathematics, 1995.

Abdellatif Medouri is Full professor of physics, with interests in Telecommunications, Information theory and Databases; professor of Statistics and Computer Sciences, Poly disciplinary Faculty of Tétouan, Abdelmalek Essaâdi university, Morocco.
University of Granada, Spain, Ph.D. in physics, 1993.

Link Quality and MAC-Overhead aware Predictive Preemptive Routing Protocol for Mobile Ad hoc Network

Moussa ALI CHERIF[1] and Mohamed Kamel FERAOUN[1]

**[1]Computer Science Department, Sidi Bel Abbes University
Sidi Bel Abbes, Algeria**

Abstract

In Ad Hoc networks, route failure may occur due to less received power, mobility, congestion and node failures. Many approaches have been proposed in literature to solve this problem, where a node predicts pre-emptively the route failure that occurs with the less received power. However, this approach encounters some difficulties, especially in scenario without mobility where route failures may arise. In this paper, we propose an improvement of AODV protocol called LO-PPAODV (Link Quality and MAC-Overhead aware Predictive Preemptive AODV). This protocol is based on new metric combine more routing metrics (Link Quality, MAC Overhead) between each node and one hop neighbor. Also we propose a cross-layer networking mechanism to distinguish between both situations, failures due to congestion or mobility, and consequently avoiding unnecessary route repair process. The LO-PPAODV was implemented using NS-2. The simulation results show that our approach improves the overall performance of the network. It reduces the average end to end delay, the routing overhead, MAC errors and route errors, and increases the packet delivery fraction of the network.

Keywords: *Ad-Hoc networks, AODV, PPAODV, QoS, Cross layer*

1. Introduction

An ad hoc network consists of mobile nodes, which communicate with each other through multi-hop routes. Nodes cooperate with their neighbors to route data packets to their final destinations. In ad hoc networks, network topology is changing continuously because of the node movement. To maintain the communication between nodes, many routing protocols have been proposed, which are classified under two categories: table-driven and on-demand routing protocols.

On-demand routing protocols discover routes only when the source needs to send packets. Therefore, there is almost no route maintenance overhead, whereas the route discovery before data transmission increases the delay. However, if the link failure happened, nodes should inform the sources to change the existing route and retransmit the packets that were lost due to link failure. Therefore, on-demand routing protocols increase delay and decrease the successful packet arrival ratio. This causes the reduction of the packet delivery ratio.

Several approaches have been proposed [3,4] to flexibly anticipate link failure by adding a function that predicts the link failure in one of the popular on-demand routing protocols which is Ad hoc On-demand Distance Vector (AODV).

Previous approaches encounter some difficulties, especially in scenario without mobility. The problem is that these approaches predict link failures based of RSS information and interpret that it happened due to node mobility, where actually it was due to congestion. Therefore, the process of route repair should not be performed since it increases even more the congestion, decreasing the overall performance of the network.

Transmitting information to a neighboring node in MAC layer is preceded by the exchange of Request To Send (RTS)/Clear To Send (CTS) frames. If this communication fails, the MAC layer waits (back off time) and retries later. After several failed attempts, the MAC layer informs the routing layer using a cross layer interaction. In our approach, the cause of that unsuccessful communication is sent to the routing layer. If the last received power of the destination node indicates that it is reachable, the routing layer is informed, using the variable xmit_reason with the value XMIT_REASON_HIGH_RSS. Depending on this information a node will decide whether it performs a route repair or not.

In this paper, we propose Link Quality and MAC-Overhead aware Predictive Preemptive Ad hoc On-Demand Distance Vector (LO-PPAODV), it is an on-demand routing protocol based on new metric combine more routing metrics (Link Quality, MAC Overhead), that aims to create congestion-free routes by making use of information gathered from the MAC layer. Also we propose a cross-layer networking mechanism to distinguish between both situations, failures due to congestion or mobility, and consequently avoiding unnecessary route repair process, where we use a “Route Failure Prediction Technique” based on the Lagrange interpolation for estimating whether an active link is about to fail or will fail.

The rest of the paper is organized as follows. Section 2 describes related works; section 3 describes an overview of AODV; the proposed protocol is presented in section 4 and its performance is evaluated and compared with that of PPAODV in section 5. Some conclusions are given in section 6.

2. Related Works

In [8] Norman and Joseph propose an energy efficient routing protocol (HLAODV) for heterogeneous sensor networks with the goal of finding the nearest base station or sink node. Hence the problem of routing is reduced to finding the nearest base station problem in heterogeneous networks.

Xiaoqin, Jones and Jayalath In [10] have proposed the Congestion Aware Routing protocol for Mobile ad hoc networks (CARM). Also they have proposed a congestion-aware routing metric which was employed data-rate, MAC overhead, and buffer queuing delay.

In [11] the authors have proposed a link availability-based QoS-aware (LABQ) routing protocol for mobile ad hoc sensor networks based on mobility prediction and link quality measurement, in addition to energy consumption estimate was proposed.

In [12] Yi and Shakkottai have developed a fair hop-by-hop congestion control algorithm with the MAC constraint was being imposed in the form of a channel access time constraint, using an optimization-based framework. They have used a Lyapunov-function-based approach.

Chen and Heinzelman [13] have proposed a QoS-aware routing protocol that were an admission control scheme and a feedback scheme to meet the QoS requirements of real-time applications was incorporated.

Chenxi and Corson [14] have developed a QoS routing protocol for ad hoc networks using TDMA. They aims to establish bandwidth guaranteed QoS routes in small networks whose topologies were changed at low to medium rate.

In [15] CRP, a congestion-adaptive routing protocol for MANETs, was proposed by Tran and Raghavendra. CRP tried to prevent congestion from occurring in the first place, rather than dealing with it reactively.

In [16] a cross-layer designs among physical, medium access control and routing (network) layers, using Received Signal Strength (RSS) was proposed by Chandran and Shanmugavel. Their object was energy conservation, unidirectional link rejection and reliable route formation in mobile ad hoc networks.

Xia, Ren and Liang [17] have introduced a method for cross-layer design in mobile ad hoc networks. They have used fuzzy logic system (FLS) to coordinate physical layer, data link layer and application layer for cross-layer design.

Authors in [19] have proposed a link availability-based QoS-aware (LABQ) routing protocol for mobile ad hoc networks based on mobility prediction and link quality measurement, in addition to energy consumption estimate.

Baboo and Narasimhan [20] have proposed a hop-by-hop congestion aware routing protocol which employs a combined weight value as a routing metric, based on the data rate, queuing delay, link quality and MAC overhead. Among the discovered routes, the route with minimum cost index is selected, which is based on the node weight of all the in-network nodes.

In [21] Bisengar, Rziza and Ouadou have proposed an improvement of AODV protocol called AMAODV (Adaptative Mobility aware AODV). This protocol is based on new metric combine more routing metrics (distance, relative velocity, queue length and hop count) between each node and one hop neighbor.

In [23] a model was proposed that extends the existing AODV routing protocol to accommodate additional QoS constraints for minimum session bandwidth required for an application. Extensions are added to the current AODV messages during route discovery which specify the QoS requirements.

In [24] Sedrati, Bilami and Benmohamed propose a new variant based on the AODV which gives better results than the original AODV protocol with respect of a set of QoS parameters and under different constraints, taking into account the limited resources of mobile environments (bandwidth, energy). The proposed variant (M-AODV) suggests that the discovering operation for paths reconstruction should be done from the source. It also defines a new mechanism for determining multiple disjoint (separated) routes.

In order to reduce the number of broken routes, the authors propose [25] a novel reliable routing algorithm using fuzzy applicability to increase the reliability during the routing selection. In the proposed algorithm source chooses a stable path for nodes mobility by considering nodes position/ velocity information. Also they propose novel method for rout maintenance, in this protocol before breaking packet transmitted path a new one is established.

2.1 Link failure prediction methods

In [3], a Predictive Preemptive AODV (PPAODV) was proposed which predicts the link failure using the Received Signal Strength (RSS) has been proposed. The prediction method uses Lagrange interpolation, which approximates the process of RSS by means of n-dimensional function with information of past RSS.

PPAODV [3] discovers a new route before the active route becomes obsolete and changes the route smoothly by predicting a RSS of data packets at the Predict Time t_{PT} from the past information of RSS. PPAODV [3] sets Discovery Period T_{DP} as the minimum time that a node can exchange one data with the neighboring node.

In [4], the authors have proposed a High Precision - PPAODV (HPPPAODV) which is an amelioration of PPAODV. HPPPAODV can improve the prediction accuracy ratio by 1) using the Newton interpolation, 2) adding the chance of acquisition of RSS to reduce the error margin of RSS that is affected by the influence of the thermal noise and fading and 3) predicting the value of the Discovery Period T_{DP} by the number of hop in a route.

3. AODV Overview

AODV [1,2] is an on-demand routing protocol. Route discovery is initiated only when a source node needs to communicate with a destination for which it does not have a route in its routing table. To discover a route to a destination, the source node broadcasts a route request message (RREQ) that contains a request ID. If a node receives a RREQ that it has already received, it drops the request. Otherwise, it stores the address of the node from which it received the request so as to establish a reverse route to the source that it uses later. If the RREQ reaches a node that has a route to the destination, the node sends, over the reverse route, a route reply message (RREP) to the source. The reply message contains the number of hops needed to reach the destination from the node. If the RREQ reaches the destination, it sends a route reply to the source over the reverse route.

Intermediate nodes that do not have a path to the destination re-broadcast the request. As the RREP is sent back to the source over the reverse path each node stores the address of the node that sent the reply. The forward path thus determined from the source to the destination is used for sending packets to their destination. AODV uses sequence numbers maintained for the different destinations so to guarantee freshness of routing information.

A link breaks when a node within an active route moves out of the transmission range of its upstream neighbor. When a link break occurs, the node upstream the break invalidates, in its routing table, all routes become unusable due to the loss of the link. It then creates a Route Error (RERR) message, in which it lists the destinations that have become unreachable because of the loss of the link. The RERR is sent to all source nodes that use the link. This procedure is named global repair. AODV also includes a local repair mechanism to locally recover from link losses. Local repair is triggered when a link break occurs between nodes within an active route. In this repair, the node upstream the break tries to find alternative sub-paths to the destinations of packets that it has received, but is unable to forward them (packets) because of the link break.

4. The proposed LO-PPAODV

4.1 Protocol Overview

4.1.1 Link Quality Estimation

In this paper, two-ray ground model is adopted.
This model [27] considers both the direct path and a ground reflection path. The model gives more accurate prediction at a long distance than the free space model. The received power is predicted by:

$$P_r(d) = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4 L} \quad (1)$$

Where *Pt* is the transmitted signal power.
Gt and *Gr* are the antenna gains of the transmitter and the receiver respectively.
L is the system loss, *d* is the distance between transmitter and receiver. *ht* and *hr* are the heights of the transmit and receive antennas respectively.
In this paper, we suppose that the transmit range of each node is equivalent.

So, the link quality $Lq = Pr$ (2)

4.1.2 Estimating MAC Overhead

We consider IEEE 802.11 MAC with the distributed coordination function (DCF). It has the packet sequence as request-to-send (RTS), clear-to-send (CTS), and data, acknowledge (ACK). The short inter frame space (SIFS) is the amount of time between the receipt of one packet and the transmission of the next. Then the channel occupation due to MAC contention is:

$$Coc = t_{RTS} + t_{CTS} + 3t_{SIFS} \quad (3)$$

Where t_{RTS} and t_{CTS} are the time consumed on RTS and CTS, respectively and t_{SIFS} is the SIFS period.
Then the MAC overhead OH_{MAC} can be represented as:

$$OH_{MAC} = C_{oc} + t_{ac} \quad (4)$$

Where t_{ac} is the time taken due to access contention.
The amount of MAC overhead is mainly dependent upon the medium access contention, and the number of packet collisions. That is, OH_{MAC} is strongly related to the congestion around a given node.

OH_{MAC} can become relatively large if congestion is incurred and not controlled, and it can dramatically decrease the capacity of a congested link.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

LO-PPAODV employs a combined weight metric in its cost function. The node weight metric *fpd* which assigns a cost to each link in the network. Weight function *fpd* combines the link quality *Lq* and MAC overhead OH_{MAC} to select optimal paths.

The *fpd* for the link from node *i* to a particular neighboring node is given by :

$$fpd = (Lq)/(OH_{MAC}) \quad (5)$$

LO-PPAODV is reactive routing protocol; no permanent routes are stored in nodes. The source node initiates route discovery procedure by broadcasting. The RREQ message is organized as detailed in Table 1.

Table 1. RREQ message in LO-PPAODV

.TYPE u_int8_t	Reserved	HOP COUNT u_int8_t
RREQ BROADCAST ID u_int32_t		
DESTINATION IP ADDRESS nsaddr_t		
DESTINATION SEQUENCE NUMBER u_int32_t		
SOURCE IP ADDRESS nsaddr_t		
SOURCE SEQUENCE NUMBER u_int32_t		
Cost *fpd* double		

When the source node issues a new RREQ (figure1), the *fpd* value in RREQ is initialized to 65536 (2^{16}).
After the destination node receives the first RREQ, it starts to wait for a period of time to receive enough RREQs. Then it selects the route with the biggest cost *fpd* value and sends back a Route Reply (RREP) to the source node via the selected route.
If there are multiple routes with the same ***cost*** the route with the smallest hop count is selected. Let p_c be the chosen path and p_a the set of all possible paths. Then the chosen path fulfills:

$$cost(p_c) = \min\ max_{p_j \in p^a}\left(cost(p_j)\right)$$

Upon receiving the RREP, an intermediate node records the previous hop and relays the packet to the next hop.
If a node detects a link break during route maintenance phase, it sends a Route Error (RERR) packet to the source node. Upon receiving the RERR, the source node initiates a new round of route discovery.

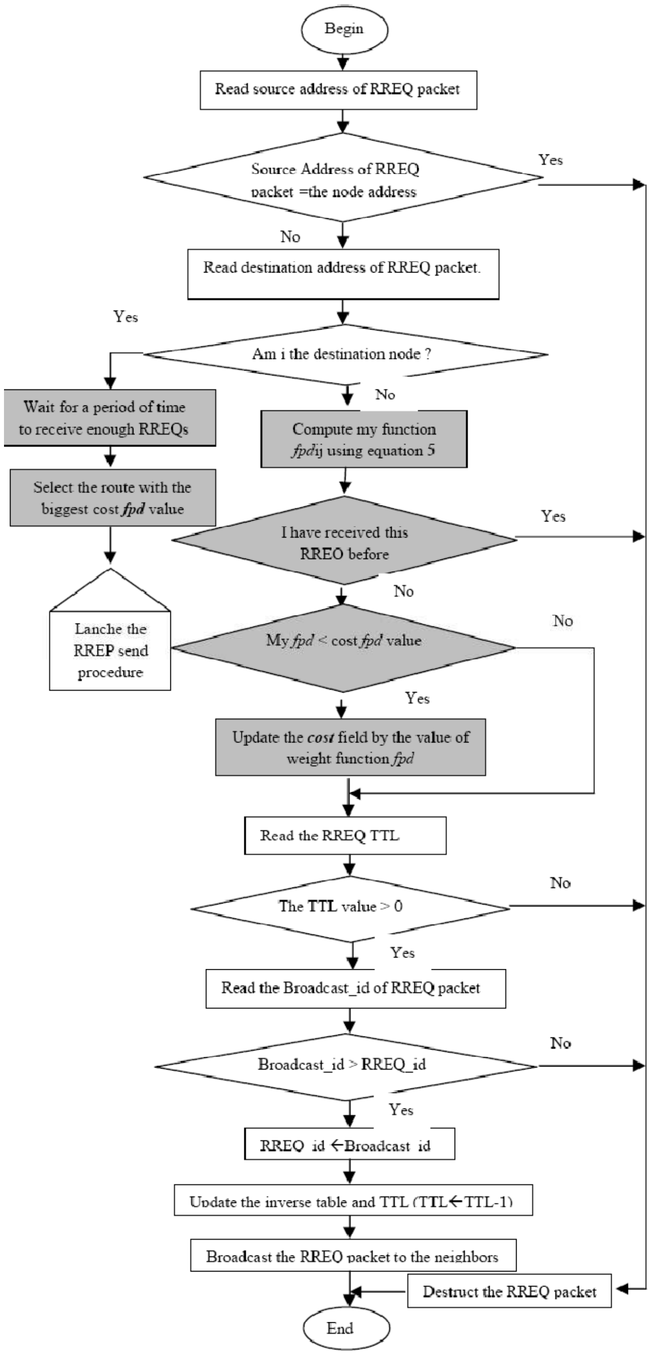


Fig. 1 flow char for RREQ in LO-PPAODV

4.2 The proposed mechanism for Congestion Control

In LO-PPAODV we implemented a cross layer approach that tracks the RSS of received data packet from each neighboring node in order to know when an adjacent node is near enough for a successful transmission.

We use a “Route Failure Prediction Technique” based on the Lagrange interpolation (6) for estimating whether an active link is about to fail or will fail, and it can distinguish between both situations; link error at MAC layer was due to congestion and due to mobility of nodes to avoid the unnecessary route repair process. The Predict Time (t_{PT}) is calculated as (7) and the Discovery Period T_{DP} can be calculated as (8).

$$P(t_{PT}) = \left(\frac{(tPT-t2)(tPT-t3)}{(t1-t2)(t1-t3)}\times P1\right) + \left(\frac{(tPT-t1)(tPT-t3)}{(t2-t1)(t2-t3)}\times P2\right) + \left(\frac{(tPT-t1)(tPT-t2)}{(t3-t1)(t3-t2)}\times P3\right) \quad (6)$$

Where P (t_{PT}) is the value of RSS at t_{PT}. P1 −P3 and t1−t3 are 1st −3rd RSS and their received time respectively.

$$t_{PT} = t3 + T_{DP} \quad (7)$$

$$T_{DP} = T_{warning}\times n_{A\text{-}S}+T_{RREQ} \times n_{S\text{-}D}+T_{RREP} \times n_{S\text{-}D} \quad (8)$$

Where, $T_{warning}$, T_{RREQ} and T_{RREP} represent the transmission time of warning packet, RREQ packet and RREP packet, respectively. Also n_{A-S} and n_{S-D} represent the number of hops between node “A” to node “S” of the active route and number of hops between node S to node D of a new route, respectively.

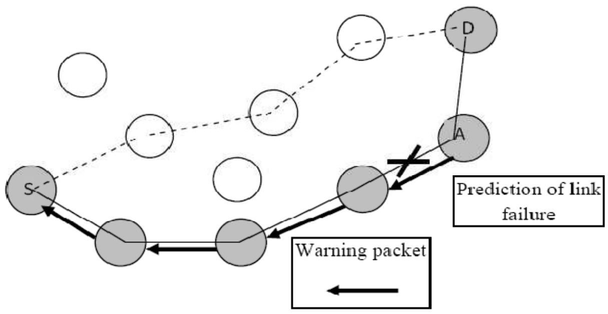


Fig. 2 Node A predicts link failure

4.2.1 Extension of MAC layer

AODV interprets a link failure (in MAC layer) as a broken link, even when it was caused by congestion at the receiver. The sender node should know why communication was impossible. We implemented an approach that tracks the RSS of received data packet from each neighboring node in order to know when an adjacent node is near enough for a successful transmission. If lost packets were due to congestion and high traffic, AODV triggers route repair, and this can affect the network performance. If lost packets is due to low signal quality or misrouted packets, then route repair is needed because the receiver is not reachable.

Afterward, the signal strength of neighboring nodes can be used to detect the reason for lost packets, distinguishing between congestion and broken links due to mobility, because in the last case, the receiver is unreachable and its signal strength is now available. The implementation is divided into two parts; the first part keeps the last three received signals from a node in an array, and computes RSS using Lagrange Interpolation (from the received data packets) if the signal is weak enough and the node moving away, the MAC layer sends a Request To Send (RTS) and the second part decides the kind of message (link failure, either due to errors or due to congestion using signal strength of neighboring nodes) to be sent to the upper layer, whenever the communication is impossible but the destination node is in the transmission range of the sender.

Transmitting information to a neighboring node in MAC layer is preceded by the exchange of Request To Send (RTS)/Clear To Send (CTS) frames. If this communication fails, the MAC layer waits (back off time) and retransmits later. After several unsuccessful attempts, the MAC layer informs to the routing layer that communication failed. In our approach, the reason for that unsuccessful communication is sent to the routing layer. If the last received power (the result of Lagrange interpolation) of the destination node indicates that it is reachable, the routing layer is informed, using the variable xmit_reason with the value XMIT_REASON_HIGH_RSS. In this case, the routing layer should interpret that communication to destination was impossible, not because of a broken link but rather congestion, therefore, route maintenance is not needed. If that is not the reason delivered to the routing layer, a route maintenance process is required.

4.2.2 Extension of AODV

When a node tries to communicate with a neighboring node and this communication failed (after several attempts, MAC layer sends an error to the routing layer). AODV [1] interprets that the neighboring node is not present anymore and communication failure was due to mobility.

In a scenario without mobility communication failures may arise, but AODV will interpret that it was due to mobility, where actually, it was due to congestion. Therefore, the process of route repair should not be performed since it increases even more the congestion, decreasing the overall performance of the network. The proposed amelioration will make AODV [1] capable to distinguish between both situations, avoiding the route repair process when the link error at MAC layer was due to congestion and not due to mobility of nodes. In our approach, when a node is not able to communicate with a neighboring node, MAC layer informs to the upper layer that there was a problem including whether the neighboring node is still reachable or not (see figure 3). Therefore, the sender node does not perform route maintenance if it was informed that the neighboring node is still reachable.

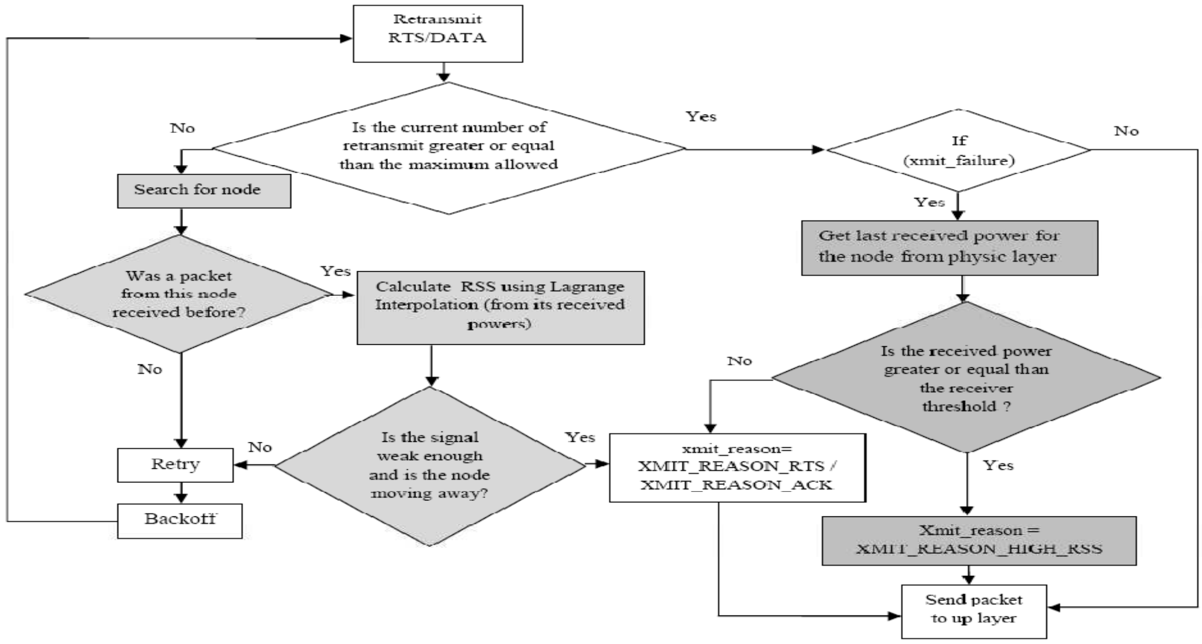


Fig. 3 The proposed approach that uses the Lagrange interpolation is shown here, this diagram shows also how MAC layer informs to the routing layer, when several attempts to communicate to the receiver node failed [28].

5. Simulation and Performance Results

We have used the implementation of AODV [1] in the NS simulator version 3.35 [5]. Our results are based on the simulation of 50 wireless nodes forming an ad hoc network moving about in an area of 1500 meters by 300 meters for 200 seconds of simulated time. The physical radio characteristics of each mobile node's network interface, such as the antenna gain, transmission power, and receiver sensitivity, are chosen to approximate the Lucent WaveLAN [6] direct sequence spread spectrum radio.

The movement scenario files used for each simulation are characterized by a pause time. Each node begins the simulation by selecting a random destination in the simulation area and moving to that destination at a speed distributed uniformly between 0 and 10 meters per second. It then remains stationary for pause time seconds. This scenario is repeated for the duration of the simulation. We carry out simulations with movement patterns generated for 5 different pause times: 0, 20, 40, 80 and 200 seconds. A pause time of 0 seconds corresponds to continuous motion, and a pause time of 200 (the length of the simulation) corresponds to limited motion. Constant bit rate (CBR) sources are used in the simulations. The packet rate is 4 packets/sec when 10, 20, 30 and 40 sources are assumed. The performance metrics used to evaluate performance are:

Packet delivery ratio: The ratio of the data packets delivered to the destination to those generated by the CBR sources. This should be maximized.

Average end-to-end delay of data packets: This includes all possible delays caused by buffering during route discovery, queuing at the interface queue, retransmission delays at the MAC layer, and propagation and transfer times. This should be minimized.

Normalized routing load: The number of routing packets transmitted per data packet delivered to the destination. This should be minimized.

Route errors: Each time PPAODV or LO-PPAODV performs a route error process at sender; it is registered and showed in the graphic (figure 18). A route error in PPAODV triggers a route maintenance process provoking more control traffic in the network. Usually these kinds of errors are due to broken links because of the mobility of nodes, but they may arise from collision of packets, as well. These errors should be minimized.

We report the results of the simulation experiments for the Predictive Preemptive AODV protocol (PPAODV) and for LO-PPAODV.

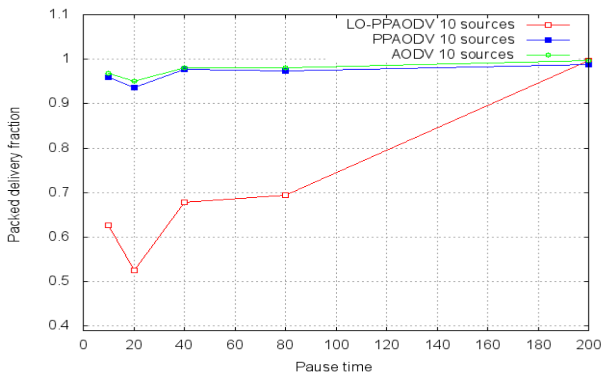


Fig. 4 Packet delivery fraction 10 sources

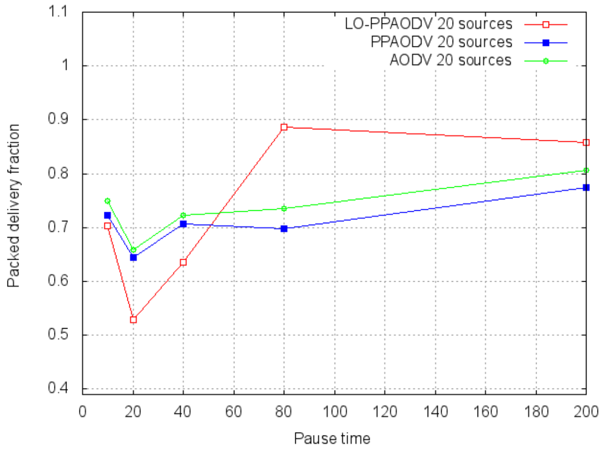


Fig. 5 Packet delivery fraction 20 sources

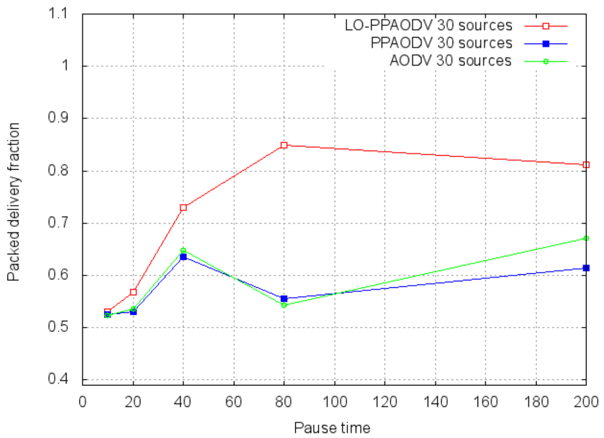


Fig. 6 Packet delivery fraction 30 sources

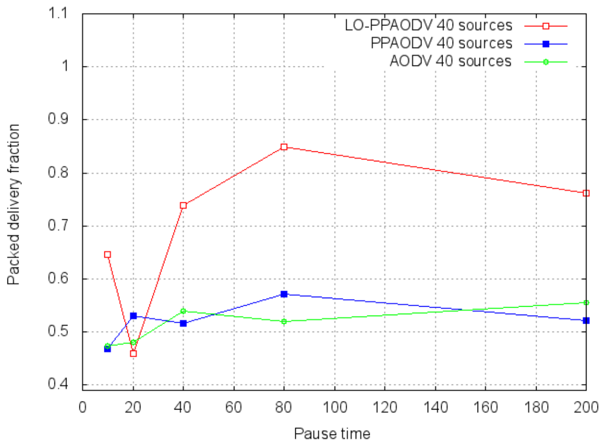


Fig. 7 Packet delivery fraction 40 sources

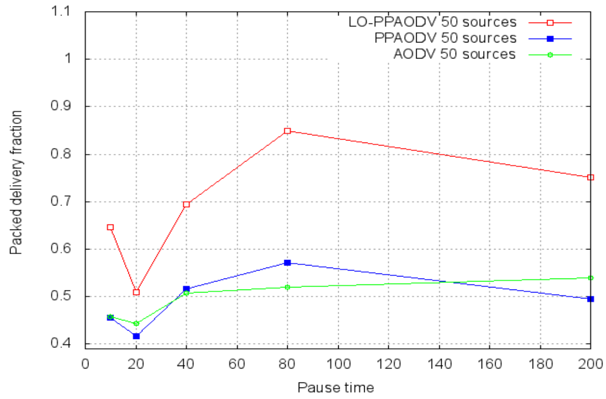


Fig. 8 Packet delivery fraction 50 sources

Figure 4, figure 5, figure 6, figure 7 and figure 8 represents the simulation results for the delivery ratio metric of 10 sources, 20 sources, 30 sources, 40 sources and 50 sources respectively. It can be seen that the method proposed can result in significant performance gains. (The results show that LO-PPAODV outperforms PPAODV significantly when the number of sources increases 30 sources, 40 sources and 50 sources see figure 5 to figure8). We observe that the packed delivery fraction increases significantly when the number of sources increases.

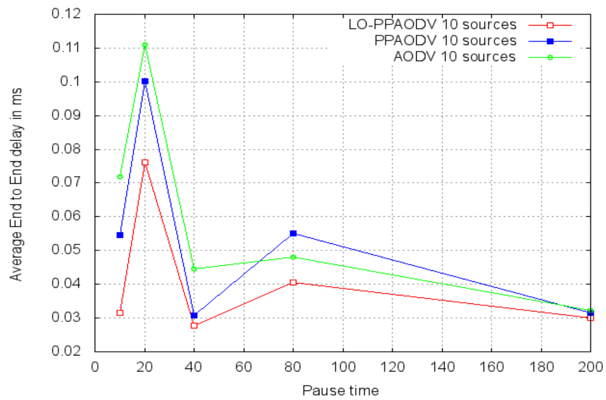


Fig. 9 Average End to end delay 10 sources

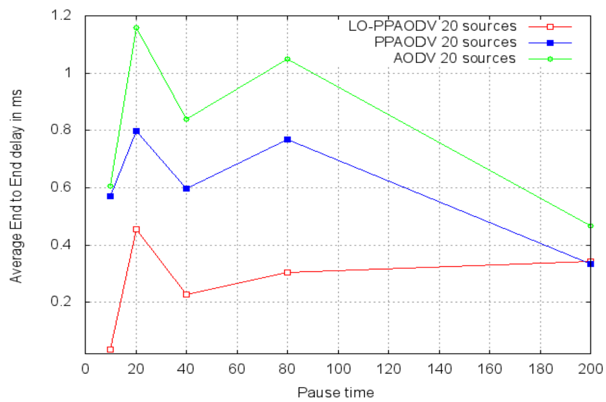


Fig. 10 Average End to end delay 20 sources

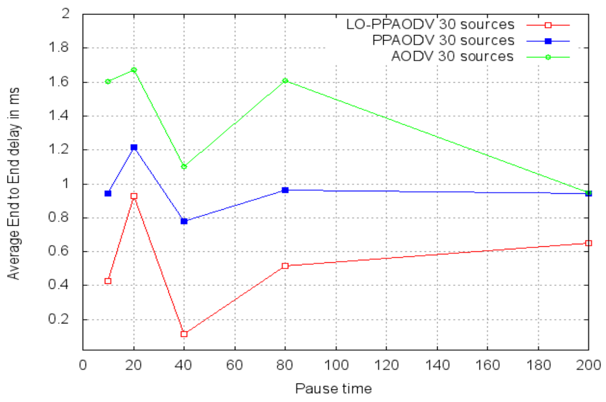


Fig. 11 Average End to end delay 30 sources

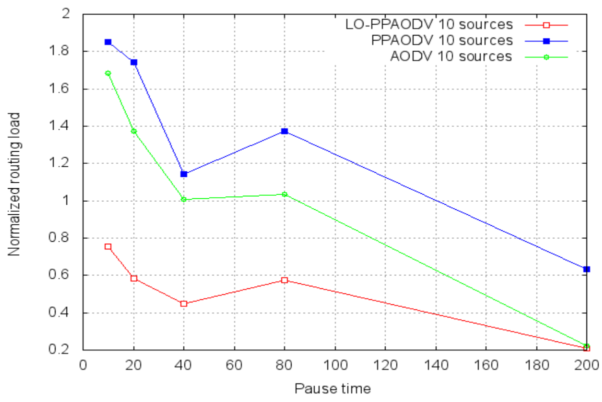


Fig. 14 Normalized routing load 10 sources

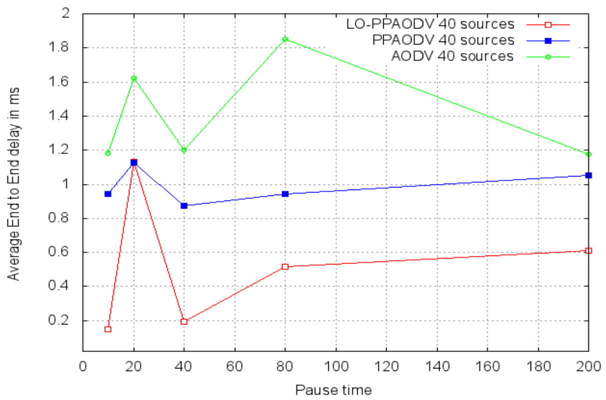


Fig. 12 Average End to end delay 40 source

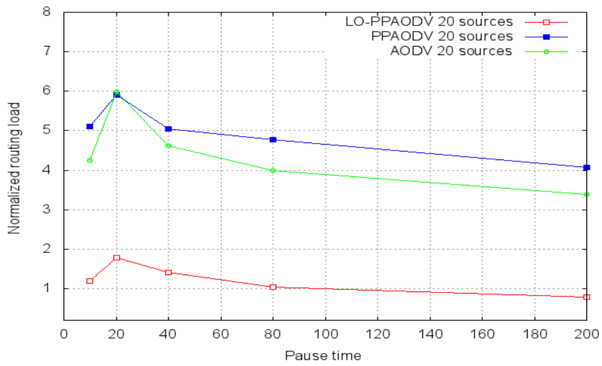


Fig. 15 Normalized routing load 20 sources

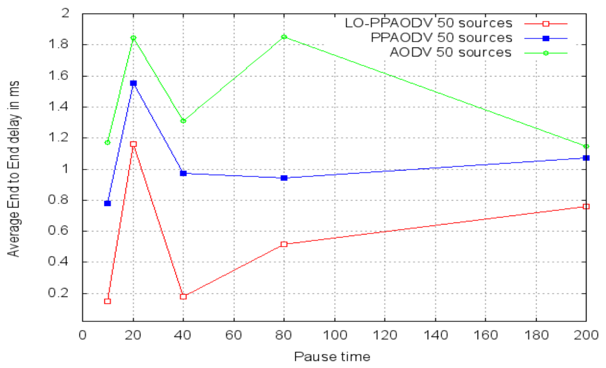


Fig. 13 Average End to end delay 50 sources

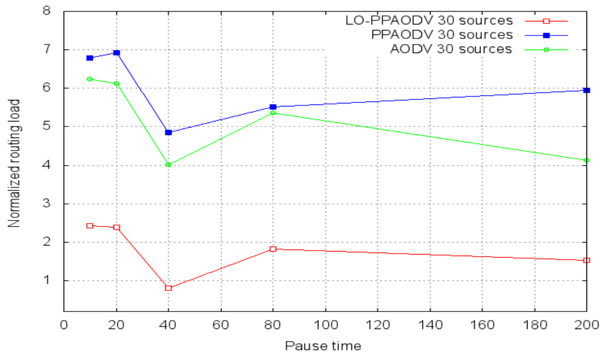


Fig. 16 Normalized routing load 30 sources

In figure 9, figure 10, figure 11, figure 12 and figure 13 the results obtained for the end-to-end delay metric of 10 sources, 20 sources, 30 sources, 40 sources and 50 sources respectively are presented. We observe that the end-to-end delay increases significantly when the number of sources increases. The delay is affected by the route repair procedure because data packets are buffered until an alternative route is found. The results show that LO-PPAODV outperforms PPAODV significantly when the number of sources increase and the motion is low. Figure 12 shows a gain of about 90% of LO-PPAODV over PPAODV, for 40 sources in the pause time 200s.

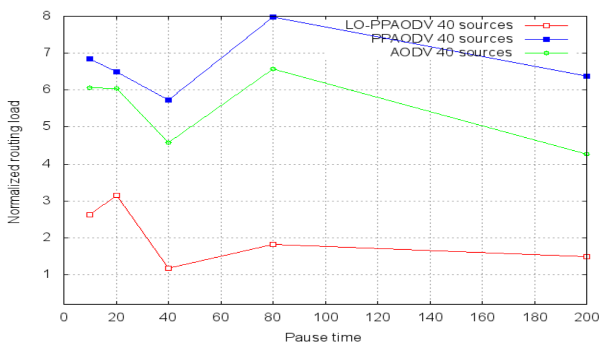


Fig. 17 Normalized routing load 40 sources

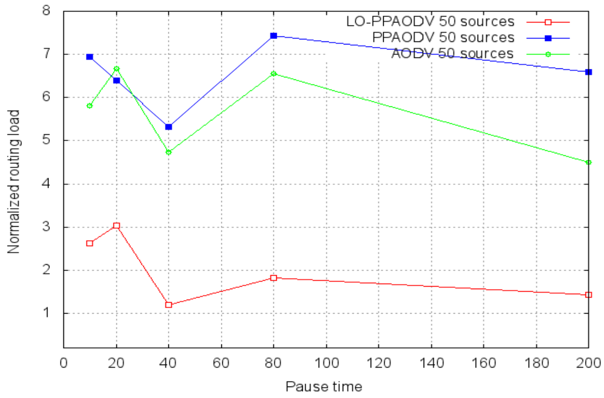


Fig. 18 Normalized routing load 50 sources

Figure 14, figure 15, figure 16, figure 17 and figure 18 show how mobility and number of sources affect the communication overhead. The overhead is high when node motion is high; this is due to the fact that it is difficult to obtain an alternative link to replace a broken one when motion is high. It is also observed that the overhead is low when the number of sources is low. This results from the fact that many sources may share one or more paths, which decreases the communication overhead. It can be observed from figure 16 that the biggest gains of LO-PPAODV over PPAODV is of 300% less and happen with 80s of pause time and 40 sources. This has a good impact on energy because the number of control packets generated is low.

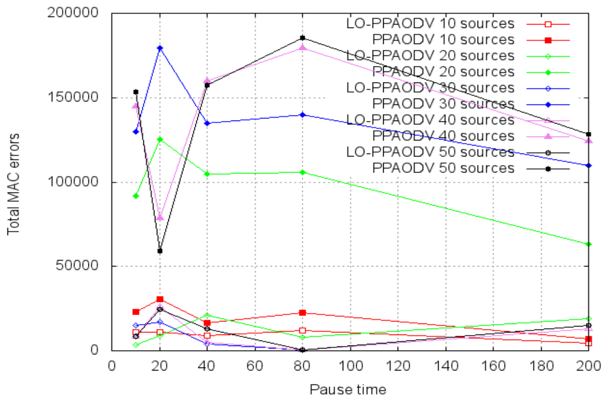


Fig. 19 Total MAC errors

In Figure 19 we observe in the vertical axis the dropped packets from the MAC layer for PPAODV and LO-PPAODV. There were found 4 different types of MAC errors: collision, retry exceed count, MAC busy or duplicate packet. These errors should be minimized. MAC errors are increasing using PPAODV (figure 19), because of the mobility of nodes (from a minimum of 0m/s to a maximum of 10m/s). This mobility, as well, causes a high number of route errors (figure 20) and therefore more routing overhead and packet loss. We see from Figure 19 that LO-PPAODV outperforms PPAODV significantly; MAC errors are decreased for all sources (up to 600% errors less for 50 sources).

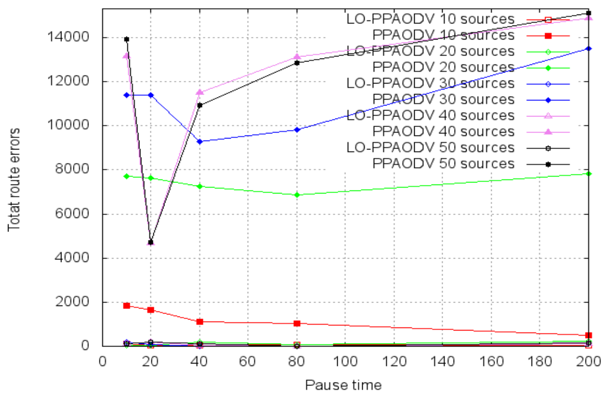


Fig. 20 Total route errors

As conclusion, it is worthwhile to point out that the graphics are related between themselves, since less MAC errors and less route errors provokes lower normalized routing load in the network. As normalized routing load is decreasing, the nodes are able to transmit more data packets; therefore, a higher packed delivery fraction is obtained (up to 300% for 50 sources and happen with 200s of pause time see figure 8).

6. Conclusions

In this paper, we have proposed a Link Quality and MAC-Overhead aware Predictive Preemptive Ad hoc On-Demand Distance Vector (LO-PPAODV). There are two main contributions in this work. One is the protocol is based on new metric combine more routing metrics (Link Quality, MAC Overhead) another is the proposition of a cross-layer networking mechanism to distinguish between both situations, failures due to congestion or mobility; by the usage of the "Route Failure Prediction Technique" based on the Lagrange interpolation for estimating whether an active link is about to fail or will fail.

Simulation results show that the average and to end delay of LO-PPAODV is less than that of PPAODV. Also normalized routing load of LO-PPAODV is smaller than that of PPAODV. It can be noticed from this study that the packet delivery fraction is more than that of PPAODV especially when the number of sources is superior to 10 sources. We can see also that MAC errors and route errors are increasing using LO-PPAODV.

References

[1] E. Belding-Royer and C. Perkins,"Evolution and future directions of the ad hoc on-demand distance-vector routing protocol", Ad Hoc Networks Journal, Vol. 1 No.1, July 2003, pp. 125–150.

[2] C. Perkins, E. Royer, and S. Das, “Ad-hoc On-Demand Distance Vector (AODV) Routing”, Mobile Ad Hoc Networking (MANET), 1999, IETF RFC 3561.

[3] S. Boukli Hacene, A. Lehireche and A. Meddahi, “Predictive Preemptive Ad Hoc On-Demand Distance Vector Routing”, Malaysian Journal of Computer Science, Vol. 19, No 2, December 2006, pp.189-195.

[4] K.Hayato, M.Naoki, H. S. Bouk and Iwao Sasase, “High Precision-Predictive Preemptive Ad hoc On-demand Distance Vector Routing in Ad hoc Networks”, in The 11th International Symposium on Wireless Personal Multimedia Communications (WPMC’08), 2008.

[5] NS-2 Network Simulator, November 2011, available at http://www.isi.edu/nsnam/ns/.

[6] B. Tuch, “Development of WaveLAN, an ISM Band Wireless LAN”, AT&T Technical Journal, July/August 1993, Vol. 4 No. 72, pp. 27–33.

[7] Internet Engineering Task Force (IETF), “Mobile Adhoc Networking (manet) Working Group”, http://www.ietf.org/html.charters/manet-charter.html, 2004.

[8] J. Norman and J.P. Joseph, “HLAODV – A Cross Layer Routing Protocol for Pervasive Heterogeneous Wireless Sensor Networks Based On Location”, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 7, July 2010, pp. 28-37.

[9] M. A. Sayid, and H. Suhaidi, “Divert Failure Route Protocol Based on AODV”. IEEE DOI10.1109/NETAPPS, 2010, pp. 67-71.

[10] C. Xiaoqin, M. H. Jones, and A .D .S. Jayalath, "Congestion-Aware Routing Protocol for Mobile Ad Hoc Networks" in proceedings of IEEE conference on Vehicular Technology, Doi. 10.1109/VETECF.2007.21., 2007, pp.21-25

[11] Y. Ming, A Malvankar, W. Su and S. Y. Foo, "A link availability-based QoS-aware routing protocol for mobile ad hoc sensor networks", Computer communications Archive,vol.30,no.18,2007, pp. 3823-3831.

[12] Y. Yi and S. Shakkottai,"Hop-by-Hop Congestion Control Over a Wireless Multi-Hop Network", in IEEE/ACM on Networking, vol.15, no.1, February 2007.

[13] L. Chen and W. B. Heinzelman," QoS-Aware Routing Based on Bandwidth Estimation for Mobile Ad Hoc Networks", IEEE on Selected Areas in Communications, Vol.23, No. 3, march 2005.

[14] Z. Chenxi and M. S. Corson, "QoS routing for mobile ad hoc networks", in proceedings of Twenty-First IEEE Annual Joint Conference of Computer and Communications Societies, Vol.2, 2002, pp. 958-967.

[15] A. D. Tran and H. Raghavendra," Congestion Adaptive Routing in Mobile Ad Hoc Networks", in IEEE Transactions on Parallel and Distributed Systems, Vol.17, No.11, November 2006.

[16] R. Chandran and Shanmugavel,"Received Signal Strength-based Cross-layer Designs for Mobile Ad Hoc Networks", IETE Technical Review, Vol.25, No.4, August 2008 pp.192-200.

[17] X. Xia, Q. Ren and Q. Liang, "Cross-layer design for mobile ad hoc networks: energy, throughput and delay-aware approach" in proceedings of the IEEE conference on Wireless Communications and Networking, Vol.2, 2006, pp. 770-775.

[18] X. Chen, H. M. Jones and Jayalath, "Congestion-Aware Routing Protocol for Mobile Ad Hoc Networks" IEEE 66th conference on Vehicle Technology, October 2005,pp.21-25

[19] Y. Ming, A. Malvankar, W. Su, and Y. F. Simon, "A link availability-based QoS-aware routing protocol for mobile ad hoc sensor networks", Computer Communications, Vol.30, , 2007,pp.3823–3831.

[20] S. S. baboo and B. Narasimhan, “A Hop-by-Hop Congestion-Aware Routing Protocol for Heterogeneous Mobile Ad-hoc Networks”, (IJCSIS) International Journal of Computer Science and Information Security,Vol. 3, No. 1, 2009.

[21] A. Bisengar, M. Rziza and M. Ouadou, “Queue Length and Mobility aware Routing Protocol for Mobile Ad hoc Network”, International Journal of Communication Networks and Information Security (IJCNIS), Vol. 4, No. 3, December 2012, pp. 207-211.

[22] Y. W. Ho, “A Simple and Effective Cross Layer Networking System for Mobile Ad Hoc Networks”, PIMRC 2002, 2002.

[23] M. Ali cherif, M.K. Feraoun, N. Doumi, and M. Kather “Integration of dynamic current bandwidth capacity calculation for existing AODV”, in Proceding of IEEE ICITeS 2012 Int. Conf. on Information Technology and e-Services, March 2012, pp. 669-675.

[24] M. Sedrati, A. Bilami and M. Benmohamed, “M-AODV: AODV variant to Improve Quality of Service in MANETs”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011, pp.429-436.

[25] A. Dana and M. H. Babaei, “A Fuzzy Based Stable Routing Algorithm for MANET”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011, pp. 367-371.

[27] T. S. Rappaport, “Wireless communications”, principles and practice: Prentice Hall, 1996.

[28] M. Ali cherif, M.K. Feraoun, and S.H. Boukli “Enhanced Predictive Preemptive Multipath Routing Protocol for Mobile Ad hoc Network”, IEEE CommSoft E-Letters, Vol. 2, No. 1, May 2013, pp. 9-12.

Semi-automatic Data Warehouse Design methodologies: a survey

Wafa Tebourski[1],Wahiba Ben Abdessalem Karâa[2] and Henda Ben Ghezala[3]

[1] Computer Science Department, High Institute of Management, University of Tunis, Bouchoucha, Bardo-Tunis, Tunisie

[2] Computer Science Department, High Institute of Management, University of Tunis, Bouchoucha, Bardo-Tunis, Tunisie

[3] Computer Science Department, National School of Computer Sciences, University of Manouba Manouba, Manouba, Tunisie

Abstract

Data warehouses are used in making strategic decisions. A data warehouse is a collection of integrated, historised data originating from heterogeneous sources which gives rise to business bases (data store). Several approaches proposed semi-automatic building data warehouses. In this paper, we present an overview of those works dedicated to the design of data warehouses and a comparison of these different approaches.

Keywords: *Data warehouses, multidimensional modeling, Bottom-up approach, Top-down approach, mixed approach.*

1. Introduction

Faced to the large amounts of data and the remarkable diversification of their sources, a scientific and economic interest to explore the reservoirs of knowledge is established. Hence, we use a process of decision support where users seek models of information interpretation, hidden knowledge extraction and potentially useful information from the available data to improve product and services quality for the company's strategic differentiation.

Systems for Decision Support (DSS) are flexible and interactive information systems that help decision makers to extract useful information, to identify and solve problems and make decisions [22]. The DSS processes the information from different sources in one place, the information are consistent and familiar to the user. The DSS combines and standardize databases, allowing analysis and decision making.

Among the decision support systems, data warehouse systems are possibly the most used in the world. Traditional information model systems are not able to analyze complex data on a large number of areas such as complex calculations, aggregations ... etc. Thus, multidimensional modeling was proposed. The multidimensional model aims at presenting the data in a structured and intuitive way to solve the transactional models' difficulties and meet the decision makers' needs. The multidimensional model is based on two fundamental concepts: fact and dimension.

Many researchers have focused on the design of data warehouse schemas. This design is a complex task. Several works have been proposed. Our purpose is to present a comparative study between these different approaches, based on several criteria.

This article is organized as follows: In Section 2, we present the different design approaches of data warehouses. In Section 3, we present the different research works related to multidimensional modeling at the conceptual and logical level. In Section 4, we propose a comparative study between the different models.

2. Data warehouses design approaches

2.1 Sources based approaches (ascending/Bottom-up)

Approaches directed by sources perform the extraction of data from heterogeneous sources. These approaches integrate the data into a storage space accessible by all decision makers. The design of data warehouse is based on a detailed analysis of data models, generally the entity-relationship model (E/R). Such approaches facilitate the ETL (Extraction-Transformation-Load) processes since each entity and relation in the source model will be presented by multidimensional concepts.

Various studies have been conducted in this context such as [6], [14], [16].

2.1.1 Golfarelli et al's approach

The authors [14] suggest a formal model: Dimensional Fact Model (DFM) which is a multidimensional graphical model, clearly differencing concepts such as facts, dimensions, measures and hierarchies. This model is presented as a set of tree structured patterns of events.

In this approach, the derivation of a pattern is performed using a two-steps process:

(i) The first step consists in establishing facts.

(ii) The second step ensures the construction of an tree attribute; it removes irrelevant attributes from the tree, identify and analyze related dimensions, measures and hierarchies for each fact.

2.1.2 The approach of Hüsemann et al.

This approach [6] is used for star logical modeling. Such study is conducting using a four-steps process:
The analysis and specification requirements: the experts select the relevant attributes of the model E / A source and specify their use (measure of fact or dimension identifier). Additional requirements are added using complex derived measures;

(i) The conceptual modeling: at this stage, a transformation of the semi-formal specification requirements of the multidimensional conceptual schema is accomplished;

(ii) The logical modeling: This step converts conceptual schemas into logical patterns respecting the logic model referred (usually relational or multidimensional) via transformation rules;

(iii) The Physical modeling: This step performs a physical implementation of logic diagrams.

2.1.3 The approach of Romero et al.

The goal of this approach [16] is to identify the multidimensional concepts from domain ontology. This approach is based on four criteria allowing exploration of the multidimensional concepts. These latter's are: (i) the multidimensional model, (ii) the constraint of multidimensional space arrangement, (iii) the integrity of constraint base, and (iv) of the additivity constraint.

2.2 Requirements based approaches (descendants / Top-down)

This type of approaches presents the phases of requirement specification and derivation of conceptual schemas. These approaches attempt to reduce the risk of failure of the decisional information system.
In this context several researches have been directed towards the design of data warehouses based on requirements such as [12], [13], [17] and [21]

2.2.1 Kimball's approach

This approach [21] is a requirements based approach aiming to derive a logical design of the data warehouse. This is an informal method, which introduces a detailed multidimensional concept to give rise to multidimensional schemas guide.
The proposed method has two axes:
The bus architecture: aims to identify all the data marts which the designer seeks to build. Data marts are defined as a pragmatic set of related facts. The next step consists in classifying the different dimensions of each data mart. Hence, an ad hoc matrix is constructed to capture the multidimensional requirements and point out associations between different data marts.
The cycle of multidimensional life is driven by a five steps process: (i) project planning, (ii) business requirements definition, (iii) choice of technology, (iv) data modeling, and (v) specification and development of an application.

2.2.2 The approach of Cabbibo and Torlone

The requirement based approach [13] presents the design methodology of the most cited. This

approach allows the generating of a logic diagram of ER (Entity Relationship or n-ary) diagrams. In addition, it can produce multidimensional schemas in terms of relational databases or multidimensional matrix. This is an informal method performs depth analysis of data sources, but does not provide formal rules.
However, this method introduces the basic foundations that will be used later in the literature, put in place the foundations used by the rest of methodologies. The proposed method consists of four steps:

(i) The first and the second stage allow restructuring facts and dimensions identification as well as the ER diagram.
(ii) The third and the fourth steps provide the multidimensional diagram.

2.2.3 The approach of Mazôn et al.

This proposal is requirement-based approach. The authors [12] integrates the business objectives of the company in the specification of the requirements using i* technology. This approach is based on three steps:

(i) Definition of business goals: consists in specifying the main objectives of a company. These goals can be classified into three levels of abstraction: strategic, decision-making and informational.

(ii) Modeling requirements using i* technology: This step identifies the data warehouse users, the business goals of the organization and the relationship between these two components.

(iii) Transformation of the obtained i* model into a multidimensional model: using heuristics [23].

2.2.4 The approach of Giorgini et al.

This approach [17] begins with the specification requirements phase and carried out using two organizational model and decision-making model. After that, a construction step is performed to provide the conceptual model which is derived from the relational model in the decision-making perspective and subsequently refined using the hierarchies of the organizational model diagram.

2.3 Mixed Approaches

This type of approach incorporates both bottom-up and top-down approaches in an attempt to take advantage of their benefits. Some researches have focused on mixed approaches, such as the [3], [4] and [18].

2.3.1 The approach of Bonifati et al.

Bonifati et al proposed a semi-automatic approach [3] based on both of requirements and sources, called as mixed approach. This method consists in three phases:

(i) Bottom-up analysis: This step examines the E/R model of the data source to construct star schemas candidates using patterns. We note that the bottom-up analysis can generate a large number of candidate patterns. An algorithm transforms each N-M association in n one to many through dealing with E/R model as a graph.
(ii) The top-down analysis: This step collects the analyzed requirements that will be refined and aggregated in a tabular report of abstraction. This step outputs the star schemas ideals.
(iii) Integration: This step makes the perfect match for each pattern from the top-down analysis with all the candidates produced by the bottom-up analysis to meet the requirements of decision diagrams.

2.3.2 The approach of Nabli et al.

The approach proposed by [4] is a mixed approach of using the automated design of data marts and data warehouse starting from semi-structured OLAP requirements expressed in a tabular form.
This method contains three steps: (i) The acquisition OLAP requirements, assisted through ontology, (ii) the generation of patterns of data marts and (iii) the generation of the warehouse schema.

2.3.3 The approach of Giorgini et al.

Giorgini et al introduced a hybrid approach consisting in three phases: (i) requirements analysis, (ii) matching requirements with sources and (iii) refinement [18]:

(i) The requirement analysis phase: this step generates a decision model and organizational model.
(ii) Matching requirements with sources: in this step, the decision model is mapped to a data source E/R through jointing on organizational model.
(iii) Refinement: The multidimensional model is enriched through the construction of hierarchies and their refinement.

3 Multidimensional modeling

Multidimensional modeling has two aspects: (i) conceptual aspect aims to make realistic modeling, (ii) the logical aspect presenting this reality.

3.1 Conceptual modeling

Several works dedicated to conceptual modeling of data warehouses and data marts are proposed. Three models are established in this context: (i) models based on the extension of the model Entity/ Relationships, (ii) models based on the UML extension, and (iii) personalized models.

3.1.1 Extension of the Model E/R

For modeling data warehouses, several extensions of the model E/R (entity-relationship) have been proposed. We discuss, such as, Starer [15], ME / R [7] and ERA [9]. The fact is a new modeled concept in E/R representing real data having the same properties. Indeed, the entity concepts present the different hierarchical levels of dimensions. However, the relations modelize associations between entities or entering facts and entities.

3.1.2 UML extension

Several methods for modeling data warehouses have been proposed in the literature, these methods are based on particular object paradigm, using UML: [1], [2], [23], [26]. In [23], [24], the authors propose a multidimensional object-oriented model (GOLD) for integrating UML in the multidimensional modeling.
They used a set of value-typed stereotypes. The fact is modeled using a class stereotyped "FACT" containing measures (atomic or calculated attributes). The dimension is represented using a class stereotyped "DIMENDION", hierarchies are represented by classes with the stereotype "BASE". The relations between two levels of a dimension in a hierarchy are modeled using the stereotype "ROLLUP UP."

3.1.3 Ad hoc or Personalized model

The Ad hoc models represent another class of modeling methods in the literature [5], [6], [8], [13], [14], [20] .The Dimensional Fact Model (DFM) proposed by [14] and extended by Rizzi [25], consists on a set of facts diagrams. The studies works on this category of approaches offer a range of constructors based on the following concepts: facts, measure, dimension, hierarchy, descriptive attribute (low attribute), multiple arcs, shared hierarchies and parameters.

These different families of proposed conceptual models are based on several paradigms (entity-relationship, object ...etc). These models mainly focus on the multidimensional data modeling.

3.2 Logic modeling

At the logic level several presentations, are feasible of multidimensional modeling. The relational schema, suitable for decision-making, is presented in a star schema shaped, in snowflake or constellation:

a. Star Schema: it is a multidimensional representation of data given data [21] describing the fact placed in the center dimensions surrounding it. Every fact is a table called fact table, consisting of a set of attributes representing activity measures and foreign keys that reference the dimensions. Each dimension corresponds to a table, called dimension table containing attributes (strong or weak) and primary key to ensure the joints with the fact table. In this context, we are talking about a star logic model.

b. Snowflake schema: it consists in building a separate table for each hierarchical level of a dimension. In this model, the joints are numerous; the fact table includes a foreign key at the hierarchy level of each dimension.

c. Constellation schema: it is a collection of star schemas that divide the common dimensions.

4 Comparative study

The comparison between these three approaches seems essential, we rely on seven criteria classified into four categories such as: (i) the inputs occur on different kinds of data sources, (ii) the outputs articulating the goals which is the data warehouses, or the data marts creation, (iii) design focuses on logical or conceptual, formal and informal modeling of data representation, (iv) methods engineering requirements summarizing the specification requirements.

Figure 1 shows a comparison between the different approaches of data warehouses design based on these criteria.

From the level of design perspectives [6], [14] and [16] use logical schema. However, [12], [13], [17], [18] and [21] introduce conceptual schema.

The majority of these approaches aim to create the data warehouse as the work [6], [12], [13], [14], [16], [17] and [18]. However, the rest leads to create data marts such work [4] and [21].

All those approaches are formal methods except the approach of [13] and [21] where an informal modeling is introduced.

The design of data warehouse is based on several data sources namely the relational schema such as the works of [6], [12], [13], [14], [17], [18] and [21].Other methods [4] and [16] use ontology as data sources.

Given the complexity of the requirements specification, some research have developed their own techniques such as i * technique proposed by Mazon et al [12], the TROPOS method Giorgini [17], [18] and the GQM method (Goals / Question / Metric) [3].

Approaches / Criteria		*Approaches directed by sources*			*Approaches directed by requirements*				*Mixed approaches*		
		Golfarelli et al.[14]	Hüsemann et al.[6]	Romero et al.[16]	Kimball et al[21]	Cabibbo et al.[13]	Mazón et al.[12]	Giorgini et al.[17]	Bonifati et al.[3]	(Nabli et al.[4]	Giorgini et al.[18]
level	Conceptual schema	.	+	+	.	.	+	+	.	+	+
	Logic schema	+	.	.	+	+	.	.	+	.	.
Goals	Data warehouse	+	+	+	.	+	+	+	+	.	+
	Data marts	.	.	.	+	.	.	.	.	+	.
Modeling	Formal	+	+	+	.	.	+	+	+	+	+
	Informal	.	.	.	+	+	.	.	.	.	.
Type of data sources	ER diagram	.	.	.	.	+	.	.	.	.	.
	Ontology	.	.	+	.	.	.	.	.	+	.
	Relational schema	+	+	.	+	.	+	+	+	.	+
Conceptual representation	Others	.	.	.	Adopting a	Adopting a	.	DFM	.	.	.
	UML Extension	.	.	+	.		+	.	.	+	+
	Personalized	+	+	.	.		.	.	+	.	.
Method used for Requirements Specification		.	.	.	.	.	Technique i*	TROPOS	GQM (Goals /Question/	.	TROPOS

Fig. 1: comparison between the different approaches to data warehouse design

At the sight of this comparative analysis, we can infer that led sources based approaches are useful if the schema of the data source is simple and available. In this category, they suffer generally from requirements engineering patterns. In contrast, the requirements based approaches, focus on the requirements specification which are frequently variable and limitedly expressed.

Thus, the design of data warehouses cannot be exclusively based on data sources or requirements. Indeed, we find that both ascending and descending approaches are complementary and can be mixed together for better results, being the subject of the third approach called hybrid approaches.

4 Conclusion

In this paper, we have presented various research approaches data warehouses design. These studies were classified according to three trends: sources based approaches; requirements based approaches and mixed approaches. Our comparative study discusses various works of data warehouses design.

In the future, we propose to study the problem of modeling data warehouses and we introduce a new method.

References

[1] A.Abello, J.Samos, and F.Saltor. YAM2 (Yet another multidimensional model),"An extension of UML". In Proceedings of the International Database Engineering & Applications Symposium. Edmonton, Canada. 2002, July 17-19, (pp. 172-181).

[2] A.Abello, J.Samos, and F.Saltor. YAM2,"A multidimensional conceptual model extending UML". Information System,2006, (pp.541-567).

[3] A.Bonifati, F.Cattaneo, S.Ceri, A.Fuggetta, and S.Paraboschi,"Designing data marts for data warehouses". ACM Trans. Softw.Eng. Methodol, 2001, (pp.452–483).

[4] A.Nabli, J.Feki, F.Gargouri, "Automatic Construction of Multidimensional Schema from OLAP Requirements", Arab International Conference on Computer Systems and Applications (AICCSA'05), Reference proceedings 0-7803-7983-7/03 © 2005 IEEE, Cairo, Egypt, January 2005.

[5] A.Tsois, N.Karayannidis, and T.Sellis, "Conceptual Data Modeling for OLAP". In Proc. DMDW. Interlaken, Switzerland, 2001.

[6] B.Hüsemann, J.Lechtenbörger, G.Vossen, "Conceptual Data Warehouse Design". Design and Management of Data Warehouses, Sweden, 2000.

[7] C.Sapia, M.Blaschka, G.Hofling, and B.Dinter, "Extending the E/R model for the multidimensional paradigm". Lecture Notes in Computer Science, 1552, 1999,(pp. 105-116).

[8] D.L.Moody, and MA.R.Kortink, "From Enterprise Models to Dimensional Models: a Methodology for Data Warehouse and Data Mart Design". In Design and Management of Data Warehouses, 2000, (pp. 5).

[9] E.Franconi, U.Sattler, "A DataWarehouse Conceptual DataModel for Multidimensional Aggregation". In Proceedings of the Intl.Workshop on Design andManagement of DataWarehouses (DMDW1999), Technical University of Aachen (RWTH), 1999, (pp. 13–1–13–10).

[10] F.Liu, C.Yu, and W.Meng, "Personalized Web Search For Improving Retrieval Effectiveness", IEEE Transactions on Knowledge and Data Engineering, vol. 16, n1,2004, (pp. 28-40).

[11] J.P.Mc.Gowan, "A multiple model approach to personalized information access", Master Thesis in Computer Science, Faculty of science, University College Dublin, 2003.

[12] J.Mazón, J.Trujillo, M.Serrano , and M.Piattini, "Designing data warehouses: from business requirement analysis to multidimensional modeling", in: K. Cox, E. Dubois, Y. Pigneur, S.J. Bleistein, J. Verner, A.M. Davis, R. Wieringa (Eds.), REBNITA Requirements Engineering for Business Needs and IT Alignment, University of New South Wales Press, 2005.

[13] L.Cabibbo and R.Torlone, "A Logical Approach to Multidimensional Databases". In VIth International Conference on Extending Database Technology (EDBT 98), Valencia, Spain, volume 1377 of LNCS, Springer, 1998,(pp. 183–197).

[14] M.Golfarelli, D.Maio, and S.Rizzi, "The dimensional fact model: conceptual model for data warehouses", international Journal of Cooperative Information Systems 7, 1998.

[15] N.Tryfona, F.Busborg, and J.Christiansen, "A conceptual model for data warehouse design". In Proc. of ACM Second International Workshop on Data Warehousing and OLAP (DOAP'99), Kansas City, Missouri, USA, November,1999, (pp. 3–8).

[12] O.Romero, and A.Abelló, "Automating Multidimensional Design from Ontologies". DOLAP'07, Lisboa, Portugal, November 9, 2007.

[17] P.Giorgini, S.Rizzi, and M.Garzetti, "Goal-oriented Requirement Analysis for Data Warehouse Design". In Proc. of 8th Int. Workshop on Data Warehousing and OLAP, ACM Press, 2005, (pp. 47–56).

[18] P.Giorgini, S.Rizzi, and M.Garzetti, "A Goal-Oriented Approach to Requirement Analysis in Data Warehouses". In Decision Support Systems (DSS) jounal, Elsevier, (pp. 4-21), Vol 45 Issue 1, 2008.

[19] P.L.Tchienehom, " Modèle générique de profils pour la personnalisation de l'accès à l'information", In INFORSID, 2005,(pp. 269–284),.

[20] P.Vassiliadis, A.Simitsis, and S.Skiadopoulos, "Conceptual modeling for etl processes", In Theodoratos, 2002,(pp. 14–21).

[21] R.Kimball," The Data Warehouse Toolkit", John Wiley and Sons, Inc., New York,1996.

[22] S.Alter,"Decision Support System": Current Practices and Continuing Challenges. Massachusetts: Addison-Wesley Publishing Co., 1980, (pp.316).

[23] S.Luján-Mora, J.Trujillo, I.Y.Song, "Extending the UML for multidimensional modeling". In Proceedings of the International Conference on the Unified Modeling Language. Dresden, Germany, 2002,(pp. 290-304).

[24] S.Luján -Mora, J.Trujillo, and I.Y.Song, "A UML Profile for Multidimensional Modeling in Data Warehouses". Data and Knowledge Engineering, 2006, (pp.725–769).

[25] S.Rizzi, "Conceptual Modeling Solutions for the Data Warehouse. Data Warehouses and OLAP: Concepts, Architectures and Solutions", edited by R. Wrembel and C. Koncilia, 2007.

[26] T.B.Nguyen, A.M.Tjoa, and R.Wagner, "An object-oriented multidimensional data model for OLAP". In Proceedings of the International Conference on Web-Age Information Management Shanghai, China, 2000, (pp. 69-82).

OBJECT RECOGNITION USING PARTICLE SWARM OPTIMIZATION AND GENETIC ALGORITHM

Mahmood ul Hassan [1], M. Sarfraz [2], Abdelrahman osman [3] and Muteb Alruwaili [4]

**[1, 3, 4] Department of Computer Science, College of Science and Arts. Al Jouf University
Tabarjal, Jouf, Saudi Arabia
[2] Department of Information Science, Kuwait University
P.O. Box 5969, Safat 13060, Kuwait**

Abstract

Object recognition is an important research field of computer vision and has its application in a broad range of problems including image retrieval, compression, surveillance and medical diagnostics. The main goal of the object recognition problem is to recognize the objects of the same type even when they are viewed from different viewpoints. This goal, however, remains a challenge for computer vision to recognize objects having invariant features such as translation, rotation and scaling.

Shape descriptors like Fourier and Moments are invariant with respect to transformation, rotation and scaling. Particle swarm optimization (PSO) is a population based soft computing technique. Particle Swarm Optimization technique shares numerous similarities with evolutionary computation techniques such as Genetic Algorithms (GA).

One of the most important tasks regarding to object recognition is how to find number of descriptors of a given object. The query that arises is what is the optimum number of descriptors to be used with maximum recognition rate? , Are descriptors having equal importance? Such reasons signify the importance of these descriptors and also selecting the best descriptor by applying optimization technique.

We have introduced an evolutionary optimization technique known as Genetic algorithm (GA) for solving the optimization problem. GA assigns, for each of these descriptors, a weighting factor that reflects the relative importance of that descriptor.

Keywords: *Object recognition, Fourier descriptors, Genetic algorithm, ORGA (Object Recognition using Genetic Algorithm), PSO.*

1. Introduction

Object recognition in computer vision is the task of finding a given object in an image or video sequence. Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different viewpoints, in many different sizes / scale or even when they are translated or rotated.

There are various disciplines of everyday life, including security, health, post, defense, surveillance, etc., where the issue of object recognition needs to be tackled fastly and accurately. Shape descriptors [1 - 6] like Fourier can be classified by their invariance with respect to the transformations allowed in the associated shape definition. These descriptors are invariant with respect to congruency, meaning that congruent shapes (shapes that could be translated, rotated) will have the same descriptor. These descriptors describe the features of an object to uniquely represent the shape of an object. In this paper we are using Genetic Algorithm technique on Fourier Descriptors then we will compare the results with PSO [6] that have been frequently used as features for image processing, remote sensing, shape recognition and classification. These Descriptors can provide characteristics of an object that uniquely represent its shape.

This paper has used Fourier descriptors, with different combinations, for the recognition of objects captured by an imaging system which may transform, make noise or can have occlusion in the images. An extensive experimental study, similar to the moment invariants [7], has been made using various similarity measures in the process of recognition. These measures include Euclidean Measure, Percentage error, Log of Euclidean and Log of square of Euclidean. Comparative study of various cases has provided very interesting observations which may be quite useful for the researchers as well as practitioners working for imaging and computer vision problem solving. Although the whole study has been made for bitmap images, but it can be easily extended to gray level images.

From the analysis and results using Fourier Descriptors, the following questions arise: What is the optimum number of descriptors to be used? Are these descriptors of equal importance? To answer these questions, the problem of selecting the best descriptors has been formulated as an optimization problem. Genetic algorithm (GA) technique has been mapped and used successfully to have an object recognition system using minimal number of Fourier Descriptors. The goal of the proposed optimization technique is to select the most helpful descriptors that will maximize the recognition rate. The proposed method will assign, for each of these descriptors, a weighting factor that reflects the relative importance of that descriptor.

The outline of the remainder of the paper is as follows. Getting of bitmap images and their outline is discussed in Sections 2 and 3 respectively. The concepts of similarity measures are explained in Section 4. Fourier theory is explained in Section 5. Section 6 explains the detail of PSO. Detail of proposed technique is explained in Section 7. Proposed algorithm is explained in section 8. Detailed experimental study and analyses are made in Section 9 whereas Section 10 deals with interesting observations during the experimental study. Finally, Section 11 concludes the paper as well as touches some future work.

2. Getting Bitmap Image

Bitmap image of a character can be obtained by creating a bitmap character on some program like *Paint* or Adobe Photoshop. Alternatively an image drawn on paper can scan and store it as bitmap. We used both methods. The quality of bitmap image obtained directly from electronic device depends on the resolution of device, type of image (e.g. bmp, jpeg, tiff etc), number of bits selected to store the image etc. The quality of scanned image depends on factors such as quality of image on paper, scanner and attributes set during scanning.

3. Finding Boundary

In order to find boundary of bitmap image, first its chain code is extracted [8, 9]. Chain codes are a notation for recording the list of edge points along a contour. The chain code specifies the direction of a contour at each edge in the edge. From chain coded curve, boundary of the image is found [10]. The selection of Boundary Points is base on their corner strength and contour fluctuations. The input to our boundary detection algorithm is a bitmap image

Fig. 1 (a) Bitmap image (b) outline of the image

Figure 1(a) shows the bitmap image of a character. Figure 1(b) shows detected boundary of the image of Figure 1(a).

4. Similarity Measures

We implement four different simple classifiers that calculate different similarity measures of the corresponding Fourier descriptors of the input shape and each of the shapes contained in the database. The similarity measures, attempted for experimental studies, are Euclidean Distance (ED), Log of Euclidean Distance (LED), Log of Square of Euclidean Distance (LSED) and Percentage Error (PE).

1. $$\sqrt{\sum_{i=1}^{n}\left(a(i)-b(i)\right)^2}$$ (Euclidean Distance (ED)) (1)

2. $$\sum_{i=1}^{n}\left|\frac{a(i)}{b(i)}\right|$$ (Percentage Error (PE)) (2)

3. $$LED = \log\sqrt{\sum_{i=1}^{n}(a(i)-b(i))^2}$$ Log of Euclidean Distance (LED) (3)

4. $$LSED = \log\left[\sum_{i=1}^{n}(a(i)-b(i))^2\right]$$ Log of Square of Euclidean Distance (LSED) (4)

In this study, n is the number of FDs considered, a (i) is the ith FD of the template image, and b (i) is the ith FD of the test image. A tolerable threshold ρ is selected to decide a test object recognized. This threshold is checked against the least value of the selected similarity measure.

5. Fourier Theory

To characterize objects we use features that remain invariant to translation, rotation and small modification of the object's aspect. The invariant Fourier descriptors of the boundary [11-13] of the object can be used to identify an input shape, independent on the position or size of the shape in the image.

Fourier transform theory has played a major role in image processing for many years. It is a commonly used tool in all types of signal processing and is defined both for one and two-dimensional functions. In the scope of this research, the Fourier transform technique is used for shape description in the form of Fourier descriptors. The Fourier descriptor is a widely used all-purpose shape description and recognition technique. The shape descriptors generated from the Fourier coefficients numerically describe shapes and are normalized to make them independent of translation, scale and rotation. These Fourier descriptor values produced by the Fourier transformation of a given image represent the shape of the object in the frequency domain. The lower frequency descriptors store the general information of the shape and the higher frequency the smaller details. Therefore, the lower frequency components of the Fourier descriptors define a rough shape of the original object

The Fourier transform theory can be applied in different ways for shape description. One method works on the change in orientation angle as the shape outline is traversed. But in our research the following procedure was implemented, the boundary of the image is treated as lying in the complex plane. So the row and column co-ordinates of each point on the boundary can be expressed as a complex number.

The advantage of using the Fourier transform is in its invariant properties. Rotating the object merely causes a phase change to occur, and the same phase change is caused to all the components.

The simple geometric transformations of the Fourier transform are as follows:

- Translation: $u(n) + t \rightarrow a(k) + t\delta(k)$
- Rotation : $u(n)ej\theta \rightarrow a(k)ej\theta$
- Scaling: $su(n) \rightarrow sa(k)$
- Starting point: $u\ (n\text{-}t) \rightarrow a\ (k)\ ej2\prod tk/N.$

6. Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a population based soft computing technique developed by Dr. Eberhart and Dr. Kennedy in 1995. Particle Swarm Optimization technique shares numerous similarities with evolutionary computation techniques such as Genetic Algorithms (GA). PSO starts initialization with a population of random solutions and searches for optimum solution by updating generations. However, if we compare PSO with GA then unlike GA, PSO has no operators such as crossover and mutation. In PSO, the possible solutions, called particles, move through the problem space by following the current optimum particles.

In PSO initial steps for recognition of an object is to check the noise from the image. If it found the noise then remove it from the image with the help of noise removing algorithm.

After cleaning the noise, boundary is calculated with the help of chain code [14]. Now apply fast Fourier transformation on the boundary points to get descriptors. Now from here PSO assigns weight to each descriptor.

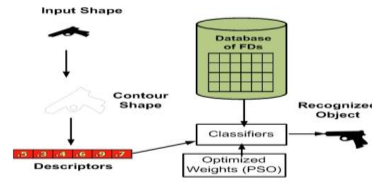


Fig. 2 General architecture of PSO

PSO starts working with a population of random solutions. Each particle move towards its fittest solution it has achieved so far. As PSO has its memory so it stores the value of fitness. PSO has also a global best value which is known as gbest from where the overall best value is stored.

At each step PSO changes the velocity of each particle in the direction of pbest and gbest.

This soft computing technique has numerous advantages over traditional techniques. It starts working with population of solutions from where it selects the best one while traditional does not work in similar way. They start searching with holding single solution. So in this regard PSO produced much better results as compare to traditional ones.

7. Optimization of the Feature Vector using Genetic algorithm

From the previous analysis and results, the following questions arise: What is the optimum number of descriptors to be used? Are these descriptors of equal importance? To answer these questions, the problem of selecting the best descriptors can be formulated as an optimization problem. The goal of the optimization is to select the most helpful descriptors that will maximize the recognition rate and to assign for each of these descriptors a weighting factor that reflects the relative importance of that descriptor.

7.1 Objective Function

Since the problem of selecting the best descriptors can be formulated as an optimization problem, one need to define an objective function. The objective function, in this case, is made up of the following two terms:

- The recognition rate,
- The number of useful descriptors.

In other words, it is required to maximize the recognition rate using the minimum number of descriptors. The function of the optimization algorithm is to assign a weight w_i for every descriptor so that the objective function is optimized, where w_i belongs to [0,1].

The mathematical formulation of the objective function is defined as follows:

$$J = -H + \alpha * \min(PE), \qquad (5)$$

Where

- H is the number of hits (the number of correct matches), and
- PE is the percentage of errors of all the training images for a given set of weights, and
- α is a factor that is adjusted according to the relative importance of the min(PE) term. In most of the simulations, $\alpha = 0.7$ is experienced as best case most of the time

The first term, the objective function Eq. (5), makes GA search for the weights that result in the highest recognition rate and the second term makes the GA reach to the highest recognition with the minimum number of descriptors.

7.2 Genetic Algorithm

GA was introduced first in 1960 by John Holland [15]. GA is also known as evolutionary algorithm (EA). Evolutionary algorithms are general-purpose stochastic methods simulating natural selection and evolution in the biological world. GA differ from other optimization methods, such as PSO, Simulated Annealing, in the fact that GA maintain a population of potential solutions to the problem, and not just one solution.

Generally, GA works as follows: a population of individual is initialized where each individual represents a potential solution to the problem. The quality of each solution is evaluated using a fitness function. A selection process is applied during each iteration of GA in order to form a new population. The selection process is biased towards the fitter individuals to insure that they will be a part of new population. Individuals are altered using two main operators of GA which are crossover and mutation. This procedure is repeated until the potential solution is reached. The best solution found is expected to be a near optimum solution [Michalewicz 1996].

7.3 Steps of the GA

Genetic algorithm work as follows: at first population is created randomly, which satisfies the environmental constraint of problem. There are four steps in the Genetic algorithm. The first step is initialization. Other three steps "evaluation of fitness function, crossover and mutation" are inside a main loop. The execution of these steps is in a sequence until the fittest population arrives at a maximum value or no conspicuous improvement is to be observed to the fitness after certain iterations.

1. Initialization: - The first step for genetic algorithm in the optimization process is initialization. In this step various parameters are initialized to their desired value. In current simulation we have set the following parameters.

- Bias is set between 0 and 0.2.
- Number of iterations are 25 but we have also set the stopping criteria. If the stopping criteria meet during the specified iterations the simulation ends otherwise it goes for the number of iterations specified.
- Numbers of trials are used in between 10 to 15.
- Stopping criteria is also set which depends on the hits used in the simulation.

2. Evaluate fitness function:- The second step of Genetic algorithm is to calculate the fitness of each individual. For this purpose we assign weights to all set of individuals, so we calculate fitness against each individual. This Fitness represents the importance of attached weights, and is given by:

$$f = 1 - \min(PE)$$

Where f represents the fitness value and PE is the percentage error of all the training objects for a given set of weights.

3. Apply crossover: - Crossover is the most important operator of GA. Initially crossover is applied on randomly selected population (weights).

4. Apply mutation: - If stopping criteria matched then terminate else apply mutation on newly created population and check for hits.

If current hit ratio is better than previous, population will be replaced with newly generated population, otherwise step (vi) will be repeated until best hit ratio is achieved.
Now again check here for stopping criteria, if found then terminate otherwise start searching from the start.

8. Proposed Algorithm

Initialization:

Set stopping criteria

Set no of iterations (Stopping criteria)

Set no of trials

Set threshold to compute goodness

1. Load database of descriptors.
2. Initialize an array of particles with random values

as initial weights.

3. Evaluate goodness & check the stopping criteria.
4. If stopping criteria does not meet then apply crossover on the weights & check for hits. If current hits are better than previous, replace new population with previous and store the hits.
5. Repeat for all Childs after crossover.
6. Check the stopping criteria.
7. Apply mutation on each particle & check for hits. If current hits are better than previous then replace.
8. Check for stopping criteria.
9. If best hit meet stopping criteria then stop Else
10. Repeat step 3 to 9
11. If stopping criteria doesn't meet for number of trials then go to step2.

End

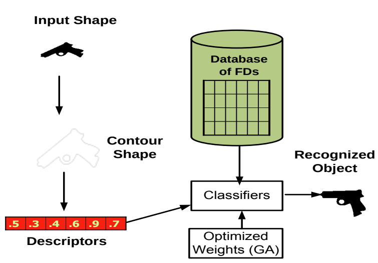


Fig. 3 Pictorial description of the proposed technique

Figure 3 shows the complete description of the proposed technique.

9. Results and Analysis

In this section we will compare proposed work with one of the most important soft computing technique known as Particle Swarm Optimization (PSO).

The genetic algorithm can be found in the current literature at many places.

The proposed GA–based approach was implemented using MATLAB 7.1.0 (simulator) for performance evaluation of proposed technique. In our implementation, the Bias is set between 0-0.2, number of iterations are 25, and the numbers of trials are used in between 10 to 15. The search process stops either desired hits are achieved or it exceeds from the specified number of iterations.

Several experiments have been attempted to use GA to search for the optimum descriptor weights. These experiments are summarized in Tables 2 to 7. In these tables, "No. of FDs" means the number of Fourier descriptors used in the optimization process. For example, if this number is *F*, the GA is supposed to search for *F* weights, a single weight for a single FD, which maximize the recognition rate with the minimum number of descriptors.

Table 1: Optimized weights for different numbers of Fourier descriptors.

Experiment No.	**1**	**2**	**3**	**4**	**5**
Training set	X	X	X	O	X, O, N
No. of FDs considered	11	7	6	6	11
Optimized Weights obtained	0.19	0.149	0.135	0.116	0.141
	0.21	0.1489	0.528	0.1457	0.7
	0.2	0.1488	0.415	0.0841	0.0118
	0.771	0.858	0.924	0.4544	0.0675
	0.897	0.941	0.942	0.4418	0.277
	0.96	0.7027	0.935	0.3533	0.4099
	0.864	0.5466			0.7199
	0.88				0.7568
	0.7313				0.1592
	0.1939				0.6508
	0.9048				0.411

Table 1 shows the optimized weight for GA.

Table 2: Recognition rates for different numbers of Fourier descriptors using Euclidean distance **(PSO)**

Experiment No.		**1**	**2**	**3**	**4**	**5**
Training set*		**X**	**X**	**X**	**O**	**X, O, N**
No. of FDs		**11**	**7**	**6**	**6**	**11**
Rec	**X**	95%	93.33%	95%	90%	98.33%

	N	93.75%	93.75%	87.5	87.5%	87.5%
	O	23.33	25%	20%	20%	25%

Table 3. Recognition rates for different numbers of Fourier descriptors using Euclidean distance (ORGA)

Experiment No		**1**	**2**	**3**	**4**	**5**
Training set		**X**	**X**	**X**	**O**	**X,O,N**
No of FDs		11	7	6	6	11
Recognition Rate	**X**	100	98.33	96.67	93.33	93.33
	N	100	100	100	100	93.75
	O	26.67	23.33	21.67	20	16.67

'Object recognition using genetic algorithm (ORGA)'

In the first experiment when a database of 60 transformed objects, in Table 3, was considered, one can see a much better recognition results than in Table 2. ORGA recognizes object 100% in some of the cases. For transformed, noisy and occluded objects, the recognition rate is 100%, 100% and 26.67% as compare to PSO. Experiment 2, when considered for 7 weighted FDs, shows generally, much better results than using 7 weighted FDs in PSO. In case of transformed, noisy and occluded objects, the recognition rate is significant as compare to PSO. Experiment 3, when considered for 6 weighted FDs, shows generally, much better results than using 6 weighted FDs in PSO. Recognition rate for transformed, noisy and occluded objects are better than PSO. Experiment 4, when considered for 6 weighted FDs and the training set is considered for occluded objects, shows generally, much better results than using 6 weighted FDs in PSO. Experiment 5, when considered for 11 weighted FDs and the training set is considered for mixed objects (combination of transformed, noisy and occluded objects), shows much better improvement in case of noisy objects.

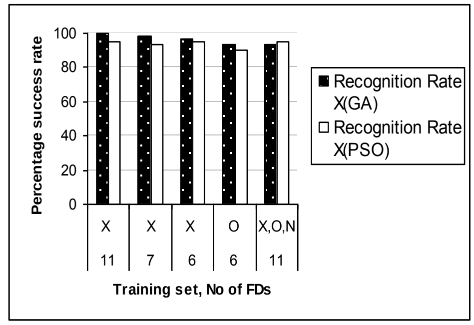


Fig. 3 Comparative graph for transformed objects (using ED)

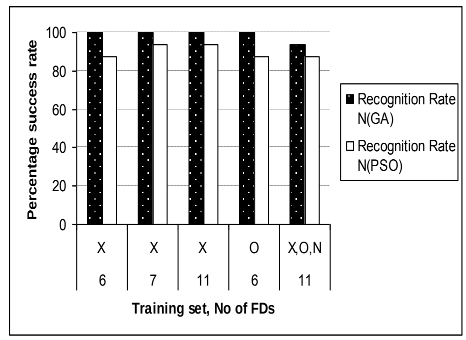


Fig. 4 Comparative graph for noisy objects (using ED)

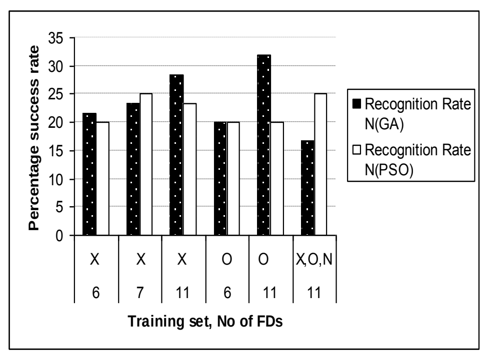


Fig. 5 Comparative graph for occluded objects (using ED)

Table 4: Recognition rates for different numbers of Fourier descriptors using log of square of Euclidean distance. (PSO)

Experiment No.		1	2	3	4	5
Training set		X	X	X	O	X, O, N
No. of FDs		11	7	6	6	11
Recognition Rate	X	95%	96.67%	96.67%	90%	98.33%
	N	93.75%	93.75%	87.5	93.75%	93.75%
	O	21.67	16.67%	20%	20%	25%

Table 4 shows the recognition rates for the different number of Fourier descriptors by using PSO.

Table 5: Recognition rates for different numbers of Fourier descriptors using log of square of Euclidean distance. (ORGA)

Experiment No		1	2	3	4	5
Training set		X	X	X	O	X,O,N
No of FDs		11	7	6	6	11
Recognition Rate	X	100	98.33	96.67	93.33	93.33
	N	100	100	100	100	93.75
	O	26.67	23.33	21.67	20	16.67

Table 4 shows the recognition rates for the different number of Fourier descriptors by using GA.
In the first experiment when a database of 60 transformed objects, in Table 4 and 5, was considered, results of ORGA are better than PSO. ORGA recognizes object 100% in some of the cases. For transformed, noisy and occluded objects, the recognition rate is much better than PSO.

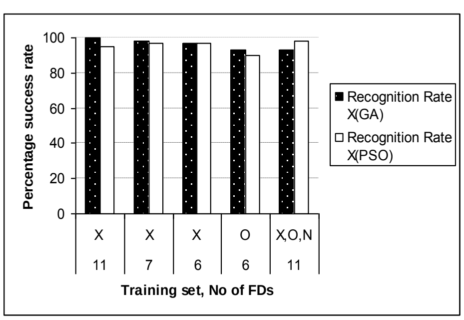


Fig. 6 Comparative graph for transformed objects (using LSED)

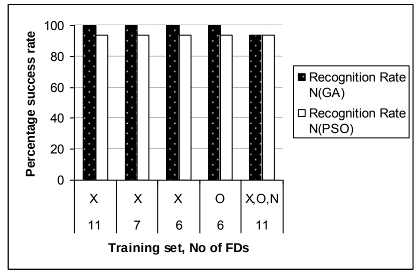


Fig. 7 Comparative graph for noisy objects (using LSED)

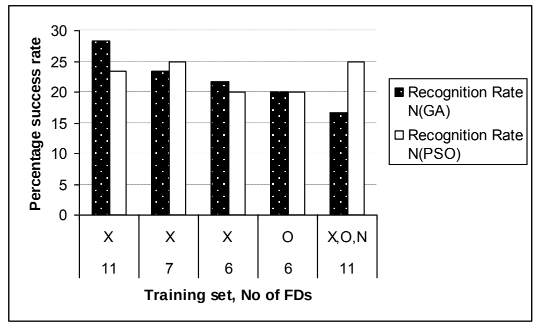


Fig. 8 Comparative graph for occluded objects (using LSED)

Recognition rate for LED is same as in LSED.

Table 6: Recognition rates for different numbers of Fourier descriptors using Percentage error (PSO)

Experiment No.		1	2	3	4	5
Training set		X	X	X	O	X, O, N
No. of FDs		11	7	6	6	11
Recognition Rate	X	75%	70%	65%	73.33%	66.67%
	N	81.25%	81.25%	68.75%	87.5%	75%
	O	18.33%	6.67%	13.33%	13.33%	8.33%

Table 6 shows the recognition rates for the different number of Fourier descriptors by using PSO.

Table 7: Recognition rates for different numbers of Fourier descriptors using Percentage error (ORGA)

Experiment No.		1	2	3	4	5
Training set		**X**	**X**	**X**	**O**	**X, O, N**
No. of FDs		**11**	**7**	**6**	**6**	**11**
Recognition Rate	**X**	80%	80%	75%	78.33%	72%
	N	87.5%	87.5%	87.5%	87.5%	81.25%
	O	15%	7%	11.67%	13.33%	16.67%

In all experiments when a database of 60 transformed objects, in Table 7, was considered, results of ORGA (proposed approach) for transformed, noisy and occluded objects are better than PSO in Table 6.

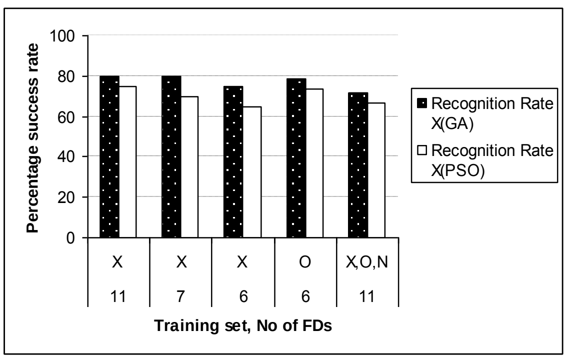


Fig. 9 Comparative graph for transformed objects (PE)

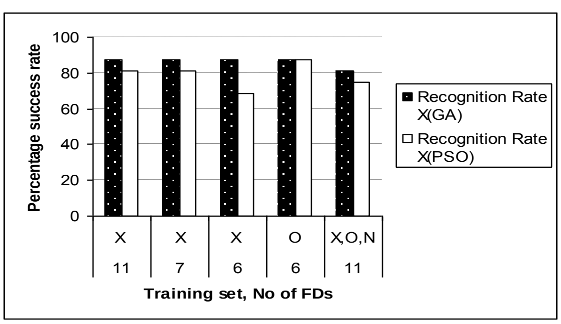


Fig. 10 Comparative graph for noisy objects (PE)

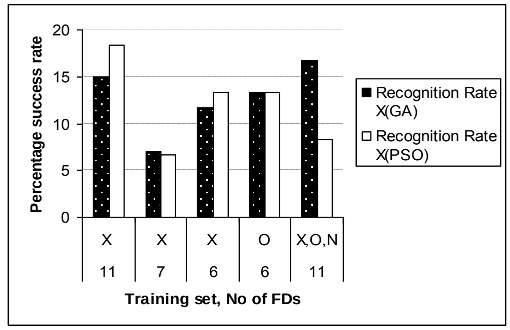


Fig. 11 Comparative graph for occluded objects (PE)

10. Some observations

Here are some observations for the whole discussion in the paper:

- The Fourier descriptors of the boundary are robust to similarity transformations.
- Fourier descriptors were found to be able to recognize at a higher rate if we use nine or more Fourier descriptors. This trend continues when the size of the database is increased from 15 to 45 to 60.
- Most cumulative combinations of Fourier descriptors are able to recognize most of the images correctly for samples without noise or occlusion.
- It is noted that if an image is recognized, it is recognized by most cumulative combinations of Fourier descriptors, and if it is not recognized, then it is not recognized by almost all cumulative combinations of Fourier descriptors.
- Noise (salt and pepper) with density of ten percent has a minimal effect on the recognition ability of Fourier descriptors. When we use eight or more Fourier descriptors add ten percent salt and pepper noise to the images, the accuracy level does not drop.
- Noise of type salt and pepper with ten percent density has similar effect on Fourier descriptors such that the decrease in recognition is noticeable but very slightly.
- Occlusion brings down the recognition rate of Fourier descriptors from 80-90 percent to around 20%.
- The Fourier descriptors show a steady increase in accuracy level as the number of Fourier descriptors used increases. It then stabilizes at same level for nine to eleven descriptors.
- Fourier descriptors perform very poorly in the

presence of occlusion in the image. The occlusion is a big issue in object recognition problem, especially, when we use Fourier descriptors. Since the method works on the boundary of edge of objects, any distortion on the shape will be affected to the recognition process. This occlusion may happen to some part of the object see Fig. 12(a). The remaining part of the object can be used for recognition process and we can guess the correct object according to the excellent part. So, it is recommended to divide the template data (the boundary of object) to four parts depending on *x* and *y* axis, see Fig. 12(b). Then, each of this part can be computed according to the Fourier approach. We can take the test object and can also divide its boundary on the axis. Each of quarter will be computed and compared to database, see Fig. 12(c). The test object will be recognized as agreeing of its parts on the specific object.

- Using GA to find the most suitable descriptors and to assign weights for these descriptors improves dramatically the recognition rate using the least number of descriptors.

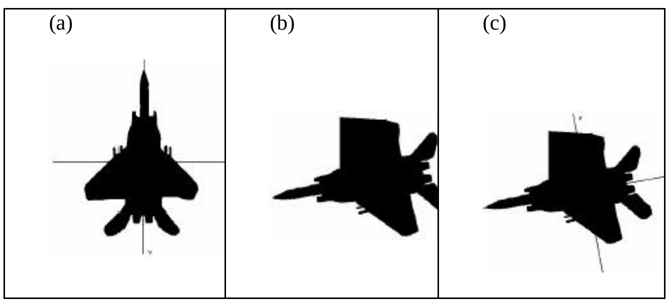


Fig. 12 Occlusion suggestion: Each quarter can be computed and compared to database.

11. Conclusion and Future work

This work has been reported to make a practical study of the Fourier descriptors to the application of Object Recognition. The implementation was done on a LAPTOP using MATLAB 7.1.0. The ultimate results have variations depending upon the selection of number of FDs, similarity transformations, noise, occlusion, and data size. The variety of similarity measures and different combinations of FD features, used in the process, make a difference to the recognition rate..

Four similarity measures, including ED, and PE, LED and LSED provided different recognition results. The images used are all bitmapped images, further investigations are being done with some more complex images.

It can be seen that, using PE with FDs results in less efficient performance than using ED. Moreover, increasing the number of FDs does not necessarily guarantee a better performance.

The images that have to be recognized but failed to be recognized by most of the FD combinations are to be analyzed further. This leads to the theory of optimization to find out appropriate features or attributes in the image that made it difficult to be recognized. The methodology of GA has been utilized successfully for this purpose. Using GA, to find the most suitable descriptors and to assign weights for these descriptors, improved dramatically the recognition rate using the least number of descriptors.

In future, author would like to treat the problem as multi-objective optimization method, also try to enhance the GA by tuning different parameters to maximize the recognition rate while minimizing the number of descriptors.

Acknowledgments

This work has been supported by **AL Jouf University**, Saudi Arabia under the Supervision of Dr Muhammad Sarfraz.

References

[1] David G. Lowe, "Object Recognition from Local Scale-Invariant Features", Vancouver, B.C.,V6T 1Z4, Canada

[2] Weina Ge," OBJECT RECOGNITION: BEYOND VISION", Ph.D. Candidate, Computer Science and Engineering, College of Engineering

[3] M. Sarfraz, "Object Recognition using Fourier Descriptors: Some Experiments and Observations", Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation 2006

[4] G. H. Granlund, "Fourier Preprocessing for hand print character recognition", IEEE Trans. Computers, Vol C-21, Febr. 1972, pp. 195-201.

[5] A Project led by Julien Boeuf and Pascal Belin, and supervised by Henri

http://www.tsi.enst.fr/tsi/enseignement/ressources/mti/descript_fourier

[6] M. Sarfraz, and A.T.A. Al-Awami, (2006), "Object Recognition using Particle Swarm Optimization on Fourier Descriptors", Proceedings of the 11th Online World Conference on Sof Computing in Industrial Applications (WSC11) September 18th - October 06th, 2006 - On the Internet (World

[7] M. Sarfraz, Object Recognition using Moments: Object Recognition using Moments: Some Experiments and Observations: Geometric Modeling and Imaging – New Advances, Sarfraz,

[8] Avrahami, G., Pratt, V.: Sub-pixel edge detection in character digitization. In: Raster Imaging and Digital Typography II, pp. 54–64 (1991)

[9] Hou, Z.J., Wei, G.W.: A new approach to edge detection. Pattern Recognition 35, 1559–1570 (2002)

[10] Richard, N., Gilbert, T.: Extraction of Dominant Points by estimation of the contour fluctuations. Pattern Recognition 35, 1447–1462 (2002)

[11] G. Avrahami and V. Pratt. Sub-pixel edge detection in character digitization. Raster Imaging and Digital Typography II, pp. 54-64, 1991.

[12] Hou Z. J., Wei G. W., A new approach to edge detection, Pattern Recognition Vol. 35, pp.1559-1570, 2002.

[13] N. Richard, T. Gilbert, Extraction of Dominant Points by estimation of the contour fluctuations, Pattern Recognition Vol. 35, pp. 1447-1462, 2002.

[14] H. Freeman and A. Saghri. Generalized chain codes for planar curves. In Proceedings of the 4th International joint conference on Pattern Recognition, pages 701-703, Kyoto, Japan, November 7-10-1978

[15] D. Goldberg, Genetic algorithms in search, optimization and machine learning. Addison-Wesley, 1989.

Mahmood ul Hassan was born in Mansehra, Pakistan, in 1985. He received the B.S. degree from Hazara university and M.S. degree from COMSATS University Abbottabad, Pakistan

From 2008 to 2010, he was a Research Assistant in COMSATS Abbottabad, Pakistan. Since 2010, he has been Lecturer in Computer department at Al Jouf University, Saudi Arabia.

He has six International Journal publications and four International conference publications.

Dr Muhammad Sarfaraz was born in Pakistan. He is the member of IEEE and the IEEE Compter Society.

Currently he is the faculty member of Kuwait university. He has a number of many journal and conference publications. He is also the author of a number of books related to the field of computer science.

Dr Abdel Rehman osman was the faculty member of Al Jouf University Saudi Arabia from 2010 to 2012.

He has a number of many journal and conference publications. He is also the author of a number of books related to the field of computer science.

Muteb Alruwaili

is the faculty member of Al Jouf University Saudi Arabia from 2011 to present.

He has a number of many journal and conference publications.

Multiple Skew Estimation In Multilingual Handwritten Documents

D. S. Guru[1], M. Ravikumar[1] and S. Manjunath[2]

[1]Department of Studies in Computer Science, University of Mysore, Mysore-570016, Karnataka, India
[2]PG Department of Computer Science, JSS College of Arts Commerce and Science, Mysore-570025, Karnataka, India

Abstract

In Indian scenario, most of the time the office documents such as forwarded notices and other documents are multilingual with multiple skews. This poses new challenges in the field of document image analysis. In this direction we are presenting a method of estimating the skew in handwritten multilingual documents. From a handwritten multilingual document image each word is segmented using morphological operations and connected component analysis. Skew of each word is estimated by fitting a minimum circumscribing ellipse. The orientation of each word is estimated and then words are clustered using adaptive k-means clustering to identify the multiple blocks present in the document and average orientation of each block is estimated. In order to corroborate the efficacy of the proposed model experimentation on our own dataset is carried out.

Keywords: *Multilingual Handwritten Document; Multiple Skew; Connected Component Analysis; Adaptive Clustering*

1. Introduction

There is a growing trend to share and exchange information electronically. The need to convert existing paper documents into electronic ones for better archival, retrieval and maintenance is therefore growing. Much knowledge is acquired from documents such as technical reports, government files, news papers, books, journals, magazines, letters, bank cheques, to name a few. The acquisition of knowledge from such documents by an information system can involve an extensive amount of handcrafting, which is generally time consuming and can severely limit the application of information systems. Thus, automatic knowledge acquisition from documents has become an important subject [1]. Conversion of paper documents into electronic forms essentially requires scanning and digitization. Many of the existing approaches of skew detection can only process pure printed document images successfully. But it is a challenging problem to process handwritten documents [2]. One of the most challenging tasks in analyzing handwritten documents is to tackle the inherent skew that is introduced due to writer's handwriting, segment the handwritten lines and estimate the skew angle and its direction [3]. Even though some works are reported in the literature on skew estimation of handwritten documents they are monolingual and generally with a single skew. In country like India with multi languages, officers generally write forwarded notes or observations in different languages with different orientations. This imposes a greater challenge in estimation of multiple skews in a document with multi linguistic. To the best of our knowledge, no work has been reported towards estimation of multiple skews in handwritten multilingual document images. This poses new challenges in the field of document image analysis and has motivated us to take up this research work.

In this work, we consider multilingual handwritten documents with multiple skews. We designed an algorithmic model by the use of morphological operations and clustering of data points. Initially different words are segmented out and then slope of each word is calculated. Words are later clustered based on their likelihood with respect to their spatial coordinates and slopes. The blocks or paragraphs belonging to each cluster are identified, the skew of which is estimated to be the average of the slope of the words belonging to that cluster. It is interesting to note that the proposed method is independent of the scripts and also the writers.

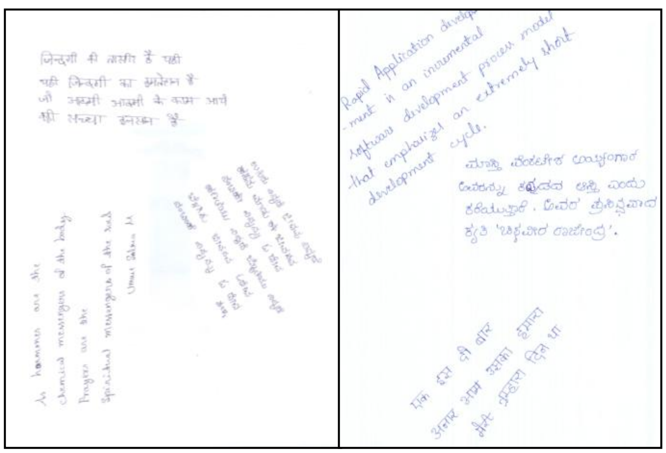
The paper is organized as follows; Section 2 presents an overview of works related to skew estimation in printed and handwritten documents. The model proposed to estimate multiple skews in multilingual document is presented in Section 3. Experimental results on our own data set are presented in the Section 4 and the conclusions are drawn in Section 5

2. Related work

There have been several techniques proposed for accurate estimation of skew angle in document images. The proposed approaches are based on Hough transform [17][18][16][15][9] projection profiles [20][7][19], connected component analysis [22] [28][23][24], cross correlation [25][26], Gradient analysis [4], Fourier spectrum [5], Morphological transforms [6], connected component blotching and linear regression [7], linear regression analysis[8], piecewise covering by

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

parallelograms [9], fuzzy run length [10], averaged block directional spectrum [11], static and dynamic thresholds [12], minimum area bounding rectangle [13] and Cohen's Class Distributions [14].
From the literature survey, we found that most of the existing methods are to estimate skew in printed documents with text as well as non-text. Though they perform well on printed documents, they cannot be employed for handwritten documents. Only few works [29][30] are reported in the literature to estimate skew in handwritten documents that too written in a single script.
The most challenging task involved in the handwritten documents with multiple skew is identifying different blocks written with different skews and correcting it. Figure 1 shows sample multilingual documents containing multiple skew. To the best of our knowledge there is no work found in the literature which can estimate skew of a multilingual document containing multiple skews and correcting it. Hence in this work we have taken up this work as the starting step in handwritten document analysis and recognition.



(a) (b)

Figure 1. Samples of multilingual handwritten documents with multiple skew

3. Proposed Model

The method works on the entire handwritten multilingual document image with multiple skews. Segmenting the handwritten document is necessary in order to identify the different blocks written in different orientations. After segmenting the blocks it would be easy to estimate the multiple skewed blocks.
To segment the blocks we carry out morphological operations (dilation followed by erosion) on the document in order to merge the characters of each word as a single component. Based on the connected component analysis, we segment each word present in the handwritten documents. The reason for identifying each word in the document is that the skew of a block depends on the skew of each word present it. In order to estimate the skew angle of each word we fit a minimum circumscribing ellipse by treating pixels of each word as data points and the slope of the major axis of the ellipse will be considered as the skew angle of the word. To select the pixels belonging to a word, we use the boundary of the connected component (word) and the pixels falling inside the boundary will be treated as data points to fit the ellipse and the angle of major axis will be the skew angle and length of the major axis will be the length of the word [21]. Instead of using directly the angle of major axis we recommend to discretize the angle into bins where each bin will span with a range of α to avoid overhead of calculations in grouping the words into blocks.
To identify the blocks of text written in different orientations we use the orientation of each word along with the coordinates of the end points of the major axis of each word. This is essential as there may be multiple blocks in the single document with different / same orientations at different locations of the document. Hence, the spatial coordinates of end points of the major axis help us in preserving the spatial location of the words. To cluster the data we suggest to store the obtained angle α and coordinates (x_1, y_1, x_2, y_2) of the end points of each axis in a two dimensional matrix where each row corresponds to a word as shown in Table 1.

Table 1. Data Structure used to preserve the orientation of words

	X_1	Y_1	X_2	Y_2	Bin Angle
W_1	X_{11}	Y_{11}	X_{12}	Y_{12}	α 1
W_2	X_{21}	Y_{21}	X_{22}	Y_{22}	α 2
W_3	X_{31}	Y_{31}	X_{32}	Y_{32}	α 3
.					
Wn	X_{n1}	Y_{n1}	X_{n2}	Y_{n2}	α n

In this work we recommend discritization of the skew angle α instead of having it in real domain. Now we suggest clustering of words based on their spatial coordinates and skew angle. Normally K-means clustering technique is used in the literature to cluster the data points. However it is quite difficult to fix up the appropriate value for the parameter k (number of clusters). This essentially requires supervised knowledge which is generally not available in a complete automation process. Hence, in this work we recommend to use the adaptive k-means clustering [27] algorithm. This algorithm is used as there is no over head of identifying the number of clusters. The adaptive k-means clustering is based on spectral based clustering which automatically identifies the number of clusters. After clustering, all the words belonging to each cluster are together treated as a single block. The orientation of each block is estimated by calculating the average orientation of each word present in that block. The obtained average orientation is treated as the skew angle of that block. The block diagram of the proposed model is presented in Figure 2.

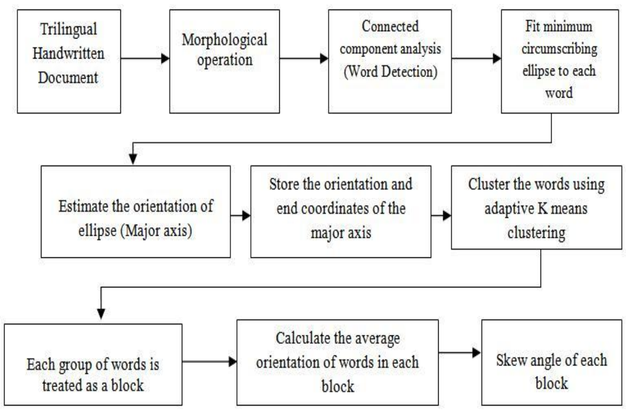


Figure 2. Block diagram of the proposed model

4. Experimentation

In order to conduct experimentation we have created our own data set containing tri lingual documents. We have created 114 handwritten documents using three languages: Kannada, English and Hindi. Also documents containing arbitrary words of their interest in different orientations were also created. The contents are written in an unconstrained manner. Table 2 shows the average number of paragraphs, lines and words containing in the entire data set and the Fig 3 shows the results obtained from the proposed model.

Table 2. Average number of paragraphs, lines and words in the data set considered for experimentation

Average number of	Kannada	Hindi	English
Paragraphs	2.35	1.95	2.25
Lines	5.2	4.2	4.05
Words	16.25	15.35	18.5

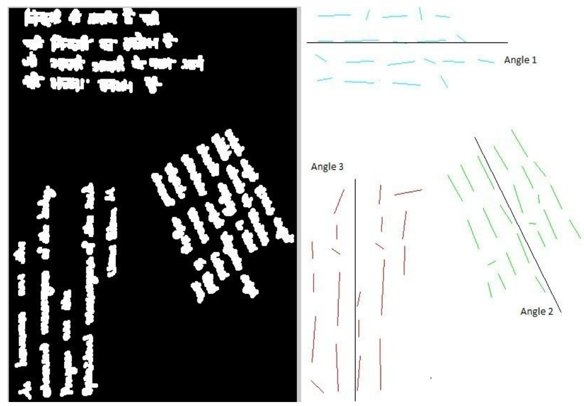


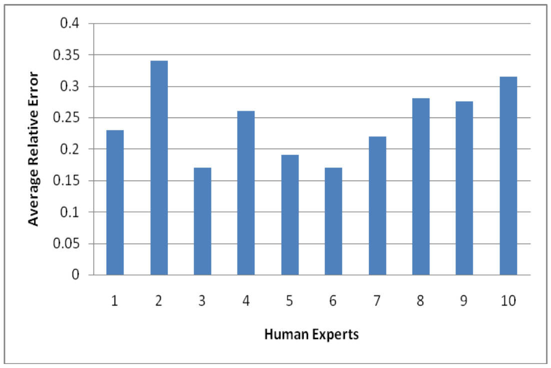
Figure 3. A sample result obtained by the proposed model

To evaluate the performance of the proposed model the skew of each block is estimated manually by drawing a line on each paragraph in the document and the orientation of each line is stored. The stored orientation of the line of each paragraph is compared with the skew obtained by the proposed model. Further to corroborate the efficacy of the proposed model ten human experts are asked to obtain the orientation of each block present in the document and the same has been used to compare the results obtained by the proposed method. Figure 4 shows average relative error in estimating the skew in multilingual handwritten documents with respect to human experts and results obtained by the proposed model on our data set

$$Average\,\mathrm{Re}lative\,Error = \left|\frac{\alpha_a - \alpha_o}{\alpha_a}\right| \qquad (1)$$

Where α_a = Actual skew and α_o = Obtained skew

Figure 4. Average relative error rate of the proposed method with respect to human experts.



In order to compare our proposed method with other state of the art techniques we have made a qualitative comparative analysis. For comparison we have considered several factors such as whether the model can handle documents written in multilingual with multiple skews. Also we have considered whether the model can work only on printed, handwritten or a document containing both handwritten as well as printed. These factors are specifically considered to show the superiority of the proposed model. Table 3 contains qualitative comparative analysis of the proposed model with that of state of the art techniques. From Table 3 it is observed that none of the state of the art techniques work on documents written in multilingual with multiple skews and also documents containing both handwritten and printed text.

Table 3. Qualitative comparative analysis of the proposed model with other state of the art techniques

	Multiple Skews	Multilingua l	Handwritten	Print ed
Kavallieratou et al., (2002)	NO	NO	**YES**	**YES**
Gatos et al., (1997)	NO	NO	NO	**YES**
Amin and Wu (2005)	NO	NO	**YES**	**YES**
Lu and Tan (2003)	NO	NO	**YES**	**YES**
Dey and Noushath (2010)	NO	NO	NO	**YES**
Proposed Model	**YES**	**YES**	**YES**	**YES**

5. Conclusions

In this paper we have presented a new model to estimate multiple skew of multilingual handwritten document .The proposed model is based on angle of major axis of each word segmented using connected component analysis. As the proposed model is based on morphological operations and connected component analysis, if the lines touch or close to each other then the estimation may not be proper. On the other hand it is observed that the proposed method works better for image blocks containing Kannada and English when compared to blocks written in Hindi. This is because the Hindi words contain elongated in the direction of minor axis. In the future we are going to use available models to identify the skews and apply our model to identify multiple skews.

References

[1]Tang, Y ., , S.,lee , Suen. C., : Automatic Document Processing: A Survey. Pattern recognition. Vol. 29. No. 12, pp 1931 – 1952 (1996)

[2]Babu, D,R., Kumat,P,M., Dhannawat,M,D.,: Skew angle estimation and correction of handwritten, textual and large areas of non- textual document images: A novel approach. International conference on image processing, computer vision and pattern recognition. pp 510-515. (2006)

[3]Kasiviswanathan, H., Ball, G, R., Srihari, S, N.,: Top down analysis of line structure in handwritten documents. 20th International conference on pattern recognition, pp 2025 – 2028 (2010)

[4]Sun, C., Si, D., : Skew and slant correction for document images using gradient direction. Fourth International Conference on Document Analysis and Recognition, pp. 142-146 (1997)

[5]Gorman, O, L.,: The document spectrum for Page layout analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 11, pp. 1162 – 1173. (1993)

[6] Najman, L.,: Using mathematical morphology for document skew estimation. Proceedings of the SPIE, Vol. 5296, pp. 182-191. (2003)

[7]Guru, D, S., Punitha, P., Mahesha, S.,: Skew estimation in digitized documents: A novel approach. Proceedings of the Fourth Indian conference on Computer Vision, Graphics and Image processing, pp. 314 – 319. (2005)

[8]Shivakumara, P., Hemanth, Kumar, G., Guru, D, S., Nagbhushan, P.,: A novel technique for estimation of skew in binary text document images based on linear regression analysis. Sadhana, Vol.30, Part I, pp. 69-85. (2005)

[9]Chou, U, C., Chu, Y, S., Chang, F.,: Estimation of skew angles for scanned documents based on Piecewise Covering by Parallelograms. Pattern Recognition, Vol. 40, issue 2, pp. 443-455 (2007)

[10]Shi, Z., Govindaraju, V.,: Skew detection for complex document images using fuzzy runlength. Proceedings of the seventh International Conference on Document Analysis and Recognition (2003)

[11]Lowther, S., Chandran, V., Sridharan, S.,: An accurate method for skew determination in document images. Digital Image Computing Techniques and Applications, pp. 21-22. (2002)

[12]Shivakumara, P., Hemanth, Kumar. G., Guru, D, S., Nagbhushan, P.,: Skew estimation of binary document images using static and dynamic thresholds useful for

document image mosaicing. Journal of society of statistics, computer and applications, Vol. 1, No. 1, pp – 12 (2003)

[13]Safabakhsh, R., Khadiv, Si.,: Document skew detection using minimum area bounding rectangle. In Proceedings of the international conference on information technology, pp. 253-258. (2000)

[14]Kavallierautou, K., Fakotakis, N., Kokkinakis, G.,: Skew angle estimation in document processing using cohen's class distributions. Pattern Recognition letters, vol. 20, pp. 1305 – 1311 (1999)

[15] Kapoor, R., Bangi, D., Kamal, T, S.,: A new algorithm for skew detection and correction. Pattern Recognition Letters, vol. 25, pp. 1215-1229. (2004)

[16] Das, K., Chanda, B.,: A fast algorithm for skew detection of document images using morphology, International Journal on Document Analysis and Recognition, Vol. 4, pp. 109 – 114 (2001)

[17] Amin, A., Fischer, S.,: Robust skew detection method using the Hough transform, Pattern Analysis and Applications, Vol. 3, pp. 243-253. (2000)

[18]Chen, K, Y., Jwang, F, J.,: Skew detection and reconstruction based on maximization of variance of transition counts, Pattern Recognition, Vol. 33, pp. 195-208. (2000)

[19]Dey, P., Noushath, S.,: e- PCP: A robust skew detection for scanned document images. Pattern Recognition. Vol. 43, pp. 937-948 (2009)

[20]Kwag, K, H., Kim, H, S.,: Jeong, H, S., Lee, S. G.,: Efficient skew estimation and correction algorithm for document images. Image and Vision Computing, Vol. 19, pp. 25-35. (2002)

[21] Messelodi, S., Modena, M, C.,: Automatic identification and skew estimation of text lines in real scene images. Patter Recognition, pp. 791-810. (1999)

[22]Liolios, N., Fakotakis, N., Kokkinakis, G.,: On the generalization of the form identification and skew detection problem. Pattern Recognition, Vol. 35, pp. 253-264. (2002)

[23]Lu, Y., Tan, C, L.,: A nearest- neighbour chain based approach to skew estimation in document images, Pattern Recognition letters, Vol. 24, pp. 2315-2323. (2003)

[24]Amin, A., Wu, S.,: Robust skew detection in mixed text / graphics documents, International conference on document analysis and recognition, Vol.1, pp. 247-251. (2005)

[25]Yan, H.,: Skew correction of document images using interline cross-correlation .Computer Vision Graph. Image Process, 55(6), pp 538-543. (1993)

[26]Gatos, B., Papamarkos, N., Chmzas, C.,: Skew detection and text line position determining in digitized documents, Pattern Recognition, Vol.30, no. 9, pp. 1505 – 1519. (1997)

[27]Sanguinetti G., Laidler J and Lawrence N. D.,: Automatic determination of the number of clusters using spectral algorithms. Proceedings of IEEE Machine Learning for Signal Processing, pp. 28-30. (2005)

[28]Pal U., Sinha S and Chaudhuri B. B ., : Multi-oriented text lines detection and their skew estimation, Third Indian conference on computer vision, Graphics and image processing. (2002)

[29]Kavallieratou K., Fakotakis N., and Kokkinakis G.,: Skew angle estimation for printed and handwritten documents using wigner-ville distribution. Image and Vision Computing, pp. 813 – 824 (2002)

[30]Basu S., Chaudhuri C., Kundu M., Nasipuri M., and Basu D K.,:Text line extraction from multi-skewed handwritten documents, Pattern Recognition , Vol. 40, pp. 1825 – 1839.(2007)

D. S. Guru received his BSc, MSc and PhD degrees in Computer Science and Technology from the University of Mysore, Mysore, India, in 1991, 1993, and 2000, respectively. He is currently an associate professor in the Department of Studies in Computer Science, University of Mysore, India. He was a fellow of BOYSCAT. He was a visiting research scientist at Michigan State University. He is supervising a couple of major projects sponsored by UGC, DST, and Government of India. He is a life member of Indian professional bodies such as CSI, ISTE, and IUPRAI. He has authored 36 research papers for journals and 145 peer-reviewed conference papers at international and national levels. His area of research interest covers image retrieval, object recognition, shape analysis, sign language recognition, biometrics, and symbolic data analysis.

M. Ravikumar obtained his BE from Kuvempu University, Karnataka and M.Tech from Visveswaraya Technological University , Karnataka during 1996 and 2001 respectively. Currently he is pursuing PhD in Document Image Analysis in University of Mysore, Mysore, India. He is also working Assistant Professor, Department of Computer Science. Kuvempu University, Karnataka, India. He has authored a few peer-reviewed papers in journals and conferences. His areas of research cover Document Image Processing and Pattern Recognition.

S. Manjunath obtained his BSc and MS degrees in Computer Science from University of Mysore, Mysore, India, respectively, in the years 2003 and 2006. He obtained PhD in video processing from University of Mysore, Mysore, India in 2013. Currently he is working as Assistant Professor, PG Department of Computer Science, JSS College of Arts, Commerce and Science, Mysore. He has authored a few peer-reviewed papers in journals and conferences. His areas of research cover image and video processing, biometrics and text mining.

A conceptual framework for the Adoption of Social Network Technologies (SNTs) in Teaching – case of Ghana

John Kingsley Arthur[1], Kofi Sarpong Adu-Manu[2] and Clement Yeboah[3]

[1] Computer Science Department, Valley View University
Accra, Ghana

[2] Computer Science Department, Valley View University
Accra, Ghana

[3] Department of Business Administration , Strayer University
Columbus, Ohio, USA

Abstract

Social Networking Technologies (SNTs) have changed lots of former procedures in this modern era. One of its greatest influences is in the area of teaching and learning. This research work seeks to identify in the Ghanaian context with unavailability of some ICT infrastructure and other determinants, how this advantageous tool could be adopted and implemented to benefit the area of teaching and learning in the private Universities. To determine the adoption factors three (3) different theoretical frameworks were considered, namely; Unified Theory of Acceptance and Use of Technology (UTAUT), Unified Theory of Acceptance and Use of Technology (UTAUT) Extension Model and Technology Acceptance Model. The research method adopted was quantitative and hence paper form questionnaire was used for the data collection. Several recommendations were made to enhance the adoption of SNTs for the purposes of supplementing teaching in the tertiary education in Ghana.

Keywords: *Social Networking Technologies, theoretical framework, ICT Infrastructure, Adopting factors.*

1. Introduction

There have been tremendous changes in the line of former process with the introduction of Information Technology tools. Technology is constantly re-shaping our experience and, in education, it is challenging the way which teaching and learning activities are structured and delivered (Mistry, 2011). This has been aided by the advances in the technologies running on the global internet.

Tertiary education students use information technology tools for several purposes such as connecting to friends, family, reading news, event notification, entertainment, etc. Technology has become an integral part of our lives, and one way that many students stay connected is through the use of social networking sites such as Facebook, MySpace, Twitter, LinkedIn, etc." (Kister, 2011).

These Social Networking Technologies when introduced into education cultivates several advantages such as faculty and staff sharing learning materials with students, making teaching and learning location independent, enhances creativity and innovation, very easy to learn to use, they are free, provision of multimedia tools for enhancing the understanding of the students. According to the Underscience Publishers(2011), Facebook usage is around 90% across campuses and many educational institutions new students are given orientation on how to capitalize on social networking to improve their experience of their course, and to enhance their final results.

However, this is not practiced in the Universities in Ghana. Therefore, this research will seek to investigate the factors that influence the adoption of Social Networking technologies in the private Universities in Ghana to enhance teaching and learning.

For the above objective to be achieved, this research would attempt to prompt the few research work done in Social Networking and its usage in the Universities in Ghana. The majority of the literature has focused on Social networking technology usage in US, UK and some Asian countries like Taiwan and China. Some of the research work done in these countries include Selwyn N. (2012), Ahmed I. and Qazi T.F. (2011), Kiser and Porter (2011), Redecker et al (2010), Bonzo (2010), Yang and Tang (2003) and many more.

However, few researchers have contributed to the knowledge of Social Networking around Africa. Some of the few done in Africa include Social Networks and Technology Adoption in Northern Mozambique (Bandiera & Rasul, 2005) and that done in Ghana is the work of Asiedu (2012) titled "A case of online social networking in the workplace in Ghana". From this perspective, this research would contribute to the knowledge base of Social Networking technologies and their usage in Ghana and Africa at large.

2. Statement of Problem

"Some educators feel that social networking is innately disruptive to the education process. Students may access them on their laptops or cell phones during class. Some educators respond by banning these electronic devices" (Kister, 2011). However, Online social networking sites, such as Facebook, can help students become academically and socially integrated as well as improving learning outcomes (Under Science Publishers, 2011), and yet they are not being used by lecturers and students for teaching and learning. Therefore, the research looks into how social networking technologies could be adopted and implemented to enhance effective teaching and learning in the private Universities in Ghana.

3. Objectives

i. To assess the factors influencing the adoption of social networking technologies for teaching and learning in Private Universities in Ghana.
ii. To propose a conceptual framework for which the Ghanaian Private Universities can consider for the adoption of SNTs for teaching.

3. Research questions

The questions the research seeks to answer are:

i. What are the factors that influence the adoption of Social Networking Technologies in private Universities in Ghana?
ii. How can Social networking technologies be adopted and implemented in Universities for effective teaching?"

4. Literature Review

In this section an in depth study of three previous works done in this area is briefly discussed with respect to their limitations as far as their proposed frameworks. Firstly, Hwee (n.d.) proposed a framework generalizing his findings found as results of studying only one secondary school in the whole of Singapore and generalize the finding. The sample size is not enough representation of all the numerous schools in Singapore. The author identified key factors of trust, comfort level, command of language, attitude towards work and image, as additional factors that would influence the adoption of Social Networks.

Secondly, Venkatesh et al. (2003) combined eight technology acceptance models to propose a framework known as the Unified Theory of Acceptance and Use of Technology (UTAUT). This model encompasses the following factors performance expectancy, expectancy effort, social influence, facilitating conditions and behavioural intention. In addition to these factors other moderators (gender, age, voluntaries, and experience) were used to measure the influences on the factors. Though extensive work was done in this area the following setbacks were identified: lacks other key contextual factors when it comes to the use and implementation of social networks. . Indicators such as policy and culture and financial support were not discussed in the work of Venkatesh et al. (2003) and the authors were silent on the acquisition and implementation factors.

Finally, the work presented by Monguatosha et al.(2011) discusses two critical factors in addition to the existing factors on the UTAUT model. In their work, budgeting and accountability (BA) and organizational culture (OC) was added as additional factors that could influence the use of SN's adding up to facilitation conditions. The Technology Acceptance Model (TAM) was used in the proposition made by Monguatosha et al.(2011).

To propose a conceptual framework for private Universities in Ghana, all the three frameworks were combined and yet there was and identified gap that needs to be filled. In the frameworks mentioned above, the researchers were silent on acquisition, implementation and maintenance of ICT infrastructures. They were also silent on culture, and on policies (i.e. Government policies, institutional policies) and its influence on the adoption of technology. Therefore a new research framework for effective learning and teaching is proposed which will cater for the gap left even when the three models have been combined. The additional factors are discussed below.

4.1Policies

Policies are very important when it comes to the issue of sustainability and implementation of lay down principles of institutions. In this regard several policies listed below affect the use of SNT in TL in private universities.

i. **Government policies:** The private universities are directly affected by polices made by the government with respect to the initiatives on ICT. According to Canaves (2011), some regimes view social networking sites as a threat to stability and restrict their citizens' access to Facebook, Twitter, YouTube, and the like. For example in China, there is a ban on usage of Facebook and Twitter. This policy does not permit any institution to implement SNT's in TL.

ii. **Institutional policies:** Policies made by institutions can permit, limit or ban the use of social media technology by all parties involved. According to ASTD Research (2011) on social learning, social tools are often held to higher standards than traditional business tools because they are new, and negative stories can go viral quickly. Rather than ban the use of social tools, educate people how to use them effectively for work. They are the future of collaboration and learning at work, so the more you prepare people for how to use the tools respectfully and how to apply good social practices, the better. Several organizational factors such as a technology strategic plan, administrative support, professional development, and an incentive policy, can influence the adoption of technology (Anderson, Vamhagen & Campbell (1998), Barone & Hagner, 1998, & Green, 1998).

4.2 Culture

According to ASTD Research (2011) on Social Learning, the major obstacle institutions normally have to adopting new technology is that "We've never done it that way. Our organization will never embrace social media. Social media can't be governed and it's against our compliance rules." In most organizations when a new strategy is to be implemented and counters the corporate culture—the corporate culture will push back and almost always win.

This type of challenge is often identified when leaders and employees say, "We've never done it that way." This attitude represents a fear of something new. Cultures may be more resistant to adoption if they are less technologically-dependent, or if the organizations are more hierarchical. Demographics, such as generational differences, may also play a role in this dynamic.

This means that culture would be one of the key factors or indicators when SNT's are to be implemented in institutions for TL. In an institution, all the parties (lecturers, students and staff) or individuals that make up the institutional community have different cultural beliefs as well as the institution itself.

There are individual factors that include the available time that university instructors can spend in learning how to use technology, their tolerance of possible failures in using technology, and their beliefs in the effectiveness of technology in enhancing teaching and learning (Anderson, Vamhagen & Campbell (1998), Adams, 2003, Ebersole & Vorndam, 2003 & Hannafin & Savenye, 1993).

4.3 ICT Infrastructure conditions

According to Digital Education Revolution official website, ICT infrastructure provides a technology foundation within a school. It enables students, teachers and school staff to access a wide range of tools, services and digital resources to support teaching, learning and school administration. The term ICT Infrastructure in the context of this research defines the enabling environment for private universities to use in the adoption of social network technologies (SNT's) for effective teaching and learning (TL).

The adoption of technology at universities tremendously influences its operations and services rendered to its customers (faculty, students and staff). Previous studies have shown that, there are three types of factors influencing the adoption of technology at universities: technical factors, individual factors, and organizational/institutional factors (Nantz & Lundgren, 1998).

In terms of ICT infrastructure development, the factors must meet three key considerations (availability, accessibility and performance) if the institution or organization wants to achieve high throughput as proposed in this framework.
With reference to this framework, the technical factors are defined to include variables like access to technology, technical support, etc. (Nantz &Lundgren, 1998; & Schifter, 2000). To effectively implement SNT's in institutions of higher learning the three indicators should be considered: security and privacy issues, hardware and software issues, and network and internet issues.

The social network environment is primarily used for socializing; the adoption of the environment for TL is a big issue to consider. Since the security and privacy issues are off concerning in the adoption of SNT for TL, institutions should increase the level of trust on the part of all the parties involved (lecturers & students). The hardware and software services should be in an institution intending to use SNT's for TL should be of minimum requirement.

The Network infrastructure according to Digital Education Revolution official website connects the access devices in the school with the required tools, services and digital resources. Many of these tools, services and digital resources will be external to the school. The network infrastructure components include: internal communications services, cabling and equipment, telecommunications services, server computers and associated storage devices, environmental management equipment and operating software for server computers, communications equipment and related hardware. For such a system a high speed Internet is required to aid in ease of accessibility (EOA) by the users of the system. The diagram below shows the conceptual framework derived from the study.

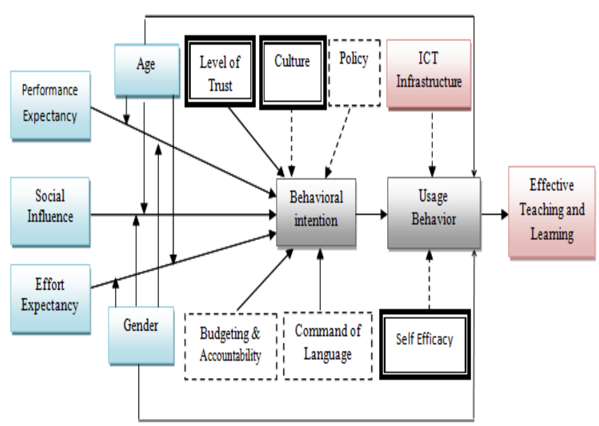


Fig. 1 Adopted Framework

5. Methodology

To achieve the said objective, the research design adopted was the descriptive research method. In this research the population is the Lecturers of all private Universities in Ghana. Sample size of eighty (80) lecturers was used. Stratified sampling method was used to group the University population into two (2) main categories: teaching staff and supporting teaching staff. Non-Random sampling was employed to select 400 respondents for data collection. This technique was more efficient because it improves accuracy of estimates.

The research instrument used was questionnaire. The questionnaire was closed-ended with a few being open-ended. This was to enable the researchers to analyze the information easily since the respondents were many and also to give the respondents the opportunity to answer the questions with ease. The data collection tools used paper based questionnaires.

Data analysis was done using Statistical Package for the Social Sciences (SPSS). SPSS was chosen because it has couple of features that would enhance display of results in text and variations on graphical representations of results. It has proven to be good and commonly used by majority of researchers for statistical representations.

6. Results/Discussions

In this section the two main objectives of the research are discussed and related to literature.

Discussion of objective 1: To assess the factors that influence adoption of Social Networking Technologies for teaching and learning.
The following were identified as factors from literature, that they could influence the adoption of SNTs. They are discussed subsequently.

6.1 Performance expectancy

The table below shows the distribution of respondents by Performance expectancy. The questions below on performance expectancy are represented by the SN numbers in the table 1.
1. I will find SNTs useful in my teaching
2. Using SNTs enables me to accomplish more tasks quickly
3. Using SNTs increases my productivity
4. Using SNTs helps me to receive good evaluation by management

Table 1: distribution of respondents by performance expectancy

SN	*CD*	*MD*	*SD*	*N*	*SA*	*MA*	*CA*
1	0	8	2	16	2	22	30
2	0	0	12	2	18	32	16
3	0	0	12	7	12	17	32
4	0	0	12	13	9	31	15
Totals	0	8	38	38	41	102	93
%	0	2.5	11.9	12	12.8	32	29.1

To ascertain the Performance expectancy factor, several questions such as SN (1 - 4) were asked. From table 1 above, it is evident that the majority of the lecturers, constituting 73.7505% agreed that the use of SNTs will enhance their teaching performance.

However 14.38% disagreed and very few 11.875% were neutral. With the statistics provided, we accept that the adoption of SNTs will enhance the performance of lecturers in their teaching and hence will influence their behavior to adopt and use SNTs. The outcome of this research is an affirmation of the findings of (Hwee, n.d, Venkatesh et al. 2003) where they identified Performance expectancy as a leading factor in the adoption of technology.

6.2 Effort Expectancy

To critically assess effort expectancy as an influential factor to the adoption of SNTs, questions SN (5 – 8) as indicated in table II below were asked the respondents. The questions below on effort expectancy are represented by the SN number in the table.

5. My interaction with SNTs is clear and understandable
6. It is easy for me to become skillful at using SNTs
7. I find SNTs easy to use
8. Learning to operate SNTs is easy

Table 2: distribution of respondents by effort expectancy

SN	*CD*	*MD*	*SD*	*N*	*SA*	*MA*	*CA*
5	0	0	19	5	10	23	23
6	0	0	12	1	14	37	16
7	0	7	6	23	22	15	7
8	0	0	15	11	7	20	27
Totals	0	7	52	40	53	95	73
%	0	1.75	13	10	13.25	23.75	18.25

From table 2, the majority constituting 55.25% of the lecturers agreed that, little input will have to be made to know how to use SNTs. 14.75% of the respondents disagreed and 10% were neutral. With the statistics provided it is evident that the lecturers are already using it in a way; possibly not direct on academic purpose and hence are able to tell that little effort is required. This result is in line with that of Hwee (n.d.); where the students from his case study confirm that the little effort required in using the technology influenced their adoption.

6.3 Attitude towards the use of SNTs

To examine the attitude of lecturers towards the use of SNTs, questions such as SN (9 - 11) as shown in table 3 below were asked. The questions below on attitude towards the use of SNTs are represented by the SN numbers in the table 3.

9. Using SNTs is a good idea
10. SNTs makes teaching more interesting
11. Teaching with SNTs is fun

Table 3: distribution of respondents by attitude towards the use of SNTs

SN	*CD*	*MD*	*SD*	*N*	*SA*	*MA*	*CA*
9	0	0	0	11	23	26	20
10	0	0	0	18	17	45	0
11	0	0	29	5	7	24	15
Totals	0	0	29	34	47	95	35
%	0	0	12.08	14.2	19.58	39.5	14.5

From the table 3 above, the majority (73.75%) agreed that it's a good idea and it will make teaching more interesting and fun. However, very few (12.08%) disagreed and 14.167% were neutral. The higher value of agreement presupposes most lecturers are interested in the use of technologies for teaching. This result is agreement with that of (Hwee,n.d., & Venkatesh et al.,2003).

6.4 Social Influence

Table 4 below shows the distribution of respondents by their social influence. The questions below on social influence are represented by the SN numbers in the table 4.

12. People who influence my behavior think that I should use SNTs
13. Fellow lecturers have been helpful in the use of SNTs
14. In general, the university has supported the use of SNTs

Table 4: distribution of respondents by social influence

SN	*CD*	*MD*	*SD*	*N*	*SA*	*MA*	*CA*
12	0	0	8	18	12	42	0
13	7	0	11	13	12	37	0
14	0	0	11	23	14	21	11
totals	7	0	30	54	38	100	11
%	2.91	0	12.5	22.5	15.83	41.66	4.58

From table 4 above, the majority of the respondents constituting 62.08% agreed that people around them such as fellow lecturers and staff influence their behavior to adopt and use SNTs; also their various schools support the use of SNTs for teaching. On the other hand, 15.417% disagreed and 22.5% were neutral. Hwee (n.d.) had a similar result when his students adopted the SN's as their fellow students were using them.

6.5 ICT Infrastructure conditions

Table 5 shows the frequency distribution of respondents by ICT infrastructure conditions in their respective Universities. The questions below on ICT infrastructure are represented by the SN numbers in table 5.

15. I have the hardware and software resources necessary to use SNTs
16. A specific person (group) is available for technical support when SNTs difficulties are encountered
17. The internet and network infrastructure available for use of SNTs is consistent and of high speed
18. My personal records are safe when using SNTs since the institution has a secure system installed
19. SNTs is not compatible with other systems I use

Table 5: distribution of respondent by ICT infrastructure conditions

SN	*CD*	*MD*	*SD*	*N*	*SA*	*MA*	*CA*
15	0	0	32	11	11	17	9
16	0	0	0	20	25	31	4
17	0	9	11	21	10	19	10
18	0	14	20	21	18	7	0
19	0	0	0	16	30	33	1
Totals	0	23	63	89	94	107	24
%	0	5.75	15.75	22.25	23.5	26.75	6

From table 5 above, the majority (56.25%) agreed that they have hardware and software resources, a good internet and secured network infrastructure in their respective universities. Very few (21.5%) disagreed and 22.25% were neutral. This result implies that implementation of SNTs for teaching would be very much easier. Teachers trusting the web will boost the usability of the system.

6.6 Self efficacy

To assess self-efficacy as an influential factor to the adoption of SNTs by lecturers, questions such as SN (20 - 23) as shown in table 6 below. The questions below on self-efficacy are represented by the SN numbers in the table 6.

20. I feel apprehensive about using SNTs
21. It scares me to think that I could lose a lot of information using SNTs by hitting the wrong key
22. I hesitate to use SNTs for fear of making mistakes I cannot correct
23. SNTs are somewhat intimidating to me

Table 6: distribution of respondents by self-efficacy

SN	*CD*	*MD*	*SD*	*N*	*SA*	*MA*	*CA*
20	15	3	10	7	9	21	15
21	6	14	28	16	4	11	1
22	0	0	5	20	21	34	0
23	13	6	15	29	9	8	0
totals	34	23	58	72	43	74	16
%	10.63	7.188	18.13	22.5	13.4	23.13	5

From table 6 the majority (41.563%) agreed that they feel apprehensive and intimidated; and also think that they could lose their information when using SNTs as opposed by 35.948% and 22.5% are neutral. This implies that the self-efficacy of the lecturers to adopt SNTs will negatively influence their behavior to accept and use SNTs for teaching. The result here is contrary to that of Venkatesh et al. (2003), where they explained that Self efficacy is not a direct determinant to behavioral intension of a person to adopt technology. Definitely if the students are afraid to use the technology it has a negative influence on the adoption.

6.7 Policies: Government and Institutional

Table 7 below shows the frequency distribution of respondents by government and institutional policies. To analyze the adopting factor, policies, several questions such as SN (24, 25, and 26) as shown below were asked. The questions below on Government and Institutional Policies are represented by the SN numbers in the table VII.

24. Government policies on ICTAD influences my use of SNTs
25. I will always wait for government policies to be implemented before usage of SNTs
26. The University has policies governing the use of SNTs
27. The University's norms affect my intent to the use of SNTs

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Table 7: distribution of respondents by policies

SN	CD	MD	SD	N	SA	MA	CA
24	20	6	4	25	21	4	0
25	5	11	24	22	7	10	1
26	21	9	7	22	15	6	0
27	20	0	0	22	17	13	8
Totals	66	26	35	91	60	60	9
%	20.63	8.125	10.94	28.44	18.75	10.31	2.81

From table 7 above, the majority (39.95%) disagreed government nor institutional policies will influence their behaviour to adopt SNTs for teaching. Few constituting 31.8755% of the lecturers, however, agreed and 28.438% were neutral.

6.8 Culture

Table 8 shows the frequency distribution of respondents by cultural factor. To assess the influencing factor, culture, questions SN (28 & 29) were asked the correspondents. The questions below on Government and Institutional Policies are represented by the SN numbers in the table 8.
28. My cultural background does not support my use of SNTs
29. "We have not done it that way" How do you relate to this statement with the use of SNTs in teaching

Table 8: distribution of respondents by culture

SN	CD	MD	SD	N	SA	MA	CA
28	17	2	19	8	2	13	19
29	4	7	22	7	19	21	
Totals	21	9	41	15	21	34	19
%	13.13	5.63	25.63	9.38	13.125	21.25	11.88

From table 8 above, the majority of the lecturers constituting 46.25% agreed that they have not used SNTs for teaching before and also, their cultural background does not support the use of SNTs. However, only 4.385% have used SNTs for teaching and their cultural background as well castigate the use of SNTs.

6.9 Trust level

The questions below on trust level are represented by the SN numbers in the table 9.
30. Someone will use my private information for something else
31. Lecturers will have a good reputation and hence only academic materials would be displayed and shared on SNTs
32. Information provided by colleagues and students cannot be verified
33. Fellow teachers will share very good academic materials related to my field of teaching.

Table 9: distribution of respondents by trust

SN	CD	MD	SD	N	SA	MA	CA
30	27	7	6	12	11	2	15
31	24	0	7	18	16	1	14
32	7	0	14	13	36	1	9
33	6	1	0	45	10	1	17
Totals	64	8	27	88	73	5	55
%	20	2.5	8.44	27.5	22.813	1.56	17.18

From table 9 above, most (41.563%) of the lecturers said they were afraid of what people would use their private information for, they do not trust that it is only academic materials that would be shared on the network. The minority constituting 30.938% of the lecturers have an opposing idea and 27.5% are not sure of the trust level of SNTs.

6.10 Budgeting and Accountability

Budgeting and accountability was identified as one of the factors influencing adoption of SNTs. Table 10 below, showcases the frequency distribution of lecturers by budgeting and accountability. The questions below on budgeting and accountability are represented by the SN numbers in the table 10.

34. Pay for the use of SNTs for effective usage of resources for teaching

Table 10: distribution of respondents by budgeting and accountability

SN	CD	MD	SD	N	SA	MA	CA
34	14	26	10	2	17	4	0
Totals	17.5	32.5	12.5	2.5	21.25	5	0

From table 10 above, the majority (62.5%) of the lecturers said they do not pay for the effective use of SNTs and their resources for teaching. On the contrary 26.25% of the lecturers said they pay for the use of the SNTs. This is actually true because, in most of the private universities, the students are charged for internet services and it covers for that of the faculty and staff.

6.11 Command of language

It was identified from literature that, by the command people have over the language for which SNTs are captioned, they are able use it very well. Table 11 above shows the distribution of respondents by command of language. The questions related to command of language is

indicated in the table 11 by SN numbers (35 and 36). These are the questions
35. English is my primary language which enhances my use of SNTs for learning
36. English is my secondary language which hinders my use of SNTs for learning

Table 11: distribution of respondents by command of language

SN	CD	MD	SD	N	SA	MA	CA
35	21	0	7	19	16	9	8
36	16	0	7	37	7	1	12
Totals	37	0	14	56	23	10	20
%	23.13	0	8.75	35	14.38	6.25	12.5

The majority (35%) of the lecturers neither agreed nor disagreed that language is a factor influencing adoption of SNTs for teaching purposes. 33.125% agreed and 31.88% disagreed. The difference between the agreed and the disagreed percentage is a close margin. Therefore, command of language is not a critical point to be considered for the adoption of SNTs in for the purpose of teaching. This is in conformity with the outcome of Hwee (n.d.) work.

6.12 Behavioral intension

Table 12 shows the distribution of respondents by their intension to adopt and use SNTs. To access behavioral intension of lecturers, the following questions, SN (37, 38 & 39) were asked. The questions below are related to behavioral intention as indicated in table 12.

37. I intend to use SNTs in the next semesters
38. I predict I would use SNTs in the next semesters
39. I plan to use SNTs next semester

Table 12: distribution of respondents by behavioural intention

SN	CD	MD	SD	N	SA	MA	CA
37	4	0	15	28	17	0	16
38	0	0	6	28	38	4	4
39	4	0	0	13	30	11	22
	8	0	21	69	85	15	42
Totals	3.33	0	8.75	28.75	35.42	6.25	17.5

From table 12 above, 59.17% of the lecturers said they intend to use the SNT for teaching as opposed to the few (12.08%) who do not intend using SNTs the next semester for teaching.

7. Conclusions

From the above discussions, three main theoretical frameworks were thoroughly reviewed and constructively critiqued to identify the gaps. Additional factors of ICT infrastructure conditions, policies and culture were considered to fill the gap. Upon data collection from the selected private Universities of study; the study came up that, all factors were influencing the adoption of SNTs for teaching with the exception of culture and trust. Policies, budgeting and command of language were identified to have a negative influence on the adoption of SNTs for the purposes of teaching.

8. Recommendation

For SNTs to be adopted by the Ghanaian private Universities the ICT infrastructure of the intuition would have to be improved. The speed of the internet should be enhanced, a much more robust network infrastructure should be established because students are particular with how secured the system is. Therefore, if ICT infrastructure is enhanced students and lecturers would have a sustained interest in the use of the SNTs for learning purposes.

University management should consider how to make laptops available for all of their students. When students have laptops to use, it will encourage them to have a strong affiliation with the technology. It boosts the frequency of usage. The absence of the laptops means, that students will have to cluster around machines in the schools computer laboratories to use these social networking technologies.

A good wireless network should be situated in the university. Access to network only through cables would affect mobility of the students and lecturers in the usage of the technology. At worse case, hotspots should be created on campus where students can cluster around and use.

The University management will have to include a comprehensive budget covering the maintenance of the entire infrastructure for running the SNTs. This will help that, for SNTs that require payment to use some features, like video conferencing and like could be enabled.

Students and faculty alike, need to be trained on how to use SNTs. The University management should therefore, regularly organize seminars and workshops that will enhance the knowledge base of the faculty and students on how to use the SNTs for teaching and learning.

A proper auditing group should be designated to measure the progress of the use of the technology. This will inform the University the performance of the faculty and students as the technology was adopted.

There should be a developed policy that will govern the use of SNTs; such that students faculty alike will not abuse its usage.

Professors should encourage the use of SNTs for teaching and learning because they were identified to have a strong social influence on their colleagues and the students they teach.

University management should consider the use of SNTs by faculty in their faculty appraisal. This will motivate all professors in the universities to use the SNTs in their teaching.

A special ICT team should be designated to render assistance to students and faculty who will have difficulty in the use of the SNTs. This team will provide guidance on how to use the SNTs and all other technical assistance.

.

References

[1] V. Mistry. “Critical care training: using Twitter as a teaciiingtooiVirendraMistry”. Uritish journal of Nursing, 20 (20), 2011.

[2] A.I.T. Kister, and Porter, T. Social Networking: Integrating student and University Professor Utilization. Global Conference on Business & Finance Proceedings, 6 (1), 2011.

[3] Underscience Publishers. Social learning: Can Facebook and related tools improve educational outcomes?. ScienceDaily. Retrieved from: http://www.sciencedaily.com/releases/2011/05/110509091557.htm, 2001.

[4] N. Selwyn. Social Media in Higher Education. Europa World of Learning, 2012.

[5] I. Ahmed, and T.F. Qazi. A look out for academic impacts of Social networking sites (SNSs): A student based perspective. African Journal of Business Management, 5(12), pp. 5022-5031, 2011.

[6] C. Redecker, K. Ala-Mutka, and Y. Punie. Learning 2.0- The impact of Social Media on Learning in Europe. JRC Technical Notes, 2010.

[7] J. Bonzo and G. Parchoma. The paradox of Social Media and Higher Education. African Journal of Business Management. 5(12). pp. 5022-5031, 2010.

[8] H. Yang and J. Tang. Effects of Social Network on Students academic performance: A web-based forum study in Taiwan, 2003.

[9] O. Bandiera, and I. Rasul. Social Networks and Technology Adoption. Mozambique, Northern, 2005.

[10] K. A. Asiedu. A Case Study of Online Social Networking at Workplaces in Ghana, 2012. StudyMode.com. Retrieved 10, 2012, from http://www.studymode.com/essays/A-Case-Study-Of-Online-Social-1179317.html

[11] T.C. Hwee. Factors Affecting the adoption of Social Media in Singapore Schools: Extending the UTAUT model, n.d.

[12] V. Venkatesh, M.G. Morris, G.B. Davis, and F.D. Davis. User Acceptance of Information Technology: Toward A Unified View. MIS Quarterly. 27(3). pp. 425-478.

[13] S. Canaves. China’s Social Network problem. IEE Xplore. 48 (6). pp. 74-77, 2011.

[13] ASTD Research. Social Learning. ASTD: VA. Alexandria, 2011.

[14] T. Anderson, S. Varnhagen, and K. Campbell. Faculty adoption of teaching and learning technologies: Contrasting earlier adopters and mainstream faculty. The Canadian Journal of Higher Education. 28(2/3). Pp. 71-98, 1998.

[15] C. A. Barone, and P. R. Hagner. Technology-enhanced teaching and learning: Leading and supporting the transformation on our campus. San Francisco: Jossey-Bass, 2001.

[16] K. Green. Colleges struggle with IT planning. The 1998 National Survey of Information Technology in Higher Education, 1998.
Retrieved from: http://www.campuscomputing.net/summaries/1998/

[17] N. Adams. Educational computing concerns of postsecondary faculty. Journal of Research on Technology in Education. 34(3). Pp. 285-303, 2003.

[18] S. Ebersole and M. Vorndam. Adoption of computer-based instructional methodologies: A case study. International Journal of E-Learning. 2(2). Pp. 15-20, 2003.

[19] R.D. Hannafin and W. C. Savenye.Technology in the classroom: The teacher's new role and resistance to it. Educational Technology. 33(1). pp. 26-31, 1993.

[20] K. Nantz, and T.D. Lundgren. Lecturing with technology. College Teaching. 46 (1). Pp. 53-56, 1998.

[21] C. Schifter.“Faculty participation in asynchronous learning networks: A case study of motivating and inhibiting factors.” Journal of Asynchronous Learning Networks, pp. 4(1), 15-22, 2000. http://www.sloan-c.org/publications/jaln/v4n1/v4n1_schifter.asp

[22] G.M. Munguatosha, P.B. Muyinda, and J. T. Lubega. "A social networked learning adoption model for higher education institutions in developing countries", On the Horizon, Vol. 19 Iss: 4, pp.307 – 320, 2011.

Modified Pattern Extraction Algorithm for Efficient Semantic Similarity Measures between Words

Pushpa C N[1], Thriveni J[1], Venugopal K R[1] and L M Patnaik[2]

**[1] Department of Computer Science and Engineering,
University Visvesvaraya College of Engineering
Bangalore, Karnataka, India**

**[2] Indian Institute of Science,
Bangalore, Karnataka, India**

Abstract

Semantic Similarity measures plays an important and significant role in information retrieval, natural language processing and various tasks on web such as relation extraction, community mining, document clustering, and automatic meta-data extraction. In this paper, we have proposed a Modified Pattern Extraction Algorithm [MPEA] to compute the semantic similarity measure between the words by combining both page count method and web snippets method. Four association measures are used to find semantic similarity between words in page count method using web search engines. We use a Sequential Minimal Optimization (SMO) Support Vector Machines (SVM) to find the optimal combination of page counts-based similarity scores and top-ranking patterns from the web snippets method. The SVM is trained to classify the synonymous word-pairs and non-synonymous word-pairs. The proposed approach aims to improve the Correlation values, Precision, Recall, and F-measures, compared to the existing methods. The proposed algorithm outperforms by 89.8 % of correlation value for Miller-Charles dataset and 75.3% of correlation value for Word Similarity dataset.

Keywords: *Information Retrieval, Semantic Similarity, Support Vector Machine, Web Mining, Web Search Engine, Web Snippets.*

1. Introduction

Search engines have become the most supportive tool for obtaining useful information from the Internet. The search results obtained by most of the popular search engines are not satisfactory. It amazes users because they do input the right keywords and search engines do return pages involving these keywords, and most of the results obtained are irrelevant. Developing Web search mechanisms depends on addressing two important questions: (1) how to extract related Web pages of user interest, and (2) how to rank them according to relevance in a given set of potentially related Web pages. Measures of semantic similarity are necessary to evaluate the effectiveness of a Web search mechanism in finding and ranking results. In traditional approaches users provide manual assessments of relevance or semantic similarity. This is very difficult and expensive.

Semantic similarity between words is the study of an integral part of information retrieval and natural language processing. Semantic similarity is a concept whereby a set of terms within term lists are assigned a metric based on the likeness of their meaning. Measuring the semantic similarity between words is an important factor in different tasks on the web such as relation extraction, community mining, document clustering, automatic meta-data extraction and Web mining applications such as, community extraction, relation detection, and entity disambiguation. The main problem in information retrieval is to retrieve a set of documents that is semantically related to a given user query. Efficient estimation of semantic similarity between words is critical for numerous natural language processing jobs such as Word Sense Disambiguation (WSD), textual entailment and automatic text summarization. Semantic Similarity is the very challenging task when it comes to web, but in dictionary this problem is solved. For example, “apple” is frequently associated with computers on the Web. However, this sense of “apple” is not listed in most general-purpose dictionaries. A user, who searches for apple on the Web, may be interested in this sense of “apple” and not “apple” as a “fruit”.

Every day the new words are being added in web and the present words are given multiple meanings i.e. polysemous words. So manually maintaining these words is a very difficult task. We have proposed a Modified Pattern

Extraction Algorithm to estimate the semantic similarity between words or entities using Web search engines. Due to the massiveness of the web, it is impossible to analyze each document separately; hence Web search engines provide the perfect interface for this vast information. A web search engine gives two important information about the documents searched, Page count and Web Snippets. Page count of a query term will give an estimate of the number of documents or web pages that contain the given query term. A web snippet is one which appears below the searched documents and is a brief window of text that is searched around the query term in the document.

Page count between two objects is accepted normally as the relatedness measure between them. For example, the page count of the query *apple* AND *compute*r in Google is 977,000,000 whereas the same for *banana* AND *computer* is only 60,200,000 [as on 20 December 2012]. The more than 16 times more numerous page counts for *apple* AND *computer* indicate that *apple* is more semantically similar to *computer* than is *banana*. The drawbacks of page count is that it ignores the position of the two words that appear in the document, hence the two words may appear in the document but may not be related at all and page counts takes into account polysemous words of the query term, hence a word for example *Dhruv* will have the page counts for both *Dhruv* as the star of fortune and *Dhruv* as a name of the Helicopter.

Processing snippets is possible for measuring semantic similarity but it has the drawback of downloading a large number of web pages which consumes time, and all the search engine algorithms use a page rank algorithm, hence only the top ranked pages will have properly processed snippets. Hence there is no guarantee that all the information we need is present in the top ranked snippets.

Motivation: The search results returned by the most popular search engines are not satisfactory. Because of the vastly numerous documents and the high growth rate of the Web, it is time consuming to analyze each document separately. It is not uncommon that search engines return a lot of Web page links that have nothing to do with the user's need. Information retrieval such as search engines has the most important use of semantic similarity. It is the main problem to retrieve all the documents that are semantically related to the queried term by the user. Web search engines provide an efficient interface to this enormous information. Page counts and snippets are two useful information sources provided by most Web search engines. Hence, accurately measuring the semantic similarity between words is a very challenging job.

Contribution: We propose a Modified Pattern Extraction Algorithm to find the supervised semantic similarity measure between words by combining both page count method and web snippets method. We have used the four association measures including variants of Web Dice, Web Overlap Ratio, WebJaccard, and WebPMI to find semantic similarity between words in page count method using web search engines. The proposed approach goals to improve the correlation values, Precision, Recall, and F-measures, compared to the existing methods.

Organization: The remainder of the paper is organized as follows: Section 2 reviews the related work of the semantic similarity measures between words, Section 3 gives the problem definition, Section 4 explains the architecture of the system, and Section 5 gives the Modified Pattern Extraction Algorithm [MPEA]. The implementation and the results of the system are described in Section 6 and Conclusions are presented in Section 7.

2. Related Work

Semantic similarity between words has been important and challenging problem in data mining. Nowadays, World Wide Web (WWW) has become a vast collection of data and documents, with available information for every single user query.

Mehran Sahami et al., [1] presents a novel method for measuring the similarity between short text snippets by leveraging web search results to provide greater context for the short texts. In this paper, a similarity kernel function is defined and provided examples of its efficacy. The proposed method works on all kernel machines. A method is proposed for measuring the similarity between short text snippets that captures more of the semantic context of the snippets rather than simply measuring their term-wise similarity. Hsin-Hsi Chen et al., [2] proposed a web search with double checking model to explore the web as a live corpus. Instead of the simple web page counts and complex web page collection, the novel model, a Web Search with Double Checking (WSDC) is used to analyse snippets. They collect snippets for both the words X and Y from a Web Search Engine. They proceed by counting the occurrences of word X in the snippets for word Y and the occurrences of word Y in the snippets for word X. These values are then combined nonlinearly to compute the similarity between X and Y.

Rudi L. Cilibrasi et al., [3] has proposed the words and phrases acquire meaning from their relative semantics to other words and phrases and from the way they are used in society. It is a new concept of similarity between words and phrases based on information distance and

Kolmogorov complexity. The proposed method is applicable to all search engines and databases. This theory is then applied to construct a method to automatically extract similarity, the Google similarity distance, of words and phrases from the world-wide web using Google page counts. The main objective is to develop a new theory of semantic distance between a pair of objects is based on the background contents consisting of a database of documents. Relations between pairs of objects are extracted from the documents by just using the number of documents in which the objects occur, singly and jointly. Authors are introduced some notions underpinning the approach: Kolmogorov complexity, information distance, and compression-based similarity metric and a technical description of the Google distribution and the Normalized Google Distance (NGD). NGD uses page counts but does not take into account the context in which the words co-occur, it suffers from the drawbacks.

Dekang Lin et al., [4] proposed that, bootstrapping semantics from text is one of the greatest challenges in natural language learning. They defined a word similarity measure based on the distributional pattern of words. They demonstrate how their definition can be used to measure the similarity in a number of different domains. Many similarity measures have been proposed, such as information content, mutual information, Dice coefficient, cosine coefficient; Distance-based measurements, and feature contrast model. The similarity measure allows constructing a thesaurus using a parsed corpus. It is a new evaluation methodology for the automatically constructed thesaurus. The similarity between two objects is defined to be the amount of information contained in the commonality between the objects divided by the amount of information in the descriptions of the objects. They use a broad-coverage parser to extract dependency triples from the text corpus. Dependency triple consists of two words and the grammatical relationship between them in the input sentence. The description of a word ' w ' consists of the frequency counts of all the dependency triples that matches the pattern (w, *, *). The commonality between two words consists of the dependency triples that appear in the descriptions of both words. The main contribution of this paper is a new evaluation methodology for automatically constructed thesaurus. While previous methods rely on indirect tasks or subjective judgments, this method allows direct and objective comparison between automatically and manually constructed thesauri. Jian Pei et al., [5] proposed a projection based sequential pattern-growth approach for efficient mining of sequential patterns.

Jiang et. al., [6] combines a lexical taxonomy structure with corpus statistical information so that the semantic distance between nodes in the semantic space constructed by the taxonomy can be better quantified with the computational evidence derived from a distributional analysis of corpus data.

Philip Resnik et al., [7]-[8] presents measure of semantic similarity in an is-a taxonomy, based on the notion of information content. Experimental evaluation was performed using a large, independently constructed corpus, an independently constructed taxonomy, and previously existing and new human subject data and the results suggest that the measure performs encouragingly well and can be significantly better than the traditional edge-counting approach. Semantic similarity, as measured using information content, was shown to be useful in resolving cases of two pervasive kinds of linguistic ambiguity. In resolving coordination ambiguity, the measure was employed to capture the intuition that similarity of meaning is one indicator that two words are being conjoined. Suggestive results of a first experiment were bolstered by unequivocal results in a second study, demonstrating significant improvements over a disambiguation strategy based only on syntactic agreement. In resolving word sense ambiguity, the semantic similarity measure was used to assign confidence values to word senses of nouns within thesaurus-like groupings. A formal evaluation provided evidence that the technique can produce useful results but is better suited for semi-automated sense filtering than categorical sense selection. Application of the technique to a dictionary/thesaurus on the World Wide Web provides a demonstration of the method in action in a real-world setting.

Bollegala et al., [9] proposed a method which exploits the page counts and text snippets returned by a Web search engine. The proposed a novel approach is to compute semantic similarity using automatically extracted lexico-syntactic patterns from text snippets. They define various similarity scores for two given words P and Q, using the page counts for P, Q and P AND Q. These different similarity scores are integrated using support vector machines, to leverage a robust semantic similarity measure. Ming Li et al., [10] proposed a metric based on the non-computable notion of Kolmogorov computable distance and called it the similarity metric. General mathematical theory of similarity that uses no background knowledge or features specific to an application area. Hence it is, without changes, applicable to different areas and even to collections of objects taken from different areas.

Ann Gledson et al., [11] describes a simple web-based similarity measure which relies on page- counts only, can be utilized to measure the similarity of entire sets of words in addition to word-pairs and can use any web-service enabled search engine distributional similarity measure which uses internet search counts and extends to calculating the similarity within word-groups. The propensity of words to appear together in texts, known as their distributional similarity is an important part of Natural Language Processing (NLP). T Hughes et al., [12] proposed a method that presents the application of random walk Markov chain theory for measuring lexical semantic relatedness. Dekang Lin et al., [13] presented the information theoretic definition of similarity that is applicable as long as there is a probabilistic model. They demonstrate how their definition can be used to measure the similarity in a number of different domains. Many similarity measures have been proposed, such as information content, mutual information, Dice coefficient, cosine coefficient, Distance-based measurements, and feature contrast model. Vincent Schickel-Zuber et al., [14] present a novel approach that allows similarities to be asymmetric while still using only information contained in the structure of the ontology.

These literature surveys showed the fact that semantic similarity measures plays an important role in information retrieval, relation extraction, community mining, document clustering and automatic meta-data extraction. Thus there is need for more efficient system to find semantic similarity between words.

3. Problem Definition

Given two words X and Y, we model the problem of measuring the semantic similarity between X and Y, as a one of constructing a function semanticsim (X, Y) that returns a value in the range of 0 and 1. If X and Y are highly similar (e.g. synonyms), we expect semantic similarity value to be closer to 1, otherwise semantic similarity value to be closer to 0. We define numerous features that express the similarity between X and Y using page counts and snippets retrieved from a web search engine for the two words. Using this feature representation of words, we train a two-class Support Vector Machine (SVM) to classify synonymous and non-synonymous word pairs. Our objectives are:

i) To find the semantic similarity between two words and to increase the correlation value.

ii) To increases the Precision, Recall and the F-measure metrics of the system.

4. System Architecture

The outline of the proposed method for finding the semantic similarity using web search engine results is as shown in Fig. 1.

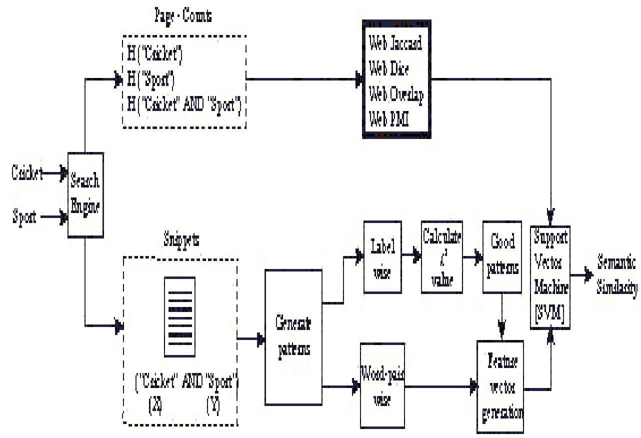


Fig. 1 System Architecture

When a query q is submitted to a search engine, web-snippets, which are brief summaries of the search results, are returned to the user. First, we need to query the word-pair in a search engine for example say we query “cricket” and “sport” in Google search engine. We get the page counts of the word-pair along with the page counts for individual words i.e. H (cricket), H(sport), H(cricket AND sport). These page counts are used to find the co-occurrence measures such as Web-Jaccard, Web-Overlap, Web-Dice and Web-PMI and store these values for future references. We collect the snippets from the web search engine results. Snippets are collected only for the query X and Y. Similarly, we collect both snippets and page counts for 200 word pairs. Now we need to extract patterns from the collected snippets using our proposed algorithm and to find the frequency of occurrence of these patterns.

We use the chi-square statistical method to find out the good patterns from the top 200 patterns of interest using the pattern frequencies. After that we integrate these top 200 patterns with the co-occurrence measures computed. If the pattern exists in the set of good patterns then we select the good pattern with the frequency of occurrence in the patterns of the word-pair else we set the frequency as 0. Hence we get a feature vector with 204 values i.e. the top 200 patterns and four co-occurrence measures values. We use a Sequential Minimal Optimization (or SMO) support vector machines (SVM) to find the optimal combination of page counts-based similarity scores and top-ranking patterns. The SVM is trained to classify the synonymous word-pairs and non-synonymous word pairs. We select synonymous word-pairs and Non-synonymous word-pairs and convert the output of SVM into a posterior probability. We define the semantic similarity between two words as the posterior probability if they belong to the synonymous-words (positive) class.

5. Algorithm

The proposed Modified Pattern Extraction Algorithm [MPEA] is used to measure the semantic similarity between words is as shown in the Table 1. Given two words A and B, we query a web search engine using the wildcard query A * * * * * B and download snippets. The * operator matches one word or none in a web page. Therefore, our wildcard query retrieves snippets in which A and B appears within a window of seven words. Because a search engine snippet contains 20 words on an average, and includes two fragments of texts selected from a document, we assume that the seven word window is sufficient to cover most relations between two words in snippets. The algorithm which is described in the Table 1 shows that how to retrieve the patterns and the frequency of the patterns.

The Modified Pattern Extraction Algorithm as described above yields numerous unique patterns. Of those patterns only 80% of the patterns occur less than10 times. It is impossible to train a classifier with such numerous parse patterns. We must measure the confidence of each pattern as an indicator of synonymy that is, most of the patterns have frequency less than 10 so it is very difficult to find the patterns which are significant so, we have to compute their confidence so as to arrive at the significant patterns. We compute chi-square value to find the confidence of each pattern.

The chi-square value is calculated by using the formula given below:

$$\chi^2 = \frac{(P+N)\,(p_v(N-n_v) - n_v(P-p_v))^2}{PN\,(p_v+n_v)(P+N-p_v-n_v)} \qquad (1)$$

Where,

P and N are the Total frequency of synonymous word pair patterns and non-synonymous word pair patterns, p_v and n_v are frequencies of the pattern v retrieved from snippets of synonymous and non-synonymous word pairs respectively.

Table 1. Modified Pattern Extraction Algorithm [MPEA]

Input: Given a set WS of word-pairs

Step 1: Extract snippets using Mozbar and store it in a text file.
Step 2: Read each snippet, remove all the non-ASCII character and store it in database.
Step 3: Retrieve each snippet, check for word pair A and B, when it encounter, then replace word pair 'A' or 'B' by 'X' and 'Y' with respectively.
Step 4: Scan the snippet until you encounter an X or Y, If you encounter an X go to Step 5. If you encounter a Y go to Step 11.
Step 5: Stop the sequence whenever you encounter a 'Y' or when the number of words encountered exceeds Maximum length L.
Step 6: Scan the sequence and replace the didn't, shouldn't, etc., by their full forms, did not, should not etc.
Step 7: Form the sub-sequences of the sequence such that each sub-sequence contains [X . . . Y . .].
Step 8: Compare each sub-sequence with the existing patterns, If its unique then Add it to the list of patterns and set its count to 1.
Step 9: If the sub-sequence is similar to existing pattern then frequency of the pattern increase by 1.
Step 10: If the length exceeds L then discard the pattern until you find an X or Y.
Step 11: If you encounter a Y replace the value of X with Y and Y with X and go to Step 5.

6. Implementation and Results

6.1 . Page-count-based Co-occurrence Measures

We compute four popular co-occurrence measures; Jaccard, Overlap (Simpson), Dice, and Point wise Mutual Information (PMI), to compute semantic similarity using page counts.

Web Jaccard coefficient between words (or multi-word phrases) A and B, is defined as:

$$\mathrm{WebJaccard}(A,B) = \begin{cases} 0 & \text{if } H(A \cap B) \le c, \\ \dfrac{H(A \cap B)}{H(A) + H(B) - H(A \cap B)} & \textit{otherwise} \end{cases} \qquad (2)$$

Web Overlap is a natural modification to the Overlap (Simpson) coefficient, is defined as:

$$\mathrm{WebOverlap}(A,B) = \begin{cases} 0 & \text{if } H(A \cap B) \le c, \\ \dfrac{H(A \cap B)}{\min(H(A), H(B))} & \text{otherwise} \end{cases} \qquad (3)$$

WebDice is defined as:

$$\mathrm{WebDice}(A,B) = \begin{cases} 0 & \text{if } H(A \cap B) \le c, \\ \dfrac{2H(A \cap B)}{H(A) + H(B)} & \text{otherwise} \end{cases} \quad (4)$$

Web PMI is defined as:

$$\mathrm{WebPMI}(A,B) = \begin{cases} 0 & \text{if } H(A \cap B) \le c, \\ \log_2\left(\dfrac{\frac{H(A \cap B)}{N}}{\frac{H(A)}{N}\frac{H(B)}{N}}\right) & \text{otherwise} \end{cases} \quad (5)$$

We have implemented this in Java programming language and used Eclipse as an extensible open source IDE (Integrated Development Environment) [15]. We query for A AND B and collect 500 snippets for each word pair and for each pair of words (A, B) store it in the database. By using the Pattern Retrieval algorithm, we retrieved huge patterns and select only top 200 patterns. After that we compare each of the top 200 patterns based on the chi-square values "χ^2" which are called as good patterns with the patterns generated by the given word pair. If the pattern extracted for the particular word pair is one among the good patterns, store that good pattern with a unique ID and store the frequency of this pattern as that of the pattern generated by the given word pair. If a pattern does not match then store it with a unique ID and with its frequency set as 0 and store it in the table.

To the same table, we add the four co-occurrence measure values of page counts are the Web- Jaccard coefficient, Web-Overlap, Web-Dice coefficient and Web-PMI which gives a table having 204 rows of unique ID, frequency and word pair ID. After that we normalize the frequency values by dividing the value in each tuple by the sum of all the frequency values. Now this 204-dimension vector is called the feature vector for the given word pair. Convert the feature vectors of all the word-pairs into a .CSV (Comma Separated Values) file. The generated .CSV file is fed to the SVM classifier which is inbuilt in Weka software [16]. This classifies the values and gives a similarity score for the word pair in between 0 and 1.

6.2 Test Data

In order to test our system, we selected the standard Miller-Charles dataset, which is having 28 word-pairs. The proposed algorithm outperforms by 89.8 percent of correlation value, as illustrated in Table 2.

The Fig. 2 shows the comparison of correlation value of our MPEA with existing methods.

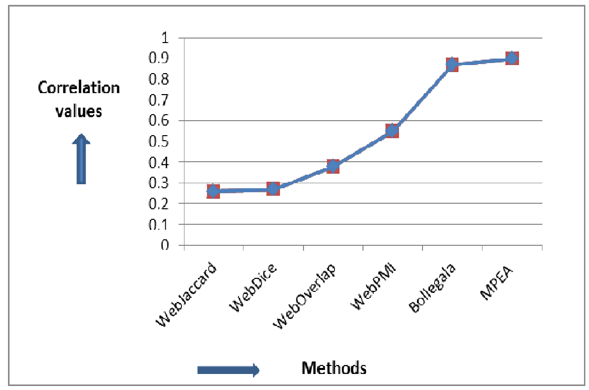


Fig. 2: Comparison of correlation value of MPEA with existing methods

The success of a search engine algorithm lies in its ability to retrieve information for a given query. There are two ways in which one might consider the return of results to be successful. Either we can obtain very accurate results or we can find many results which have some connection with the search query. In information retrieval, these are termed precision and recall, respectively [17].

The precision is the fraction of retrieved instances that are relevant, while Recall is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance. In even simpler terms, high Recall means that an algorithm returned most of the relevant results. High precision means that an algorithm returned more relevant results than irrelevant. A measure that combines both precision and recall is the harmonic mean of precision and recall which is called as the F-measure or balanced F-score.

$$F = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} \quad (6)$$

Table 2. Comparison of Correlation value of MPEA with existing methods

Word Pair	**Miller-Charles**	**Web Jaccard**	**Web Dice**	**Web Overlap**	**Web PMI**	**Bollegala**	**MPEA**
automobile-car	1.00	0.65	0.66	0.83	0.43	0.92	0.98
journey-voyage	0.98	0.41	0.42	0.16	0.47	1	0.93
gem-jewel	0.98	0.29	0.3	0.07	0.69	0.82	0.98
boy-lad	0.96	0.18	0.19	0.59	0.63	0.96	0.95
coast-shore	0.94	0.78	0.79	0.51	0.56	0.97	0.98
asylum-madhouse	0.92	0.01	0.01	0.08	0.81	0.79	0.86
magician- wizard	0.89	0.29	0.03	0.37	0.86	1	0.94
midday-noon	0.87	0.1	0.1	0.12	0.59	0.99	0.71
furnace-stove	0.79	0.39	0.41	0.1	1	0.88	0.94
food-fruit	0.78	0.75	0.76	1	0.45	0.94	0.97
bird-cock	0.77	0.14	0.15	0.14	0.43	0.87	0.91
bird-crane	0.75	0.23	0.24	0.21	0.52	0.85	0.65
implement-tool	0.75	1	1	0.51	0.3	0.5	0.74
brother-monk	0.71	0.25	0.27	0.33	0.62	0.27	0.54
crane- implement	0.42	0.06	0.06	0.1	0.19	0.06	0.18
brother-lad	0.41	0.18	0.19	0.36	0.64	0.13	0.68
car-journey	0.28	0.44	0.45	0.46	0.2	0.17	0.26
monk-oracle	0.27	0	0	0	0	0.8	0.7
food-rooster	0.21	0	0	0.41	0.21	0.02	0.36
coast-hill	0.21	0.96	0.97	0.26	0.35	0.36	0.18
forest-graveyard	0.2	0.06	0.06	0.23	0.49	0.44	0.77
monk-slave	0.12	0.17	0.18	0.05	0.61	0.24	0.08
coast-forest	0.09	0.86	0.87	0.29	0.42	0.15	0.07
lad-wizard	0.09	0.06	0.07	0.05	0.43	0.23	0.03
cord-smile	0.01	0.09	0.1	0.02	0.21	0.01	0.03
glass-magician	0.01	0.11	0.11	0.4	0.6	0.05	0.04
rooster-voyage	0	0	0	0	0.23	0.05	0.06
noon-string	0	0.12	0.12	0.04	0.1	0	0.03
Correlation	**1.0**	**0.26**	**0.27**	**0.38**	**0.55**	**0.87**	**0.898**

Table 3. Precision, Recall and F-measure values for both Synonymous and Non-synonymous classes.

Class	Precision	Recall	F-Measure
Synonymous	0.9	0.947	0.923
Non-synonymous	0.83	0.714	0.769

In this paper, F-measure is computed based on the precision and recall evaluation metrics. The results are better than the previous algorithms, the Table 4 shows that the comparison of Precision, Recall and F-measure improvement of the proposed Algorithm.

Table 4. Comparison of Precision, Recall and F-measure values of MPEA with previous method

Method	Precision	Recall	F-Measure
Bollegala	0.7958	0.804	0.7897
MPEA	0.9	0.947	0.923

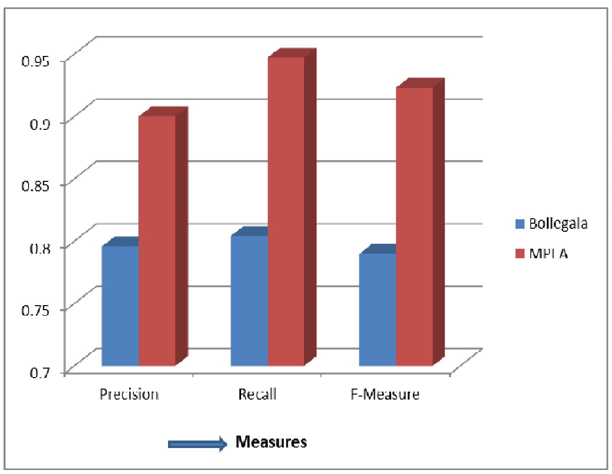


Fig. 3. Comparison of Precision, Recall and F-measure values of MPEA with previous method.

The Fig. 3 shows the comparison of Precision, Recall and F-measure values of PRA with previous method on the Miller- Charles dataset proves that our results better than the previous methods. We have tested our algorithm on the Word Similarity dataset, which contains 353 word pairs and comparison result of PRA with existing methods on the Word Similarity is as shown in the Table 5.

Table 5. Comparison result of MPEA with existing methods on the Word Similarity dataset.

Method		Correlation
Jarmasz [18]	Wordnet	0.35
Wikirelate! [19]	Wikipedia	0.48
Hughes & Ramage [20]	Wordnet	0.55
Jarmasz [18]	Roget's	0.55
Finkelstein et al. [21]	Corpus+Wordnet	0.56
Gabrilovich [22]	ODP	0.65
Gabrilovich [22]	Wikipedia	0.75
Bollegala [9]	Web snippets+ Page counts	0.74
MPEA (Proposed)	Page counts + Web snippets	0.753

7. Conclusions

Semantic Similarity measures between words plays an important role in information retrieval, natural language processing and in various tasks on the web. We have proposed a Modified Pattern Extraction algorithm to extract numerous semantic relations that exist between two words and the four word co-occurrence measures were computed using page counts. We integrate the patterns and co-occurrence measures to generate a feature vector. These feature vectors are fed to a 2- Class SVM to classify the data into synonymous and non-synonymous classes. We compute the posterior probability for each word-pair which is the similarity score for that word-pair. The proposed algorithm outperforms by 89.8 % of correlation value for Miller-Charles dataset and 75.3% of correlation value for Word similarity dataset. The Precision, Recall and F-measure values are improved compared to previous methods.

References

[1] M Sahami and T Heilman, "A Web-based Kernel Function for Measuring the Similarity of Short Text Snippets", 15th International Conference on World Wide Web, Year 2006, pp. 377-386.
[2] H Chen , M Lin and Y Wei , "Novel Association Measures using Web Search with Double Checking", International Committee on Computational Linguistics, Year 2006, pp. 1009-1016.
[3] R Cilibrasi and P Vitanyi, "The Google Similarity Distance", IEEE Transactions on Knowledge and Data Engineering, Vol. 19, No. 3, Year 2007, pp. 370-383.
[4] Lin D, "Automatic Retrieval and Clustering of Similar Words", International Committee on Computational Linguistics and the Association for Computational Linguistics, Year 1998, pp. 768-774.

[5] J Pei, J Han, B Mortazavi-Asi, J Wang, H Pinto, Q Chen, U Dayal and M Hsu, "Mining Sequential Patterns by Pattern growth: the Prefix Span Approach", IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 11, Year 2004, pp. 1424-1440.
[6] Jay J Jiang and David W Conrath, "Semantic Similarity based on Corpus Statistics and Lexical Taxonomy", International Conference Research on Computational Linguistics.
[7] P Resnik, (1995) "Using Information Content to Evaluate Semantic Similarity in a Taxonomy", 14th International Joint Conference on Artificial Intelligence, Vol. 1, Year 1995, pp. 24-26.
[8] P Resnik, "Semantic Similarity in a Taxonomy: An Information based Measure and its Application to problems of Ambiguity in Natural Language", Journal of Artificial Intelligence Research, Vol. 11, Year 1999, pp. 95-130.
[9] Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka, "A Web Search Engine-based Approach to Measure Semantic Similarity between Words", IEEE Transactions on Knowledge and Data Engineering , Vol. 23, No.7, Year 2011, pp.977-990.
[10] Ming Li, Xin Chen, Xin Li, Bin Ma, Paul M and B Vitanyi, "The Similarity Metric", IEEE Transactions on Information Theory, Vol. 50, No. 12, Year 2004, pp. 3250-3264.
[11] Ann Gledson and John Keane,"Using Web Search Results to Measure Word-Group Similarity", 2nd International Conference on Computational Linguistics),Year 2008, pp. 281-288.
[12] T Hughes and D Ramage, "Lexical Semantic Relatedness with Random Graph Walks", Conference on Empirical Methods in Natural Language Processing Conference on Computational Natural Language Learning, (EMNLP-CoNLL07), Year 2007, pp. 581-589.
[13] D Lin, "An Information-Theoretic Definition of Similarity", 15th International Conference on Machine Learning, Year 1998, pp. 296-304.
[14] Schickel-Zuber V and Faltings B, (2007) "OSS: A Semantic Similarity Function Based on Hierarchical Ontologies", International Joint Conference on Artificial Intelligence, pp. 551-556.
[15] http://onjava.com/onjava/2002/12/11/eclipse.html
[16] www.cs.waikato.ac.nz/ml/**weka/**
[17] C N Pushpa, J Thriveni, K R Venugopal and L M Patnaik, "Enhancement of F-measure for Web People Search using Hashing Technique", International Journal on Information Processing (IJIP), Vol. 5, No. 4, Year 2011, pp. 35-44.
[18] M. Jarmasz, "Roget's Thesaurus as a Lexical Resource for Natural Language Processing", University of Ottowa, Technical Report, Year 2003.
[19] M. Strube and S. P. Ponzetto, "Wikirelate! Computing Semantic Relatedness using Wikipedia, "In the Proceedings of AAAI'06, Year 2006, pp. 1419–1424.
[20] T. Hughes and D. Ramage, "Lexical Semantic Relatedness with Random Graph Walks," in Proceedings of EMNLP-CoNLL'07, Year 2007, pp.581–589.
[21] L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, z. Solan, G. Wolfman, and E. Ruppin, "Placing Search in Context: The Concept Revisited," ACM Transactions on Information Systems, Vol. 20,Year 2002, pp. 116–131.
[22] E. Gabrilovich and S. Markovitch "Computing Semantic Relatedness uses Wikipedia-based explicit Semantic Analysis," in Proceedings. of IJCAI'07, Year 2007, pp. 1606–1611.

Pushpa C N has completed Bachelor of Engineering in Computer Science and Engineering from Bangalore University, Master of Technology in VLSI Design and Embedded Systems from Visv esvaraya Technological University. She has 13 years of teaching experience. Presently she is working as Assistant Professor in Department of Computer Science and Engineering at UVCE, Bangalore and pursuing her Ph.D in Semantic Web.

Thriveni J has completed Bachelor of Engineering, Masters of Engineering and Doctoral Degree in Computer Science and Engineering. She has 4 years of industrial experience and 18 years of teaching experience. Currently she is an Associate Professor in the Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore. Her research interests include Networks, Data Mining and Biometrics

Venugopal K R is currently the Principal, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. He obtained his Bachelor of Engineering from University Visvesvaraya College of Engineering. .He received his Master's degree in Computer Science and Automation from Indian Institute of Science Bangalore. He was awarded Ph.D. in Economics from Bangalore University and Ph.D. in Computer Science from Indian Institute of Technology, Madras. He has a distinguished academic career and has degrees in Electronics, Economics, Law, Business Finance, Public Relations, Communications, Industrial Relations, Computer Science and Journalism. He has authored 31 books on Computer Science and Economics, which include Petrodollar and the World Economy, C Aptitude, Mastering C, Microprocessor Programming, Mastering C++ and Digital Circuits and Systems etc. During his three decades of service at UVCE he has over 350 research papers to his credit. His research interests include Computer Networks, Wireless Sensor Networks, Parallel and Distributed Systems, Digital Signal Processing and Data Mining.

L M Patnaik is a Honorary Professor in Indian Institute of Science, Bangalore. During the past 35 years of his service at the Institute he has over 700 research publications in refereed International Journals and refereed International Conference Proceedings. He is a Fellow of all the four leading Science and Engineering Academies in India; Fellow of the IEEE and the Academy of Science for the Developing World. He has received twenty national and international awards; notable among them is the IEEE Technical Achievement Award for his significant contributions to High Performance Computing and Soft Computing. His areas of research interest have been Parallel and Distributed Computing, Mobile Computing, CAD for VLSI circuits, Soft Computing and Computational Neuroscience.

Plastic Surgery Face Recognition: A comparative Study of Performance

Rehab M. Ibrahim[1], Prof. F.E.Z Abou-Chadi[2] and Prof. A. S. Samra[3]

[1] Dept. of Electronics and Communications Engineering, Mansoura University, Faculty of Engineering Mansoura – Egypt

[2] Dept. of Electronics and Communications Engineering, Mansoura University, Faculty of Engineering Mansoura – Egypt

[3] Dept. of Electronics and Communications Engineering, Mansoura University, Faculty of Engineering Mansoura – Egypt

Abstract

The paper presents a comparative study of performance for face recognition algorithms in order to select the algorithms that have the highest performance and overcome the problems faced in recognition due to plastic surgery. A plastic surgery database that contains face images with different types of surgeries is used.. The work reports the performance evaluation of eleven photometric illumination techniques, five histogram normalization techniques and four feature extraction techniques. A minimum distance classifier has been adopted and four distance similarity measures were used. Face identification/verification techniques are considered in the present work. Experimental results were carried out and it can be concluded that for face identification, the best illumination technique is gradient-face (GRF) normalization technique, histruncate histogram normalization and the best feature extraction technique is gabor kernel fisher analysis GKFA. For face verification, the best illumination technique is gradient-face (GRF) normalization technique, and the best feature extraction technique is gabor principal component analysis (GPCA). For both face identification/verification the minimum distance classifier using Mahalanobis Cosine (MAHCOS) distance gives the best results.

Keywords: *photometric illumination, histogram normalization, holostic feature extraction techniques, mininmum distance classifier.*

1. Introduction

Face recognition technology has become one of the most important biometric technologies, for its non-intrusive nature and its potential applications like personal identification, security access control, surveillance systems, telecommunications, digital libraries, human-computer interaction , military and so on [1].

Five factors can significantly affect the performance of face recognition system: illumination, pose, expression, occlusion and ageing [2]. Another challenging factor that is not much considered before is Plastic surgery[3].

Plastic surgery is generally used for improving the facial appearance, for example, removing birth marks, moles, scars and correcting disfiguring defects. However, it can also be misused by individuals to conceal their identities with the intent to commit fraud or evade law enforcement. Face recognition after plastic surgery can lead to rejection of genuine users or acceptance of impostors. While face recognition is a well studied problem in which several approaches have been proposed to address the challenges of illumination [4], pose [1,5], expression [4], aging [2] and disguise [6, 7], the use of plastic surgery introduces a new challenge to designing future face recognition systems.

In general, plastic surgery can be classified into two distinct categories.

1. Disease Correcting Local Plastic Surgery (Local Surgery): This is the kind of surgery in which an individual undergoes local plastic surgery for correcting defects, anomalies, or improving skin texture.. Such changes may cause errors in automatic face recognition and degrade the system performance.

2. Plastic Surgery for Reconstructing Complete Facial Structure (Global Surgery): Apart from local surgery, plastic surgery can be done to completely change the facial structure which is known as full face lift. This medical procedure is recommended for cases such as patients with fatal burn or trauma. In this type of surgery, the appearance, texture and facial features of an individual are reconstructed and are usually not the same as the original face.

The main objective of the work is to investigate the performance of existing face recognition algorithms on a face database that contains images before and after

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

surgery. In general, any face recognition system consists of four stages: capturing the face images, segmentation and preprocessing of the face, face feature extraction, and classification. The present work presents a detailed performance evaluation for algorithms most commonly used in each stage. In the present study considers the performance evaluation of eleven photometric illumination techniques, five histogram normalization techniques, four feature extraction techniques and a minimum distance classifier using four distance similarity measures .

The organization of this paper as follows: Section II presents the plastic surgery database used in the present work and the preprocessing steps carried out.. Section III presents the photometric illumination techniques utilized. Section IV presents the histogram normalization techniques. The four selected feature extraction techniques and the classifier are described in Sections V and VI, respectively. Section VII presents performance evaluation metrics. Experimental results are reported in Section VIII. Finally a conclusion is drawn in Section IX.

2. Plastic Surgery Database

The plastic surgery database [8] used in the present work consists of 1800 full frontal face images of 900 subjects. Table 1 summarizes the details of the plastic surgery database, for each individual, there are two frontal face images (before and after plastic surgery). The database contains 519 image pairs corresponding to local surgeries, 381 pairs of global surgery, the local surgery images consists of 194 pairs of nose surgery (Rhinoplasty), 101 pairs of eye-lid-lift-surgery (blepharoplasty) images, 74 pairs of ear-surgery (otoplasty) images , 56 cases of brow-lift (forehead-lift) images, 32 pairs of laser-skin (resurfacing) images, 18 pairs of fat-injections images and 44 pairs of Others (Mentoplasty, Malar Augmentation, Craniofacial, Lip augmentation) images. While the global surgery images consists of 320 pairs of Rhytidectomy (face lift) images and 60 pairs of skin peeling (Skin resurfacing) images. Examples of images from the database are shown in Fig. 1. The images need to be preprocessed for accurate face detection and extraction, resizing as well as reducing the influence of reflections and artifacts. Hence, for each image, the face region in the image is detected and extracted, then a resizing procedure is performed.

Two pre-processing steps were applied to the images:

1. A background removal process where the face region in the image is detected and cropped to extract the face from the surrounding background.
2. Image size normalization where the size of the detected and cropped face image is set to200 x200 as shown in Fig.2.

Table 1: Plastic Surgery Database [8]

Type	Plastic Surgery Procedure	Number of subjects
Global	Face-Lift Surgery (rhytidectomy)	321
	Skin peeling (skin resurfacing)	60
Local	Nose Surgery (Rhinoplasty)	194
	Eye-Lid-Lift surgery (blepharoplasty)	101
	Ear-Surgery (otoplasty)	74
	Brow-Lift (Forehead)	56
	Laser-skin (resurfacing)	32
	Fat-injections	18
	Others(Mentoplasty, Lip augmentation)	44

Fig. 1. Examples of images from the plastic surgery database before and after plastic surgery.

Fig.2. a)Example of face images before processing , b)Face images after cropping and c) resizing

IJCSI
www.IJCSI.org

3. Photometric Illumination Normalization Techniques

As the appearance of a face image is severely affected by illumination conditions that hinder the face recognition process, Photometric normalization techniques are used to compensate for illumination induced appearance changes.
Eleven photometric illumination techniques were selected and their performance was evaluated in the present study. These are [9]: the single-scale-retinex algorithm (SSR), the multi-scale-retinex algorithm (MSR), the single-scale self quotient image (SSQ), the multi-scale self quotient image (MSQ), the homomorphic-filtering based normalization technique (HOMO), a wavelet-based normalization technique (WAV), the DCT- based normalization technique (DCT), a normalization technique based on steerable filters (SF), the Gradient-faces approach (GRF), wavelet-denoising-based normalization technique (WD), adaptive single scale retinex technique (ASR).

4. Histogram Normalization Techniques

Five histogram normalization techniques were considered. These are [9]: Rank Histogram (RH), Histogram Truncation and Stretching (HT), Normal Distribution (ND), Lognormal Distribution (LN), Exponential Distribution (EX).

5. Feature Extraction Techniques

In the present work, four techniques were considered for feature extraction. The principal component analysis (PCA) which is one of the most powerful and popular linear methods for dimensionality reduction and feature extraction [10], the kernel principal component analysis (KPCA) [11], the kernel Fisher analysis (KFA) [12], and the Gabor face representation [13-17] were selected and their performance on the face verification/identification was investigated.

6. Classifier

The *k*-Nearest Neighbor classifier (*k* =1) was adopted for classification [18], four distance similarity measures were used: Euclidean distance (EUC), Cosine distance (COS), city block (CTB) distance and the Mahalanobis Cosine (MAHCOS) distance.

7. Performance Evaluation

Assessment of the different techniques carried out in the present work was performed as follows. For the verification experiments, Detection Error Trade-off (*DET*) curves, which plot the false acceptance error (*FAR*) against the false rejection error rate (*FRR*) using different thresholds were used as well as the half total error rate (*HTER*). The FAR and FRR are defined as follows [20]:

$$FAR = \frac{n_{ai}}{n_i} 100\% \quad (1)$$

$$FRR = \frac{n_{rc}}{n_c} 100\% \quad (2)$$

where *HTER* is given by:

$$HTER= 0.5(FAR+FRR) \quad (3)$$

where n_{ai}: number of accepted impostor, n_i: number of all impostor identity claims made, n_{rc}: number of rejected geniune, and n_c: number of all genuine identity claims made.
For the identification experiments, results are provided in the form of the rank-one recognition rate (*ROR*):

$$ROR = \frac{n_{si}}{n_s} 100\% \quad (4)$$

where n_{si}: number of images successfully assigned to the right identity and n_s is the overall number of images trying to assign an identity to.

8. Experimental Results

Four experiments are conducted in this work and the results are reported here.

Experiment1: Without Any Processing.
In this experiment, the performance of the three holistic feature extraction techniques (PCA, KPCA and KFA) was evaluated. The three holistic feature extraction techniques were performed directly on the preprocessed image sets without any illumination normalization technique. The HTER and ROR rates are depicted in Table 2 using MAHCOS distance which gives the highest rates. Fig.3. shows the DET curves using MAHCOS distance. The results obtained for both verification/ identification systems showed that PCA gives the best recognition rates in case of Blepharoplasty, Otoplasty, fat-injections and Rhtidectomy plastic surgeries using MAHACOS distance measure, while KFA gives the best rates in case of resurfacing and forehead-lift plastic surgeries using MAHACOS distance measure. An overall best rates are obtained using PCA and MAHCOS distance measure.

Experiment 2: Using Gabor Filters.
The effect of using Gabor filters was investigated.. In this experiment Gabor representation of

Table 2: HTER and ROR rates obtained from Experiment 1 using MAHCOS. (PS= Plastic Surgery Type) and (HFE = Holistic Feature Extraction).

	HFE \ PS	Resur-facing	Forehead-Lift	Blepharoplasty	Otoplasty	Fat-Injections	Rhytidectomy
HTER	PCA	0.080	0.072	0.090	0.172	0.258	0.161
	KPCA	0.184	0.152	0.101	0.223	0.400	0.149
	KFA	0.077	0.068	0.093	0.182	0.270	0.187
ROR	PCA	68.9%	60.4%	59.6%	49.3%	43.8%	32%
	KPCA	48.3%	52.8%	54.3%	47.9%	37.5%	34%
	KFA	68.9%	60.4%	58.5%	47.9%	43.8%	25%

the face image was calculated followed by one of the holistic feature extraction techniques without any illumination normalization, HTER and ROR rates are depicted in Table3- MAHCOS distance (highest performance). While Fig.4. shows the DET curves for the best results obtained using MAHCOS distance. The results obtained for both verification and identification systems showed that the best rates are obtained using GPCA in case of resurfacing, Blepharoplasty, Otoplasty and Rhytidectomy using MAHCOS distance, while KFA gives best rates in case of forehead-lift and fat-injections using MAHCOS distance measure. The overall highest rates have been obtained using GPCA feature extraction technique and MAHCOS distance.

Experiment3: Using Photometric Illumination Techniques

Experiment 3 was carried out to investigate the effect of using photometric illumination normalization techniques on face recognition, a photometric illumination normalization technique was first applied to each set of the six plastic surgery datasets. Then Gabor representation of the face image calculated and then followed by the use of holistic feature extraction technique. Table 4 and Table 5 shows the HTER and ROR results respectively, of performing 11 photometric illumination techniques: GRF, WAV, MSQ, SSQ, DCT, SF, HOMO, MSR, WD, SSR, ASR each time with one of the three holistic feature extraction techniques and MAHCOS distance which gives the highest performance. The results obtained for verification system shows that the best illumination normalization and feature extraction techniques are GRF and GPCA respectively, in case of resurfacing and rhytidectomy using MAHCOS distance measure, while GRF and GKFA are best illumination normalization and feature extraction respectively, in case of forhead-lift, blepharoplasty, otoplasty and fat-injections using MAHCOS distance measure. For verification system, the best illumination normalization and feature extraction techniques are GRF and GPCA respectively, in case of resurfacing, blepharoplasty, fat-injections and rhytidectomy using MAHCOS distance measure, while GRF and GKFA are best illumination normalization and feature extraction respectively, in case of forhead-lift and otoplasty using MAHCOS distance measure.

Experiment4: Using Histogram normalization Techniques

Experiment 4 investigates the effect of performing five histogram normalization techniques: normal, rank, exponential, lognormal and Histruncate histogram respectively. It consists of four stages: Firstly, photometric illumination normalization was performed. Secondly, Gabor representation was calculated from the resulted images, followed by histogram normalization then the holistic feature extraction technique was applied. The results obtained for verification system shows that rank histogram is the best normalization technique in case of forehead-lift, fat-injections and rhytidectomy plastic surgery, lognormal histogram in case of blepharoplasty and otoplasty while exponential histogram is the best choice in case of resurfacing. For verification system , rank histogram is the best normalization technique in case of forehead-lift, blepharoplasty and rhytidectomy plastic surgery, histruncate histogram in case of otoplasty plastic surgery. In case of resurfacing plastic surgery, normal, rank and lognormal histograms give the same best result.

Table 3: Experiment 2: HTER and ROR rates using MAHCOS. (PS= Plastic Surgery Type) and (HFE = Holistic Feature Extraction).

	HFE \ PS	Resur-f-acing	Foreh-ead-Lift	Bleph-aroplasty	Otopl-asty	Fat-Inject-ions	Rhyti-decto-my
HTER	PCA	0.015	0.071	0.036	0.040	0.108	0.067
	KPC-A	0.016	0.088	0.056	0.085	0.158	0.087
	KFA	0.016	0.071	0.046	0.055	0.104	0.079
ROR	PCA	89.6%	79.2%	73.4%	65.7%	68.7%	68.6%
	KPC-A	89.6%	69.8%	71.2%	61.6%	68.7%	63.4%
	KFA	89.6%	81.1%	72.3%	63.0%	68.7%	21.5%

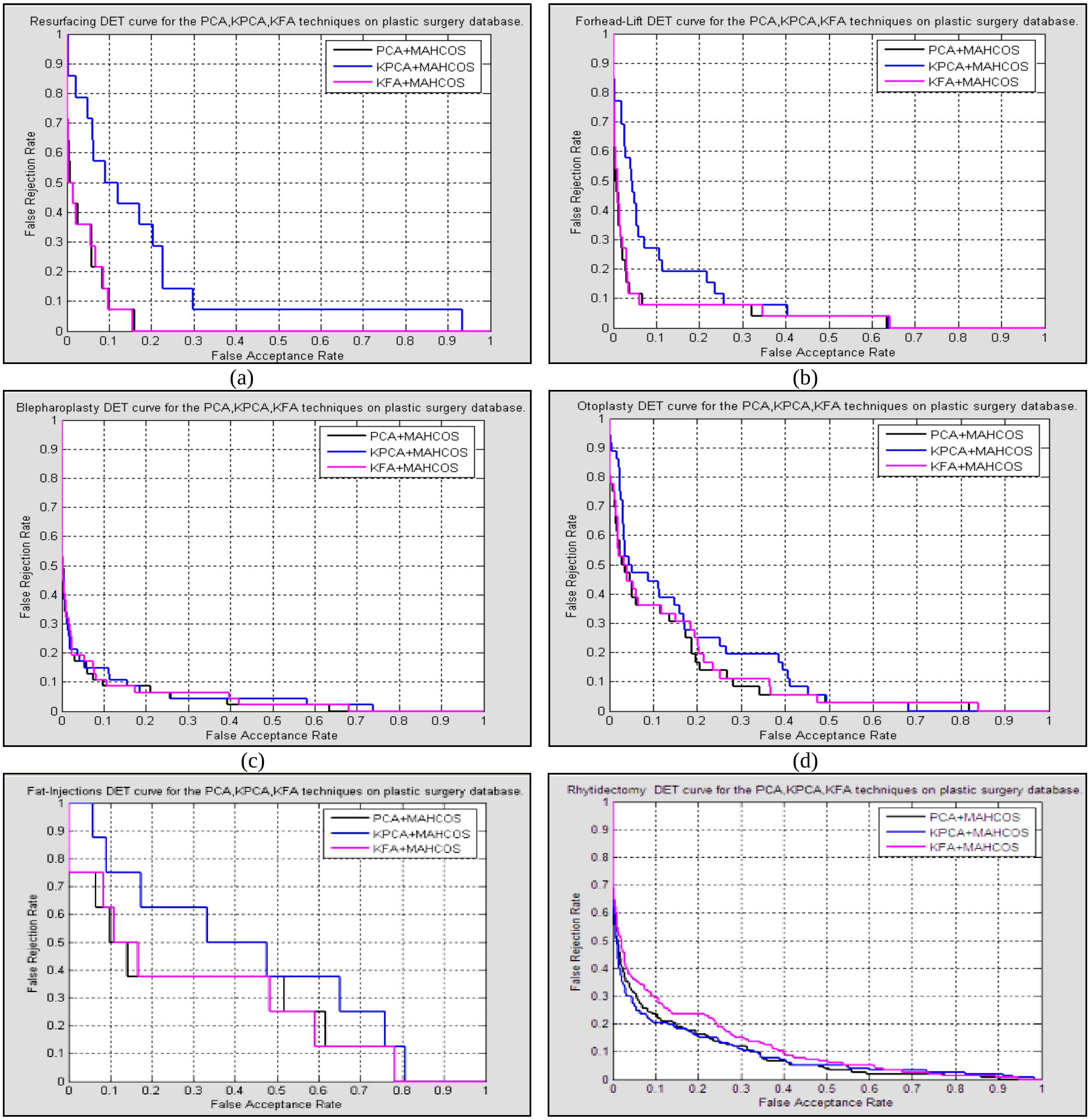


Fig.3. DET plots of the 3 holistic algorithms(a)Resurfacing, (b)Forehead- Lift, (c) Blepharoplasty, (d)Otoplasty,(e)Fat-injection and (f) Rhytidectomy surgery images using MAHCOS distance.

9. Conclusions

This paper has been concerned with the performance evaluation of the existing identification/verification approaches used in the recognition of human face. Specifically, the main objective is to identify the best face identification/verification techniques that give the highest performance for human face images for different plastic surgery.

The investigated techniques for each stage are not novel techniques; they are previously proposed and studied in literature. The selected techniques were applied onto a Plastic surgery database representing 6 types of facial

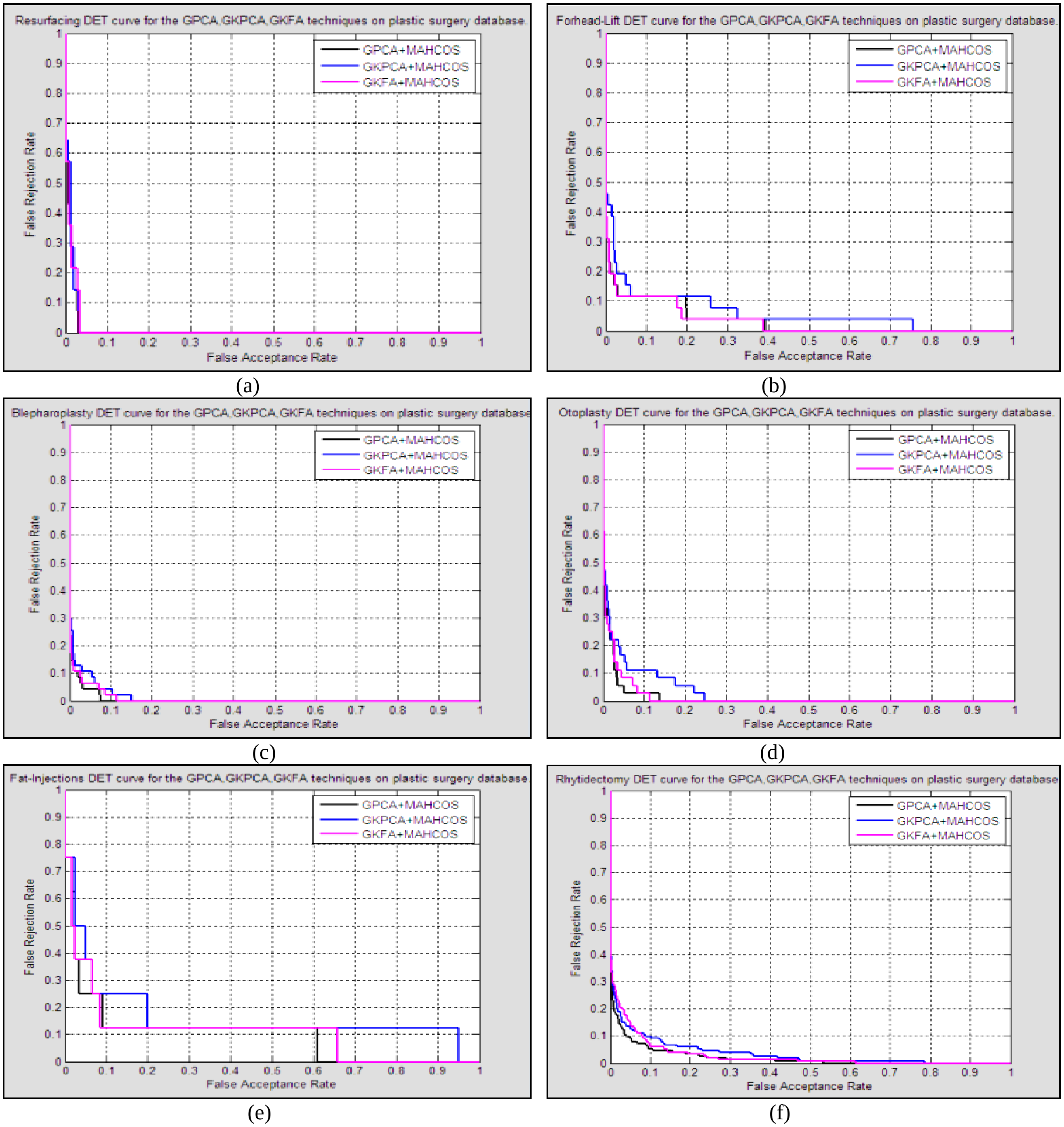


Fig.4. DET plots of the GPCA, GKPCA and GKFA (a)Resurfacing, (b)Forehead- Lift, (c) Blepharoplasty, (d)Otoplasty,(e)Fat-injection and (f) Rhytidectomy surgery images using MAHCOS distance.

plastic surgery: Brow-lift (Forehead-Lift), Eye-Lid-Left (blepharoplasty), ear (Otoplasty) ,Fat-Injections and Face-Lift (Rhytidectomy) surgery.

A face preprocessing step was first performed where the face region in the image is detected and extracted. Then a resizing procedure was performed.

As for the feature extraction stage. four feature extraction techniques were selected. These are: the principal component analysis (PCA), the kernel principal component analysis (KPCA), the kernel fisherfaces analysis (KFA) and the Gabor technique.

In general, identification and verification modes are two main goals of every security system based on the needs of the environment. In the verification mode, the system checks if the user data that was entered is correct or not but in the identification stage, the system tries to discover who the subject is without any input information. Hence, verification is a one-to-one search but identification is a one-to-many comparison. For verification purposes, a decision with high confidence level is made to verify whether the user is authenticated or imposter. To identify a person's identity, the face image needs to be matched with the stored images. Therefore, the nearest neighbor classifier was adopted for both identification and verification modes of the proposed system using four distance similarity measures. These are: Euclidean distance, cosine distance, city block distance, and Mahalanobis Cosine distance similarity measure.

The performance metrics adopted in the present study are as follows. For face verification experiments, the false acceptance error rate (FAR), false rejection error rate (FRR) and the half total error rate (HTER)were computed, as well as Detection Error Trade-off (DET) curves. However, for the identification experiments results were compared in the form of the so-called rank one recognition rate (ROR).

Four experimental studies were carried out in the present comparative study. In the first experiment, the three holistic feature extraction techniques (PCA, KPCA and KFA) were applied directly on facial plastic surgery. The results obtained for both verification/ identification systems showed that PCA gives the best recognition rates in case of Blepharoplasty, Otoplasty, fat-injections and Rhtidectomy plastic surgeries using MAHACOS distance measure, while KFA gives the best rates in case of resurfacing and forehead-lift plastic surgeries using MAHACOS distance measure. An overall best result is obtained using PCA and MAHCOS distance measure.

Extracting the Gabor features prior to performing holistic approaches in the second experiment gives a noticeable improvement in the recognition rates on all facial plastic surgery sets. The results obtained for both verification and identification systems showed that the highest recognition rates and minimum error rate are obtained using GPCA in case of resurfacing, Blepharoplasty, Otoplasty and Rhytidectomy using MAHCOS distance, while KFA gives best rates in case of forehead-lift and fat-injections using MAHCOS distance measure. An overall best results have been obtained using GPCA feature extraction technique and MAHCOS distance.

In an attempt to reduce the effect of illumination conditions that hinder the face recognition process, two experiments were carried out to investigate the effect of using photometric illumination normalization and histogram normalization on facial plastic surgery.

This effect was investigated in the third and fourth experiments. The third experiment investigates the effect of eleven photometric illumination techniques (Gradient (GRF), Wavelet Based (WAV), Multi-Scale Self Quotient (MSQ), Single Scale Self-Quotient (SSQ) , DCT Based (DCT), Steerable Filter Based (SF), Homomorphic Filtering Based (HOMO), Multi-Scale Retinex (MSR), Wavelet Denoising Based (WD), Single Scale Retinex (SSR), Adaptive Single Scale Retinex (ASSR)) using Gabor face representation before using the three holistic feature extraction techniques on facial plastic surgery. It has been found that an overall best rates were obtained, For verification system using GRF illumination

TABLE 4. HTER rates resulted from Experiment 3 using MAHCOS. (PS= Plastic Surgery Type),(HFE = Holistic Feature Extraction) and (PIT= Photometric Illumination Technique) .

PIT	PS / HFE	Resurf-acing Surgery	Forehead-Lift	Blepharoplasty	Otopla-sty	Fat-Injections Surgery	Rhytid-ectomy
GRF	PCA	0.0140	0.0592	0.0142	0.0469	0.083	0.0444
	KPCA	0.0536	0.0876	0.0304	0.0801	0.162	0.0919
	KFA	0.0140	0.0473	0.0137	0.0471	0.066	0.0487
WAV	PCA	0.0166	0.0666	0.0245	0.0592	0.100	0.0603
	KPCA	0.0395	0.1076	0.0358	0.1047	0.141	0.0747
	KFA	0.0140	0.0643	0.0341	0.0617	0.095	0.0687
MSQ	PCA	0.0242	0.0839	0.0487	0.0733	0.108	0.0830
	KPCA	0.0217	0.0954	0.0633	0.1209	0.175	0.0959
	KFA	0.0293	0.0917	0.0542	0.0826	0.104	0.1051
SSQ	PCA	0.0408	0.0740	0.0528	0.0835	0.112	0.0854
	KPCA	0.0255	0.0847	0.0645	0.1368	0.170	0.1012
	KFA	0.0434	0.0843	0.0620	0.0885	0.112	0.1186
DCT	PCA	0.0816	0.1006	0.0673	0.1121	0.229	0.1208
	KPCA	0.0816	0.1087	0.0729	0.1593	0.225	0.1209
	KFA	0.0842	0.0962	0.0824	0.1196	0.229	0.1686
SF	PCA	0.0536	0.1039	0.0581	0.0891	0.158	0.0992
	KPCA	0.0472	0.1254	0.0686	0.1308	0.150	0.1306
	KFA	0.0536	0.1084	0.0599	0.0912	0.158	0.1214
HOMO	PCA	0.0714	0.1043	0.0501	0.0741	0.166	0.0961
	KPCA	0.0829	0.1398	0.0503	0.1190	0.212	0.1156
	KFA	0.0689	0.1087	0.0556	0.0870	0.162	0.1310
MSR	PCA	0.1237	0.1302	0.0844	0.1921	0.175	0.1898
	KPCA	0.1454	0.1609	0.1197	0.2133	0.175	0.1803
	KFA	0.1237	0.1379	0.0984	0.2118	0.183	0.2893
WD	PCA	0.0485	0.0965	0.0755	0.1005	0.216	0.1065
	KPCA	0.0842	0.1187	0.0835	0.1429	0.258	0.1179
	KFA	0.0536	0.0947	0.0738	0.1049	0.225	0.2175
SSR	PCA	0.1237	0.1287	0.0880	0.1931	0.179	0.1869
	KPCA	0.1403	0.1564	0.1194	0.2110	0.179	0.1765
	KFA	0.1237	0.1394	0.0972	0.2087	0.187	0.2837
ASR	PCA	0.0268	0.1257	0.0773	0.1269	0.279	0.1302
	KPCA	0.0804	0.1391	0.1030	0.1775	0.262	0.1319
	KFA	0.0319	0.1232	0.0795	0.1306	0.283	0.1798

TABLE 5. ROR rates resulted from Experiment 3 using MAHCOS. (PS= Plastic Surgery Type),(HFE = Holistic Feature Extraction) and (PIT= Photometric Illumination Technique) .

PIT	HFE \ PS	Resurf acing-	Foreh ead-Lift	Bleph aropl-asty	Otopl a-sty	Fat-Injec-tions	Rhyti decto-my
GRF	PCA	96.6%	84.9%	85.1%	78.1%	81.2%	73.2%
	KP C-A	93.1%	77.4%	80.9%	72.6%	75.0%	67%
	KF A	89.7%	84.9%	83%	79.5%	81.3%	70.3%
WAV	PCA	96.6%	77.4%	77.7%	74%	75.0%	71.9%
	KP C-A	93.1%	71.7%	75.5%	71.2%	62.5%	66.7%
	KF A	96.6%	79.3%	72.3%	76.7%	75.0%	25.2%
MSO	PCA	89.7%	69.8%	73.4%	69.9%	75.0%	60.8%
	KP C-A	82.8%	69.8%	72.3%	63.0%	68.7%	60.1%
	KF A	93.1%	67.9%	73.4%	71.2%	75.0%	35.6%
SSO	PCA	89.7%	71.7%	74.5%	61.6%	75.0%	59.5%
	KP CA	82.8%	64.2%	72.3%	60.3%	68.7%	57.2%
	KF A	89.7%	71.7%	74.5%	63.0%	75.0%	18%
DCT	PCA	72.4%	62.3%	66%	57.5%	68.7%	40.9%
	KP CA	58.6%	62.3%	67.0%	53.4%	56.2%	42.5%
	KF A	75.9%	56.6%	58.5%	57.5%	68.7%	0.7%
SF	PCA	75.9%	67.9%	72.3%	61.6%	75.0%	51.6%
	KP CA	75.9%	67.9%	72.3%	54.8%	68.7%	49.7%
	KF A	75.9%	66.0%	70.2%	64.4%	75.0%	37.9%
HOMO	PCA	82.8%	62.3%	69.2%	61.6%	62.5%	46.4%
	KP CA	72.4%	56.6%	67.0%	56.2%	56.3%	47.7%
	KF A	86.2%	62.3%	63.8%	56.2%	62.5%	11.8%
MSR	PCA	41.4%	35.9%	47.9%	38.4%	43.7%	23.2%
	KP CA	37.9%	35.9%	40.4%	35.6%	43.7%	23.5%
	KF A	41.4%	39.6%	44.7%	32.9%	43.7%	0.7%
WD	PCA	72.4%	50.9%	67.0%	54.8%	68.7%	44.4%
	KP CA	55.2%	56.6%	69.1%	46.6%	50.0%	45.4%
	KF A	79.3%	52.8%	64.9%	53.4%	62.5%	1.3%
SSR	PCA	41.4%	35.9%	47.9%	38.4%	43.7%	23.5%
	KP CA	37.9%	35.9%	40.4%	38.4%	43.7%	23.5%
	KF A	41.4%	39.6%	44.7%	34.3%	43.7%	1.3%
ASR	PCA	72.4%	64.2%	62.8%	45.2%	56.2%	37.9%
	KP CA	72.4%	67.9%	59.6%	43.8%	50.0%	38.9%
	KF A	72.4%	66.0%	58.5%	43.8%	56.2%	9.2%

normalization technique with GKFA feature extraction technique and the MAHCOS distance measure, while for identification system using GRF illumination normalization technique with GPCA feature extraction technique and the MAHCOS distance measure.

In the fourth experiment, the effect of using five histogram normalization techniques (Normal histogram**,** Rank histogram, Exponential histogram, Lognormal Histogram, Histruncate Histogram) after the application of the photometric illumination normalization techniques on facial plastic surgery recognition rates was investigated. It has been found that the four histogram normalization: Normal, Rank and Lognormal histogram showed higher recognition rates than the exponential histogram normalization technique. Histruncate histogram proves to be the best For face verification systems, while rank histogram proves to be the best For face identification systems.

Finally, it can be concluded that for face verification systems, the best results in case of resurfacing are obtained using steerable filters (SF) illumination normalization technique, exponential histogram normalization and gabor principal component analysis (GPCA) feature extraction technique. For forhead-lift, best results are obtained using Gradient faces (GRF) illumination normalization technique, rank histogram and gabor principal component analysis (GPCA) feature extraction technique. For Blepharoplasty, best results are obtained using GRF illumination normalization technique, no histogram normalization and gabor kernel fisher analysis (GKFA**).** For Otoplasty, best results are obtained using SF illumination normalization, Lognormal histogram and GPCA feature extraction technique. For Fat injection, best results are obtained using GRF illumination normalization, rank histogram and GKFA feature extraction technique. And for Rhytidectomy, best results are obtained using GRF illumination normalization, no histogram normalization and GPCA feature extraction technique.

As a general conclusion for the verification problem and for all types of plastic surgery, best illumination technique is Gradient faces (GRF) Normalization Technique, the best histogram normalization is Histruncate histogram and the best feature extraction technique is GKFA using MAHCOS distance. . MAHCOS distance gives the best results. The minimum error rates reach 0.0140, 0.0173, 0.0451, 0.0463, 0.0549 and 0.0667 in case of resurfacing, Bleharoplasty, forhead-lift, Otoplasty, Rhytidectomy and fat-injections, respectively.

As for face identification systems: best results of resurfacing surgery, are obtained using GRF illumination normalization technique, all types of histogram normalization except exponential and Histruncate histogram normalization and GPCA feature extraction technique. For forhead-lift, best results are obtained GRF illumination normalization technique, rank histogram and gabor GPCA or GKFA feature extraction techniques. For Blepharoplasty, best results are obtained using wavelet-based (WAV) illumination normalization technique, rank histogram normalization and GPCA or GKFA**.** For Otoplasty, best results are obtained using GRF illumination normalization, histruncate histogram and GPCA or GKFA feature extraction techniques. For Fat injection, best results are obtained using GRF illumination normalization, all types of histogram and GPCA or GKFA feature extraction techniques. And for Rhytidectomy, best results are obtained using WAV illumination normalization, rank histogram normalization and GPCA or GKFA feature extraction techniques. A conclusion can be drawn that for identification purpose and for all types of plastic surgery, best illumination technique is GRF normalization Technique, no histogram normalization and the best feature extraction technique is GPCA. MAHCOS distance gives the best results. This results in a correct recognition rate of 96.55%, 85.11%, 84.91%, 81.25%, 78.08% and 73.20% in case of resurfacing, Blepharoplasty, Forehead-lift, Fat-injections, Otoplasty and Rhytidectomy plastic surgery, respectively.Therefore according to the present comparative study of performance of plastic surgery face recognition algorithms, it is recommended to use GKFA as a feature extraction technique, GRF for illumination normalization technique, histruncate histogram for histogram normalization for face verification systems. As for face identification system, it is recommended to use GPCA for feature extraction technique, GRF as an illumination normalization technique, rank histogram for histogram normalization. For both verfication and identification the 1-nearest neighbour classifier using MAHCOS distance gives satisfactory results..

In conclusion, all the face identification/verification techniques adopted in this paper are promising and could be used and developed to improve a face recognition system performance and to find the nearness between the pre-plastic surgical face to the-post plastic surgical face.

References

[1] V. Blanz, S. Romdhami, and T. Vetter, “Face identification across different poses and illuminations with a 3d morphable model,” in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, pp.202–207, 2002.

[2] A. Lanitis, C. J. Taylor, and T. F. Cootes, “Toward automatic simulation of aging effects on face images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–450, Apr. 2002.

[3] Singh, R., Vatsa, M., Noore, A.: Effect of plastic surgery on face recognition: A preliminary study. In: Proceedings Workshops of Computer Vision and Pattern Recognition (CVPR),pp. 72–77, 2009.

[4] S. Li, R. Chu, S. Liao, and L. Zhang, “Illumination invariant face recognition using near-infrared images,” *IEEE Trans. Pattern Anal* , vol. 29, no. 4, pp. 627–639, Apr. 2007.

[5] R. Singh, M. Vatsa, A. Ross, and A. Noore, “A mosaicing scheme for pose-invariant face recognition,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1212–1225, Oct. 2007.

[6] N. Ramanathan, A. R. Chowdhury, and R. Chellappa, “Facial similarity across age, disguise, illumination and pose,” in *Proc. Int. Conf. Image Processing*, 2004, vol. 3, pp. 1999–2002.

[7] R. Singh, M. Vatsa, and A. Noore, “Face recognition with disguise and single gallery images,” *Image Vis. Comput.*, vol. 27, no. 3, pp. 245–257, 2009.

[8] Richa Singh, Himanshu S. Bhatt, Samarth Bharadwaj, Afzel Noore and Shahin S.Nooreyezdan, “Plastic Surgery: A New Dimension to Face Recognition”, IEEE TRANS. ON INFORMATION FORENSICS AND SECURITY, VOL. 5, NO. 3, 2010.

[9] Mariusz Leszczyński, "Image Preprocessing for Illumination Invariant Face Verification", journal of telecommunication and information technology, Vol. 2010, No.4, pp.19, 2010.

[10] Keramat Allah Ghaffary , Fardin Akhlaghian Tab , Habibollah Danyali "Profile-based Face Recognition using the Outline Curve of the Profile Silhouette", IJCA Special Issue on “Artificial Intelligence Techniques - Novel Approaches & Practical Applications” AIT, 2011.

[11] P. N. Bellhumer, J. Hespanha, and D. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 7, pp. 711–720, 1997.

[12] Ming Hsuan Yang, “Face Recognition Using Kernel Methods”, http://www.face-rec.org/ algorithms/ kernel /nips01 .pdf .

[13] A. Eleyan, H. O¨ zkaramanli, and H. Demirel, “Complex wavelet transform-Based face recognition,” *EURASIPJournal on Advances in Signal Processing*, vol. 2008, Article ID 185281, 13 pages, 2008.

[14] V. Kyrki, J.-K. Kamarainen, and H. K¨alvi¨ainen, “Simple Gabor feature space for invariant object recognition,” *Pattern Recognition Letters*, vol. 25, no. 3, pp. 311–318, 2004.

[15] M. Lades, Jan C. Vorbrueggen, J. Buhmann, et al., “Distortion invariant object recognition in the dynamic link architecture,”*IEEE Transactions on Computers*, vol. 42, no. 3, pp. 300–311,1993.

[16] C. Liu, “Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance,”*IEEE Trans. on Pattern Analysis andMachine Intelligence*, vol. 28, no. 5, pp. 725–737, 2006.

[17] L. Shen and L. Bai, “Information theory for Gabor feature selection for face recognition,” *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 30274, 11 pages, 2006.

[18]"Nearest Neighbour Classifier", http: //www.robots.ox.ac.uk /dclaus/ digits/ neighbour.htm.

[20] Vitomir Štruc, and Nikola Pavešić, "The Complete Gabor-Fisher Classifier for Robust Face Recognition", EURASIP Journal on Advances in Signal Processing, Vol.2010. No.31, 2010.

[21] Ki-Chung, “Face Recognition using principal component analysis of Gabor filter responses”, RATFG-RTS ’99 Proc. Of the international workshop on recognition, Analysis andTracking of faces and Gestures in Real-Time Systems, pp.53-57, 1999.

Cluster Head Recovery Mechanism for Hierarchical Protocols

NAHLA S. ABD EL-AZEEM[1], SHERINE M. ABD EL-KADER[2], MOHAMED M. ZAHRA[3].

[1] System Engineer at Computers and Systems Dept., Electronics Research Institute Cairo, 12622, Egypt

[2] Associate Professor at Computers and Systems Dept., Electronics Research Institute Cairo, 12622, Egypt

[3] Communications Dept., Faculty of Engineering, Al-Azhar University Cairo, Egypt

Abstract

Due to the big and essential role played by the cluster head in hierarchical protocols of wireless sensor network, it consumes an amount of energy more than the member nodes of the cluster. Therefore, the hierarchical protocols usually suffer from a packet loss due to frequent cluster head failure during the data transmission. This decreases the communication reliability in wireless sensor network applications. Errors caused by data loss inevitably affect the data analysis and subsequent decision making. This paper introduces, the Low-Loss Low-Energy Adaptive Clustering Hierarchy Centralized Protocol (LLEACH-C); LLEACH-C provides a cluster head recovery mechanism, which recovers the cluster head's death during the protocol round operation. This is a missing mechanism in LEACH-C protocol, the protocol that is one of the most popular in hierarchical protocols. The simulation results showed that, LLEACH-C succeeds in decreasing the data packet loss by on average 43% over LEACH-C, and in increasing the lifetime of the network by on average 69.3% more than VLEACH, the protocol that also proposes a mechanism for cluster head death recovery. Furthermore the behavior of the LLEACH-C, in both of the End to End delay and Throughput, are analyzed and compared to VLEACH.

Keywords: *Wireless sensor network, Clustering, Hierarchical protocols, LEACH-C, VLEAH, Packet loss.*

1. Introduction

To establish the connection between the physical world, the computing world and human society, a new class of short-range wireless communication networks has appeared to facilitate the needs of monitoring and controlling the physical environment especially in remote and unreachable areas. These networks are called Wireless Sensor Networks (WSNs). It can be used in wide-range of applications like Agriculture [1], Medicine [2], Military [3], Intelligent Buildings [4], Natural Disasters discovery [5], Traffic control [6], etc. A WSN generally consists of a large number of small and smart Sensor Nodes (SNs), SN contains one or more sensors, embedded processors, limited memory, low-power radio, and is normally operated with AA battery or coin cell, they are capable of monitoring a desired event in a particular geographical area, processing this event related data, and sending it to a remote location which is called Base Station (BS) using a certain routing protocol. Due to the small size of sensors, they cannot be equipped with large power supplies, thus small batteries are used to provide their energy. Battery replacement is impossible in most applications. Therefore designing an energy-aware routing protocol is considered to be one of the most significant challenges for these networks because most of the sensor's energy is consumed by the communication process when sending data from SN to the sink. By summarizing routing challenges in WSNs [7], many ways for routing had been developed for energy optimization in WSNs [8-9]. They can be categorized into flat [10], hierarchical based (cluster based) [11] and location based [12]. Cluster-based routing algorithm has a better energy utilization rate compared with non-cluster routing algorithm. In hierarchical (cluster-based) routing protocols, SNs are divided to clusters each cluster has a leader which called Cluster Head (CH). Using this scheme reduces the energy consumption, during sending data to the BS, by decreasing the number of messages sent to the BS through data aggregation and fusion. Balanced distribution of energy consumption tasks among cluster nodes cannot be achieved, because the CH has many energy consuming tasks. CH tasks are (i) overhearing Member Nodes (MNs) for the round time, (ii) creating Time Division Multiple Access (TDMA) scheme and broadcasting it to all MNs to use it in a communication process, (iii) Setting up communication channels between itself and all MNs, (iv) receiving data from all MNs, and (v) aggregating the received data and sending them to the BS that might be remotely located. On the other hand, MNs have only two tasks, data sensing and sending data to CH then sleeping till next time slot. This unbalanced energy distribution tasks cause CH to consume its energy faster than other MNs, and may fail during the round whereas, other MNs are alive and sensing data. If the CH fails during the round, the collected data from MNs will be lost because it will never reach the BS. Also MNs consume their energies for sensing the surrounded environment while there will be no actual data transmission, so the entire cluster will be useless. Which means that the CH represents a single point of failure for the cluster.

This paper proposes a new cluster head recovery mechanism over LEACH-C protocol, this mechanism recovers cluster head before failure to save the collected data from being lost.

The rest of this paper is organized as follows: Section 2 overviews the related works. Section 3 discusses the LEACH-C protocol as a base for LLEACH-C. Section 4 discusses the LLEACH-C in details. Section 5 represents LLEACH-C system model, Section 6 views the performance evaluation for LLEACH-C, and finally section 7 represents the conclusions and future work.

2. Related Work

Many algorithms have been developed using clustering techniques in routing to increase the energy efficiency of the network [13]. Low Energy Adaptive Clustering Hierarchy (LEACH) protocol[14], is the first hierarchical routing protocol proposed for WSNs since 2000. It has become one of the most well known energy efficient clustering algorithms for WSNs. The whole operation of LEACH is divided into rounds. Each round consists of setup phase followed by steady state phase. During set up phase clusters are formed by organizing SNs into clusters, each cluster has its own CH which has been selected randomly based on the threshold generated by the algorithm. Each CH sets up its own TDMA scheme which determines a certain Time Slot (TS) for each MN for data transmission in the next phase. In the steady state phase CHs aggregate the data received from their MNs and send the aggregated data to the BS by single hop communication. One of the major benefits of LEACH is reducing energy consumption than flat protocols. This is achieved by decreasing the number of transmissions to the BS through sending the aggregated data from CHs, instead of each node. Further, the MNs entered sleeping mode during the transmission phase until its predetermined TS is active. In spite of these benefits, there are many drawbacks in LEACH protocol such as (i) CHs are not uniformly distributed; as CHs can be located at the edges of the network, (ii) random CHs selection, which does not take energy consumption into consideration, (iii) small coverage area because it assumes that all nodes over the network that can communicate with each other, and are also able to reach the BS, (iv) CH will die faster than other cluster members because of unbalanced distribution of energy consumption tasks among cluster nodes.

Since LEACH has many drawbacks, many researchers tried to enhance the performance of the protocol. Some of these enhancements are briefly described in the following paragraph.

LEACH-C protocol [15-16] solves the CHs uniformly distribution issue; by using a BS central control algorithm to form the clusters, It differs from LEACH only in the cluster setup phase. LEACH-C uses an unlimited energy BS, when the network is initialized, all SNs send their energy level and positions to the BS, which in turn uses a centralized algorithm to select all CHs depending on the larger residual energy and non bordered positions, then sends a message to all nodes informing each of them with the CH responsible for each one, then the rest of the operation will be the same as in LEACH.

The main advantage of LEACH-C over the LEACH is the knowledge that the BS has about the entire network topology, which makes the selection of CH more efficient, more accurate, better distributed and less energy consuming. Consequently, LEACH-C enhances the entire network lifetime. However, the selected CH still represents single point of failure for the cluster.

To save the CH from being a single point of failure for the cluster, researches had presented an idea for back up the CH, by placing a Vice Cluster Head (VCH) to take the role of the CH in case of failure.

VCH idea had been applied for the first time for hierarchical protocols in [17] over the LEACH protocol. In the Vice Low Energy Adaptive Clustering Hierarchy (VLEACH) protocol, randomly selects any of cluster MNs to become a VCH after the CH failure, and takes the role of the CH. Although in VLEACH the CH is not a single point of failure, it still has the same drawbacks of LEACH as described above.

A new version of VLEACH called Improved V-LEACH protocol [18] was proposed to reduce the drawbacks of VLEACH by enhancing the CH selection process based on minimum distance, maximum residual energy, and minimum energy, also besides having a CH in the cluster, there is a VCH selected randomly to take the role of the CH when the CH dies. The main advantage of this protocol is to increase the steady state phase to reduce the reclustering process, thus increasing the network life time. The above mentioned protocols used the VCH for increasing the network lifetime, also many protocols had used the VCH in different ways to achieve different goals.

As in [19] the Energy-aware Data Gathering and Routing Protocol Based on Double Cluster-heads (EAGRDC) which used the VCH for sharing the energy consumption of CH. not for CH recovery. Also in [20] Multiple Cluster-Heads Routing Protocol (MCHRP) used two VCHs in two ways, one for sharing the energy consumption with the CH and the other for CH back up to achieve the load balancing in the cluster.

It should be obvious that, none of the previous ideas has paid attention to the packet losses caused when CH fails and VCH takes the designed role. This paper highlights the importance of VCH in loss reduction over the network.

3. Low-Energy Adaptive Clustering Hierarchy Centralized Protocol (LEACH-C)

LEACH-C enhanced both the poor CH selection in LEACH, and the LEACH network lifetime by using the BS central algorithm, which can determine better clusters than the distributed algorithm.

This section will describe LEACH-C protocol as a base for the proposed LLEACH-C protocol. LEACH-C operation is divided into rounds each round consists of a setup phase and a steady state phase.

3.1 Setup phase

Each SN sends information about its residual energy and location to the BS. Then BS uses its central algorithm to compute the average node energy, and any node having energy below this average cannot be a CH for the current round. The BS finds clusters using the simulated annealing algorithm. This algorithm attempts to minimize the amount of energy for the non-cluster head nodes to transmit their data to the CH, by minimizing the total sum of squared distances between all the non-cluster head nodes and the closest CH. Once the CHs and associated clusters are found, the BS broadcasts a message that contains the CH ID for each node. If a node’s CH ID matches its own ID, the node is a CH; otherwise, the node determines its’ TS for data transmission and goes to sleep until it is time to transmit data.

3.2 steady state phase

The steady-state phase of LEACH-C is identical to that of LEACH. The steady-state operation is broken into frames, where MNs send their data to the CH at most once per frame during their allocated TS using a minimal amount of energy. This allows the radio components of each MN to be turned off at all times except during its’ TS. The CH must be awake to receive all the data from the MNs in the cluster. Once the CH receives all the data, it performs data aggregation. Thus, compressed data is sent from the CH to the BS at the end of its frame. This process is repeated till the designed number of frames per round is finished.

LEACH-C performance is better than LEACH because, it delivers about 40% more data per unit energy than LEACH, which enhances the total network life time.

Although LEACH and LEACH-C provide significant energy saving, they still suffer from loss of data packets due to frequent CH failure during the data transmission, because it represents a single point of failure for the cluster. This issue creates the main motivation of this research, how to save the collected data packets before the CH failure.

4. Low-Loss Low-Energy Adaptive Clustering Hierarchy Centralized Protocol (LLEACH-C)

This section describes in details the functionality of the proposed LLEACH-C protocol which is a modified version from LEACH-C. Section 4.1 represents the description of LLEACH-C. Section 4.2 explains the LLEACH-C operation.

4.1 Description of LLEACH-C

LLEACH-C is the first protocol that applies the idea of VCH Mechanism (VCHM) over LEACH-C protocol to achieve fault tolerance for the CH. It reduces losses by preventing the packets from being lost when the CH energy decreases.

As in hierarchical protocols, the LLEACH-C protocol divides the network into clusters, as shown in Fig.1(a), each cluster has a CH and VCH selected by the BS at the beginning of each round, CH responsibilities are overhearing, receiving, aggregating, and sending data to BS. Each frame consumes an amount of energy equal to EF reduced from the CH energy to complete the transmission to BS. For each frame received at BS, it tests the residual energy for each CH, if any CH has residual energy (ECH) lower than the EF required for the next frame like cluster X, the BS employs the VCHM, as shown in Fig.1 (b).

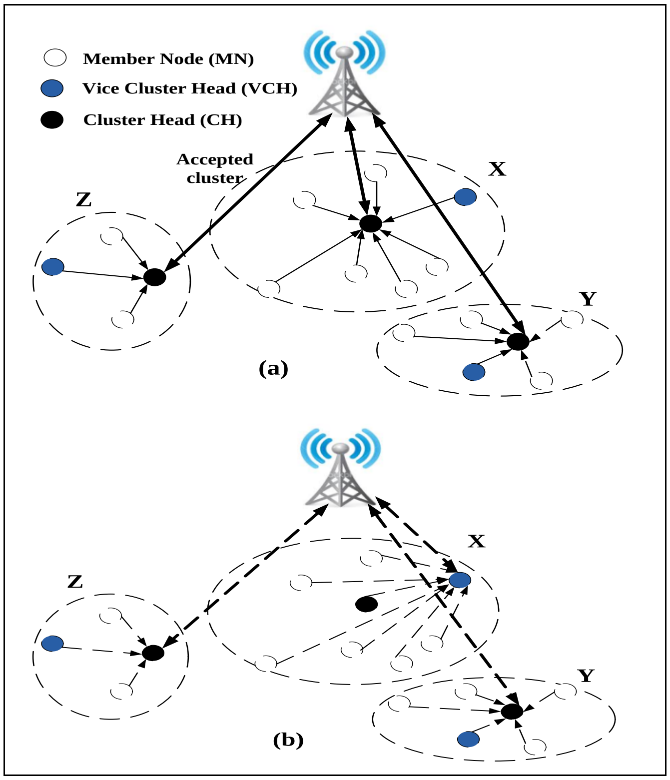


Fig.1 The network structure of LLEACH-C: (a) before applying VCHM, and (b) after applying VCHM on cluster X

4.1.1 Vice Cluster-Head Mechanism

Because of CH numerous responsibilities, its energy is consumed faster than MNs, therefore the probability of CH failure during the round is increased. Causing all cluster failure, accordingly loss of all collected data of that cluster. So to solve this problem, the VCHM had been proposed.

At the begging of each round, and during the setup phase the BS selects CHs and forming the cluster, after that it sorts MNs in each cluster according to the residual energy, and then selects the largest residual energy MN as a VCH. After that the BS sends a message to all nodes to announce their CH and VCH.

During the data transmission in the steady state phase, when the BS receives a frame, it tests the ECH for each CH, if any CH has ECH lower than the required energy for the next frame EF, the BS starts VCHM over that cluster by sending a message to all cluster nodes, informing them that the roles between CH and VCH are exchanged, therefore MNs establish a connection with VCH, CH becomes a normal MN and establishes a connection with VCH, and VCH will start overhearing, receiving, aggregating, and sending data to the BS till the end of the current round.

4.2 LLEACH-C Protocol operation

LLEACH-C achieves a good performance in fault tolerance by avoiding the CH from being a single point of failure for the cluster, when its energy decreases to a level lower than the energy required for the next frame.

LLEACH-C operation is divided into sequenced rounds; each round as shown in Fig.2 consists of two main phases, setup phase and steady state phase.

4.2.1 Setup phase

This phase is subdivided into three phases, base station cluster formation phase, cluster send/receive phase and TDMA scheme phase as shown in Fig. 3. In the cluster formation phase each node sends its ID, current location, and current level of residual energy to the BS, then the BS uses its central algorithm to select the CHs according to their remaining energy levels, the selected CHs are 5% of alive nodes. This approach has been concluded by authors of [10] to obtain the optimal number of CHs for energy minimization. Before sending the cluster information to the nodes, the BS needs to select VCHs for each cluster, by sorting each cluster MNs descendingly according to their remaining energy, and selecting the largest energy MN as a VCH. While in cluster send/receive the BS sends

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

message to all nodes, and announces their corresponding CH and VCH.

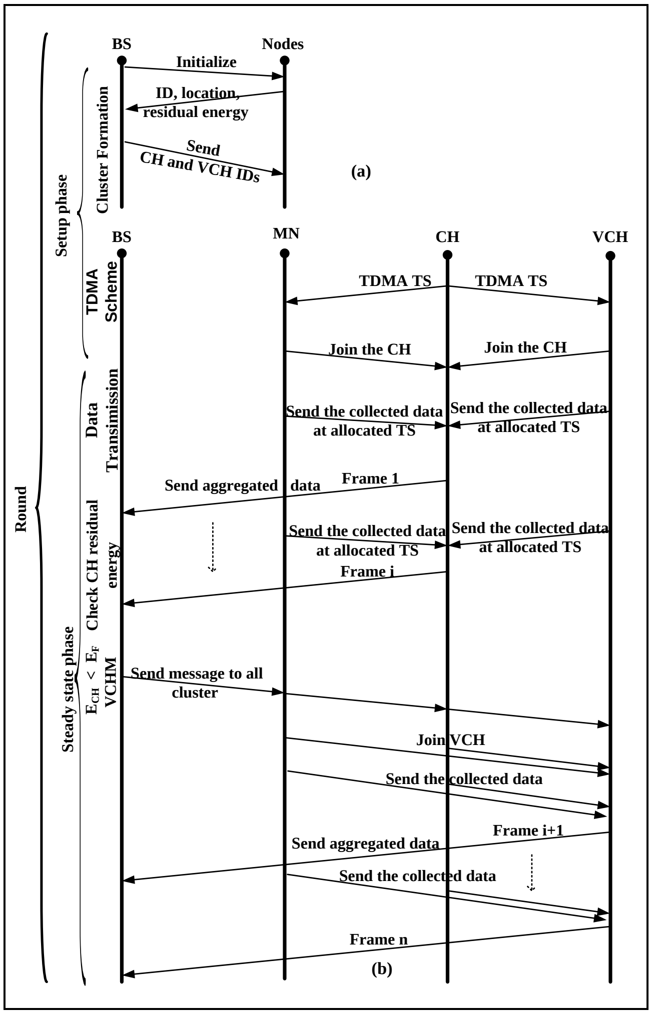


Fig.2 LLEACH-C protocol Round: (a) Round initialization (b) Round operation

These messages contains CH_{ID} and VCH_{ID}, when a node receives this message it compares its own ID with the received IDs, if it matches with the CH_{ID} the node will become a CH, else if it matches with the VCH_{ID} the node will become a VCH, otherwise, the node will be a MN and will establish the connection with the corresponding CH.

Then TDMA schema phase starts, by the CH creating the TDMA schema. Fig.4 (a) shows the TDMA schema Time Slots (TS), TSs are equals to no. of MNs + VCH + 2, CH arranges TDMA scheme as follows, (i) TSs for MNi, (ii) TS for VCH, (iii) the last two TSs are assigned to CH Data Aggregation (CH_{DA}) and CH Data Sending (CH_{DS}).

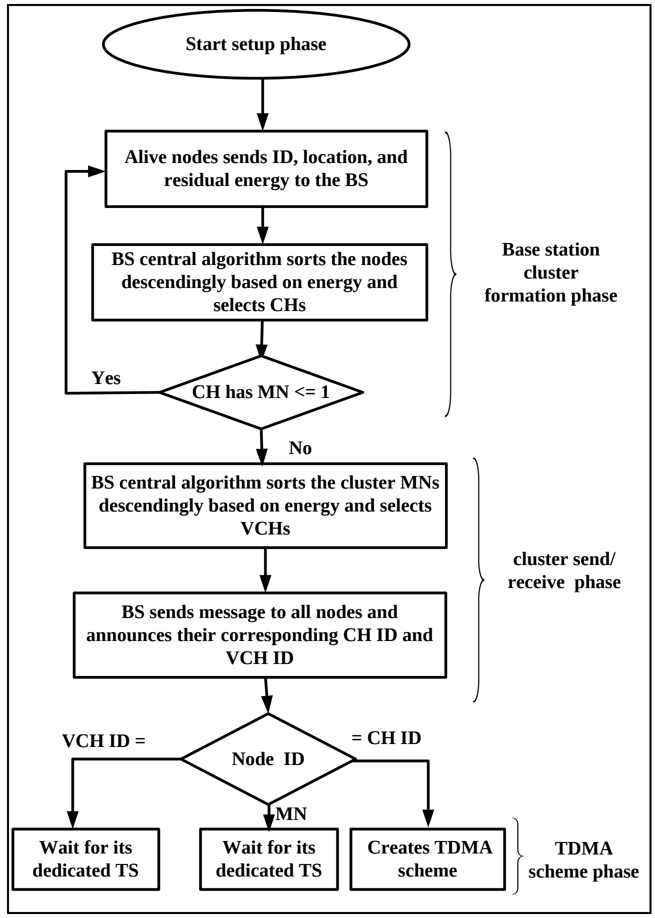


Fig.3 LLEACH-C set up phase flow chart

During the steady state phase when the BS decides to employ the VCHM over a certain cluster, instead of creating a new TDMA the VCH uses the same TDMA already created, but only exchanges the last three TSs between himself and CH, as shown in Fig. 4 (b), that is as a result of exchanging the roles between them.

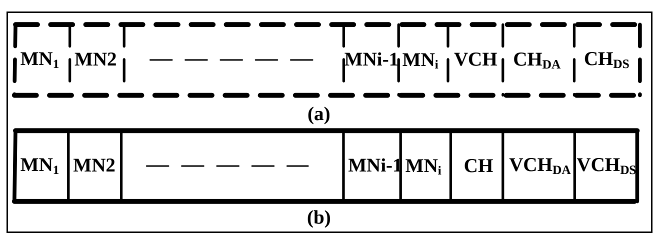


Fig. 4 LLEACH-C TDMA Scheme (a) before applying VCHM (b) after applying VCHM.

Then the CH announces its TDMA scheme by a message, sends it to its MNs containing the allocated TS for each one. When MNs receive this message each node knows its allocated TS for data transmission and goes to sleep until it is time to transmit data. Once the clusters are formed and TDMA schemes are constructed and distributed, the steady state phase is started.

4.2.2 Steady state phase

All MNs through the network sense the surrounding environment to collect required data and send it to the corresponding CH, once the CH receives packets from all MNs; it aggregate the collected data in one packet and sends it to the BS. As shown in Fig. 5.

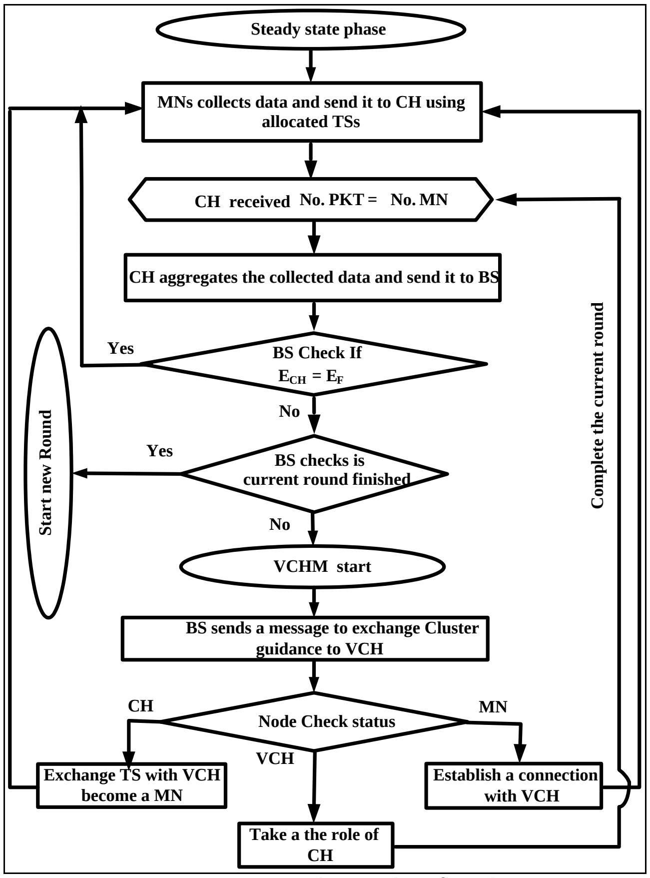


Fig. 5 LLAECH-C steady state phase flow chart

During data transmission, the VCHM at the BS central algorithm checks the residual energy for each CH, if the ECH for a certain CH is larger than EF, the VCHM confirms that this CH is able to send in the next frame, otherwise, the BS starts to employ the VCHM to save the next frame from being lost. The VCHM starts by sending a message from BS to the cluster that suffers from the above condition, to exchange the roles between the CH and VCH, till the end of the active round, the CH becomes a MN under the leadership of VCH.

Then the CH and VCH exchange the allocated TSs for each of them, as described above.

5. LLEACH-C System model

This section describes in details the assumptions of LLEACH-C system model. Section 5.1 explains the proposed network model in designing the protocol. Section 5.2 describes the adopted radio model in the protocol operation. Section 5.3 views the VCHM.

5.1 Network Structure for LLEACH-C

N sensor nodes are assumed to be randomly distributed in 1000*1000 square areas and periodically collect data. LLEACH-C network has the following design assumptions: (i) sensor nodes and BS are immobile after deployment and the BS is located far from all sensors, (ii) all nodes in the network are identical and have limited energy with an identifying ID, (iii) all nodes are able to reach BS and can communicate with each other, (iv) CHs perform overhearing, receiving, aggregating, and sending of data to the BS, (v) propagation channels are symmetric; that is the energy consumption of sending data from node A to node B equals to the consumption from node B to node A, (vi) all nodes are sensing the surrounding environment in a fixed rate and send data periodically to CH, and finally (vii) all nodes start with the same energy level and the BS has unlimited energy resource.

5.2 Radio Signal Propagation and energy consumption Model

LLEACH-C use a simple energy model as in LEACH-C, Where energy dissipation 50nJ/bit for both transmitter ETx-elec and receiver ERx-elec to run the radio electronics and the energy dissipated in the transmitter amplifier in a free-space channel εfs = 10pJ/bit/m2. Thus to send an (I-bit) message over a distance d the energy consumed by the radio can be calculated as:

$$\mathbf{E_{TX}\ (I,\ d) = E_{TX\text{-}elec}\ (I) + E_{TX\text{-}amp}\ (I,\ d)\ E_{TX}\ (I,d) = I{*}Eelec + I * \varepsilon_{Fs} {*}d2} \qquad (1)$$

To receive this message at destination the energy consumed by the radio can be calculated as:

$$\mathbf{E_{Rx}\ (I) = ER_{TX\text{-}elec}\ (I) = Eelec * I} \qquad (2)$$

5.3 Vice Cluster Head Mechanism VCHM

VCHM computes the $\mathbf{E_F}$ required from each CH, during the data transmission in steady state phase, by the following equation:

$$\mathbf{E_F = E_{DR} + E_{DA} + E_{DS}} \qquad (3)$$

Equation (3) explains the components of $\mathbf{E_F}$. Where $\mathbf{E_{DR}}$ is the CH consumed energy for data receiving, $\mathbf{E_{DA}}$ is the CH consumed energy for data aggregation, and $\mathbf{E_{DS}}$ is the CH consumed energy for data sending to BS.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

For any CH has MNs and receives packets of I data bits (PKT (I)), equation (3) can be expanded as follows:

$$E_F = E_{elec} * I * MN + P_{KT}(I) * E_{DA} * MN + I * E_{elec} + I * \varepsilon_{Fs} * d^2 \quad (4)$$

Equation (4) is the criteria for the VCHM activation. Because when the BS received a frame, the VCHM compares the E_{CH} for each CH, with the calculated E_F for that cluster, if $\mathbf{E_{CH} < E_F}$, the VCHM will exchange the roles between the CH and VCH.

6. Performance evaluation for LLEACH-C

The LLEACH-C simulation model is built by using OMNET++ discrete event Simulator [21], to evaluate its performance from the point of view of the packet loss and network lifetime.

6.1 Simulation parameters

The values of the used parameters in the simulation are listed in Table.1.

Table 1: simulation parameters

Parameters	***Value***
Field size (M ·M)	1000 · 1000
Initial energy of sensor node	0.5 J
Transmitter /receiver Electronics ETX and ERx (Eelec)	50 nJ/bit
Transmitter amplifier ε_{Fs}	0:0013pJ / bit /m2
The energy for aggregation EDA	5 nJ /bit/ signal
The data packet size	4000 bits

6.2 Results

The simulation results which were carried out to compare the performance of the LLEACH-C with both, LEACH-C as a base protocol, and VLEACH as a basic idea, are in sections 2, 3 and 4. In simulation, the number of the deployed nodes is increased from 100 to 1000 node with unit step 100. Most of the results obtained in this paper were obtained by averaging ten independent simulation runs, where each run uses a different randomly-generated topology of sensor nodes.

6.2.1 Packet loss

The packet loss is the failure of one or more transmitted packets to arrive to their destination. This event can cause noticeable effects in all types of digital communications. This paper cares to give attention to the packet loss due to the CH failure during the round.

Therefore this paper is concerned with reducing the losses caused by applying the VCHM over LEACH-C. Simulation results compare the LEACH-C average packet loss with the LLEACH-C average packet loss, to find out the loss reduction as a result of adding the VCHM.

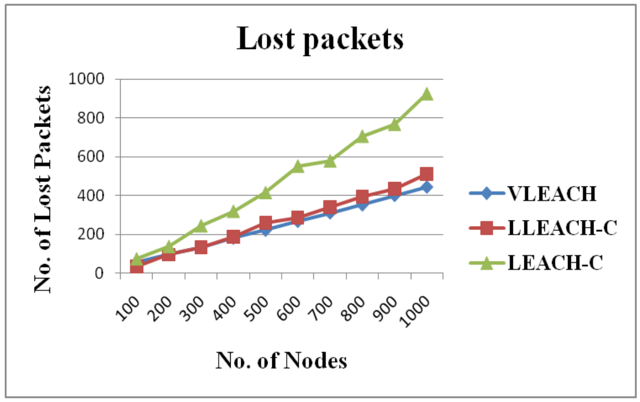


Fig.6 Average packet loss vs. number of nodes

Fig.6 demonstrates the average packet loss for LLEACH-C, VLEACH and LEACH-C. It is obvious that the average packet loss of LLEACH-C and VLEACH are approximately the same, because in both protocols there is a vice cluster head to take over in case of the CH failure, when this happens, it saves the collected data packets from being lost during the round, and thus solves the CH from being a single point of failure in the cluster. While in the LEACH-C, the packet loss increases dramatically as the number of nodes increases in the deployed area, which is due to the CH single point of failure problem, where; when the cluster head fails during the round, there is no recovery for its role, and all the data packets in the cluster are lost. It's clear that LLEACH-C has reduced the packet loss compared to that of LEACH-C by 43% on average, by applying the VCHM.

6.2.2 Network life time

The simulation calculates the network life time by counting the number of successive finished rounds till the last alive 10 nodes.

As shown in Fig. 7 the VLEACH lifetime is lower than LLEACH-C, by on average 69.3%, that is because the LLEACH-C uses a central algorithm, which creates optimal clusters, requires less energy for operation than the clusters formed using the distributed algorithm in VLEACH.

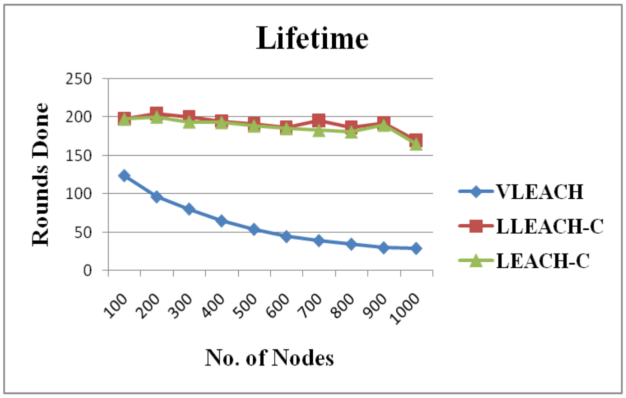


Fig.7 Network lifetime vs. number of nodes

Although the LLEACH-C reduces the packet loss of LEACH-C by approximately 43%, by applying the VCHM, the lifetime of both is the same, which means that the VCHM, saves the packets from loss, and does not consume extra network energy.

6.2.3 Throughput

The number of successive data bits received at the BS per second.

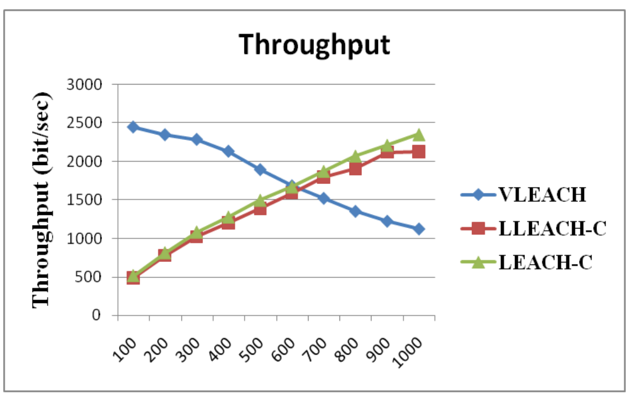


Fig.8 Throughput vs. number of nodes

Fig.8 demonstrates the average throughput for LEACH-C, LLEACH-C and VLEACH. From Fig.9, it is obvious that the throughput of LEACH-C and LLEACH-C are approximately the same, in both the throughput increased with the increase in number of deployed nodes, due to the increments of generated data packets. On the other hand, the throughput of VLEACH decreases with the number of deployed nodes. That is because VLEACH shows the worst performance of the three in lifetime as has been demonstrated before. Accordingly the number of data packets will be affected by this collapse in lifetime values. Therefore in large networks the LEACH-C and LLAECH-C perform better than VLEACH in throughput values.

6.2.4 End to end Delay

The time taken for a packet to be transmitted across a network from source to destination.

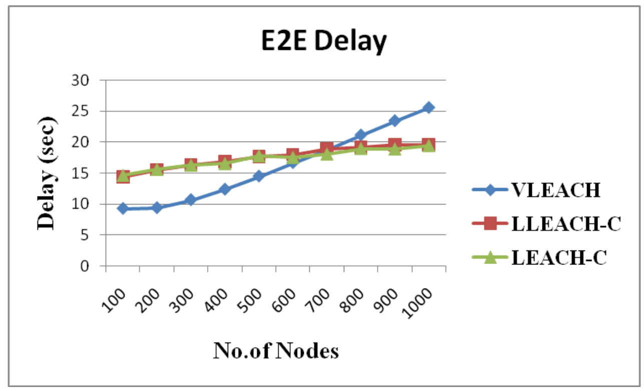


Fig.9 End to End delay vs. the number of nodes.

Fig.9 demonstrates the end to end delay for LEACH-C, LLEACH-C and VLEACH; it's obvious that the end-to-end delay of both LEACH-C and LLEACH-C is increased slowly with the increments of deployed nodes, due to a few increments of clusters members, which increases the required time from CH for overhearing and data aggregation. While in VLEACH the end-to-end delay increases rabidly with the number of deployed nodes, due to the increment of nodes density per cluster, that is because of the distributed algorithm applied for node choice in clusters, which may result in non uniform sized clusters. This problem increases the required time from CH for overhearing, and thus its failure will be accelerated and a new VCH will be chosen, and the overall network delay will increase. Therefore with large networks, LEACH-C and LLAECH-C protocols perform better end-to-end delay than VLEACH.

7. Conclusions and future work

High loss applications in WSN, suffer from degradation in network performance, caused by packet loss during the network operation, which makes the whole system unreliable, accordingly a lot of applications can bear occasional packet loss only with a certain upper bound of allowable packet loss, this paper addresses this important issue.

Generally, WSNs involve a large number of sensors ranging in hundreds or even thousands. Clustering is an effective mean for managing such high population of nodes. LEACH is one of the most popular clustering algorithms for WSNs, but its centralized version LEACH-C is able to deliver more effective data than it, even though cluster formation is more expensive, because the centralized algorithm can use network topology information to form good clusters that require less energy for operation than the clusters formed by distributed algorithm in LEACH.

Therefore this paper has proposed the VCHM mechanism for cluster head recovery to prevent the cluster head from being a single point of failure, this is done by selecting the biggest energy MN in the cluster as a VCH, to take the role of CH when its energy is not enough to finish the round, that reduces the packet loss occurred during the network operation.

Thus this paper cares about studying the performance of LEACH-C, before and after applying the VCHM to become LLEACH-C, also comparing the performance of LLEACH-C with the VLEACH, which applies the CH recovery over LEACH, in packet loss, network lifetime, End-to-End delay and throughput.

Simulation results showed that LLEACH-C offers improvements over LEACH-C in loss reduction by approximately 43%, while saving the same lifetime. Also it performs better lifetime compared to the VLEACH by on average 69.3%, with the same loss reduction, for large networks both of LAECH-C and LLEACH-C perform lower end-to-end delay than VLEACH, and higher throughput.

Renewable energy sources are the next hope to increase the WSNs lifetime, energy harvesting is the process by which energy is derived from natural resources, captured, and stored for small, wireless autonomous devices, like sensor nodes in WSN. Adding solar cells to sensor nodes gives them a continuous energy source, this leads to strengthening the lifetime of the nodes as well as the network. Studying the system performance after adding the solar cells is the goal for the future work.

References

[1] S.M.Abd El-kader and B.M.El-Basioni, “Precision Farming Solution in Egypt Using the Wireless Sensor Network Technology”, Egyptian Informatics Journal, Volume 14, Issue 3, 2013, pp 1-13.

[2] A.J. Punsawad, and Y. Wongsawat, “Wireless sensor network-based smart room system for healthcare monitoring” Robotics and Biomimetics (ROBIO), IEEE International Conference on 2011, pp. 2073 – 2076.

[3] M.P.Durisic, Z.Tafa, G. Dimic, and V. Milutinovic, “A survey of military applications of wireless sensor networks” Mediterranean Conference Embedded Computing (MECO), 2012, pp. 196 – 199.

[4] B.M.El-Basioni, S.M.Abd El-kader, and M.A.Fakhreldin, “Smart Home Design using Wireless Sensor Network and Biometric Technologies”, International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 3, March 2013, pp. 413–429.

[5] C.Yawut and S.Kilaso. “A Wireless Sensor Network for Weather and Disaster Alarm Systems”, International Conference on Information and Electronics Engineering, IPCSIT, vol.6, IACSIT Press, Singapore, 2011, pp. 155-159.

[6] M.M.Nabeel, M.Fakher el Deen, and S.Abd El-Kader, "Intelligent Vehicle Recognition based on Wireless Sensor Network", International Journal of Computer Science Issues (IJCSI), Volume 10, Issue 4, July 2013, pp. 164-174.

[7] A.Boukerche, “Algorithms and protocols for wireless sensor networks”, Wiley-IEEE Press, 2008, DOI: 10.1002/9780470396360.

[8] M.Eiaaasser, and M.Ashour, “Energy Aware Classification for Wireless Sensor Networks Routing”, International Conference on Advanced Computing Technologies, 2013, pp. 66-71.

[9] D.Goyal and M.Tripathy, “Second International Conference on Advanced Computing and Communication Technologies”, 2012, pp.474-480.

[10] P.Shenoy and L.Patnaik, “A Flat Routing Protocol for Sensor Networks”, International Conference on Methods and Models in Computer Science, 2009, pp 1-5.

[11] Z. Manap, B. M. Ali , C. K. Ng , N. K. Noordin and A. Sali, “A Review on Hierarchical Routing Protocols forWireless Sensor Networks ” Springer Science+Business Media New York Volume 72, Issue 2, September 2013, pp 1077-1104.

[12] P.Kumar, A.Chaturvedi, and M. Kulkarni “Geographical location based hierarchical routing strategy for wireless sensor networks”, International Conference on Devices, Circuits and Systems (ICDCS), IEEE, 2012, pp.9-14.

[13] B.M.El-Basioni, S.M. Abd El-kader, H.S. Eissa, and M.M. Zahra. “An Optimized Energy-aware Routing Protocol for Wireless Sensor Network”, Egyptian Informatics Journal, volume number 12, issue number 2, July 2011, pages: 61–72, DOI information: 10.1016/j.eij.2011.03.001.

[14] W.Heinzelman. “Application-specific protocol architectures for wireless networks”. PhD thesis. Massachusetts Institute of Technology; June 2000.

[15] WB.Heinzelman, AP.Chandrakasan, and H.Balakrishnan, “An application-specific protocol architecture for wireless

microsensor networks", IEEE Trans Wire Commun;1(4):660–70; 2002.

[16] A.A. Hady, S.M. Abd El-kader, and H.S. Eissa. "Intelligent Sleeping Mechanism for wireless sensor networks", Egyptian Informatics Journal , Volume 14, Issue 2, July 2013, pp 109–115.

[17] M.B.Yassein, A. Al-zou'bi, Y. Khamayseh, and W. Mardini "Improvement on LEACH Protocol of Wireless Sensor Network (VLEACH)", International Journal of Digital Content Technology and its Applications Vol. 3, No. 2, June 2009, pp 132-136.

[18] A.Ahlawat and V.Malik. "An Extended Vice-Cluster Selection Approach to Improve V Leach Protocol in WSN", Advanced Computing and Communication Technologies (ACCT), Third International Conference, 2013, pp. 236-240, DOI 10.1109/ACCT.2013.60.

[19] F.Li, X.Wang, and T.Xu. "Energy-aware Data Gathering and Routing Protocol Based on Double Cluster-heads", Communications in Information Science and Management Engineering (CISME), IEEE, Vol.1 No.4 , 2011 , PP.24-29.

[20] D.Tang, X.Liu , Y.Jiao, and Q.Yue. "A Load Balanced Multiple Cluster-heads Routing Protocol for Wireless Sensor Networks", 13th International Conference on Communication Technology (ICCT), IEEE, 2011. PP. 656 – 660.

[21] Varga A. The OMNeT++ discrete event simulation system. European simulation multiconference. Prague, Czech Republic, June 2001.

NAHLA S. ABD EL-AZEEM gained Bsc. in communication & electronics engineering from Zagazig University in 2002. In 2003 year, she joined the Professional Training Project (Network Infrastructure, security and administration Track) which is organized by the Egyptian Ministry of Electricity and Information Technology, and finished the MCSE certificate. Since 2004 she is a System Engineer in Computers and Systems Department at the Electronics Research Institute (ERI) in Egypt.

SHERINE M. ABD EL-KADER has her MSc, & PhD degrees from the Electronics & Communications Dept. & Computers Dept., Faculty of Engineering, Cairo University, at 1998, & 2003. Dr. Abd El-kader is an Associate Prof., Computers & Systems Dept., at the Electronics Research Institute (ERI). She is currently supervising 3 PhD students, and 10 MSc students. Dr. Abd El-kader has published more than 25 papers, 4 book chapters in computer networking area. She is working in many computer networking hot topics such as; Wi-MAX, Wi-Fi, IP Mobility, QoS, Wireless sensors Networks, Ad-Hoc Networking, realtime traffics, Bluetooth, and IPv6. She is an Associate Prof., at Faculty of Engineering, Akhbar El Yom Academy from 2007 till 2009. Also she is a technical reviewer for many international Journals. She is heading the Internet and Networking unit at ERI from 2003 till now. She is also heading the Information and Decision making support Center at ERI from 2009 till now. She is supervising many automation and web projects for ERI. She is supervising many Graduate Projects from 2006 till now. She is also a technical member at both the ERI projects committee and at the telecommunication networks committee, Egyptian Organization for Standardization & Quality since February 2007 till 2011, She is currently the head of Technology Innovation Support Center (TISC) at ERI.

MOHAMED M. ZAHRA received the B. S. and M. S. degrees from Al-Azhar university, Cairo, Egypt, in 1979 and 1987, respectively, all in electrical engineering. He received the Ph. D. degree from AGH, Krakow, Poland in 1993. His doctoral research focused on bandwidth allocation techniques in ATM networks. He was an Assistance Professor of communication networks from 1993 to 2006 and Associate Professor from 2006 to present in Al-Azhar University. He is an IEEE member.

Human Eye Tracking Using Particle Filters

Elham Ghasemi-Dehkordi[1], Mahmoud Mahlouji[2] and Hossein Ebrahimpour- Komleh[3]
[1]Department of Computer Engineering, Dezfoul Branch, Islamic Azad University, Dezfoul, Iran

[2]Department of Electrical and Computer Engineering, Kashan Branch, Islamic Azad University, Kashan, Iran

[3]Department of Computer Engineering, the University of Kashan, Kashan, Iran

Abstract
In this paper, an effective method for human eye tracking and also decreasing the current challenges and problems in its algorithms has been proposed. In this method, firstly face has been detected and segmented from the remaining parts to make the searching area, in tracking stage, narrower and processing speed higher. Then eye area is determined and eye pupils are detected in the specified area. In the proposed method, to support tracking in eyes closure state, corner detection has been additionally used. Experimental results were considered on a variety of images with existence head and face rotation, make up and eyes closure show the ability of this method in human eye tracking.
Keywords: *Eye Detection, Face Detection, Eye Tracking, Particle Filter, Corner Detection, Harris Algorithm.*

1. Introduction
Nowadays by ever-increasing development of scientific discussions in the content of computer vision and pattern recognition, automatic eye tracking in video images has been significantly considered because of eye organism and structure. Since eye is an important and critical organ of the body and manifests some internal moods and their reflexes to physical and non-physical factors, by using it some moods like fatigue, sleepiness, drunkenness, happiness, and sadness, stare, etc can be distinguished. So, design and implementation of eye tracking systems are applicable to the following cases: safety systems to control peoples entrance and exit, tracking, detecting and determining social misdemeanant identity, traffic systems to recognize drunken and narcotic-taken drivers, lie detector systems and in cars to detect driver fatigue and sleepiness. An eye tracking system includes the following stages: image acquisition; face detection; eye detection; and eye tracking.
Eye detection methods are divided into four classes [1]: shape-based methods; feature-based methods; appearance-based methods and hybrid methods. When eye is open, it has some special form. Eyes' corners, pupil and eyelids cause it to be distinguished from other parts of face. Shape-based methods use these characteristics and express some models for eye. This model can be a simple ellipsoid or a complex structure [2, 3]. Feature-based methods use some distinguishing features around the eye. Iris, pupil, eyelids, eyelashes, limbus (border between iris and sclera), pupil expression and reflection of light in cornea are some features used for detecting the eye [4, 5]. Appearance-based methods which are known as pattern matching methods or general methods detect and track the eyes directly based on their appearance states. These methods are independent of the considered objects and can model every other object beside the eye [6,7]. The main purpose of hybrid methods is to compose various advantages of eye models at a system to overcome relative constraints. To improve these algorithms, the proposed algorithms try to reduce search area for pattern matching by combining methods, or propose an approach reducing the time of this pattern matching [8,9].
The available approaches for eye tracking can be divided into three classes [10]: knowledge-based methods, learning-based methods and motion estimation methods. In knowledge-based methods, tracking methods are defined and developed based on the rules obtained by studies and researches on face components. Simply we can utilize some rules to describe face features and their relations [11,12]. Learning-based methods can be classified into the three following categories [13,14]: neural networks, Adaboost classifiers and support vector machines. Motion estimation methods estimate the object position in current frame, and then determine its exact position by a local searching. In the next stages, by a correct estimation in the first stage, the accuracy and effectiveness of the method are increased [15-17].
In this paper, an effective method is presented for eye tracking. To perform this, in eye and face detection stages, color feature has been used. Then in tracking stage, pupil position estimation method and particle filter have been used. To support eye closure in tracking stage, shape-based method is combined with pupil position estimation method and parallel to particle filter, corner detection has been used. The remainder of this paper is as follows: in the second section, and eye tracking system will be introduced. In the third section, the proposed method is presented. In the fourth section, the results obtained by experiments are offered and the fifth section is for conclusion.

2. Eye Tracking System
In general, an eye tracking system can be outlined in four stages:
- Image acquisition
- Face detection
- Eye detection
- Eye tracking

Image acquisition stage is data or image entry to system which can be a video capturing by a camera. The output of this stage is an image in RGB space. After image acquisition, face detection is performed. One current method for face detection is by using color feature [18]. As the RGB space, besides having pixel color data, has light intensity data, and also face brightness is different in various people and environments, use this space to detect face color is not effective and makes some problems. So we should find a space not dependent on light intensity. One option is YCbCr chromatic color space. Firstly the image is transformed to YCbCr. RGB color space is transformed to YCbCr as:

$$\begin{cases} Y = 0.299R - 0.587G - 0.11B \\ Cb = R - Y \\ Cr = B - Y \end{cases} \quad (1)$$

After transforming to YCbCr, the mean and covariance matrix of Cb and Cr components are computed. Eq. (2) shows how to compute the mean and covariance matrix.

$$\text{mean}: \mu = E(x),$$
$$\text{cov}: c = E\left((x-\mu)(x-\mu)^T\right) \quad (2)$$

in which E is mathematical expectation, μ is mean, c is covariance matrix, x is the main matrix and $(x-\mu)^T$ is $(x-\mu)$ transposed. By using mean and covariance measures, a Gaussian model for skin color can be determined. After obtaining skin color model, you can use Gaussian function and Mahalanobis distance for face detection and recognize a pixel belong or non-belong to the considered model.

After face detection, eye detection is performed. Since eye color differs from other parts of face, this feature can be used for eye detection. For eye detection based on color feature, we can use horizontal and vertical projection [19]. Horizontal projection for an image with M rows and N columns is obtained from adding pixels intensities in each row according to Eq. (3) and vertical projection is obtained from adding pixels intensities in each column according to Eq. (4) as:

$$hp = \{hp_x \mid 1 \le x \le M\}$$
$$hp_x = \sum_{y=1}^{N} f(x,y) \quad (3)$$

$$vp = \{vp_y \mid 1 \le y \le N\}$$
$$vp_y = \sum_{x=1}^{N} f(x,y) \quad (4)$$

Another method of eye detection is by using eye corners. In [20], Harris algorithm is used to specify the pupil's center. In this algorithm, a mask in each point of image is considered. By displacing mask in different directions the mean variation rate of image intensity in each window, compared with the main window, is computed and the minimum variation is considered as corner response. Depending on the position of each point, three states are created:

- If the windowed image patch is flat (approximately constant in intensity), then all shifts will result in only a small change.
- If the window straddles an edge, then a shift along the edge will result in a small change, but a shift perpendicular to the edge will result in a large change.
- If the windowed patch is a corner or isolated point, then all shifts will result in a large change. A corner can thus be detected by finding where the minimum change produced by any of the shifts is large.

Harris algorithm is mathematically expressed as:

$$E_{x,y} = \sum_{u,v} w_{u,v} \mid I_{x+u,y+v} - I_{u,v} \mid^2 \quad (5)$$

in which w is a mask imposed on image, which is considered as circle mask with coefficients 1. I is the image gradient and E is the variation made by (x, y) displacement. (x, y) displacement in four directions includes the following set:

{(1 ,0),(1,1),(0, 1),(0,0)}

Harris corner detector introduces local maximum of $\min_{x,y}\{E_{x,y}\}$ that is more than a special threshold, as a corner. Equation (6) includes all possible little displacement as:

$$E(x,y) = Ax^2 + 2Cxy + By^2$$
$$A = (I_x)^2 \otimes w,$$
$$B = (I_y)^2 \otimes w,$$
$$C = (I_x I_y) \otimes w \quad (6)$$

in which, w is Gaussian mask, I_x, I_y is image gradient in is convolution $\otimes$x and y directions, respectively, and operator. E can be expressed as:

$$E(x,y) = (x,y)M(x,y)^T,$$
$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \quad (7)$$

in which, M is a 2*2 matrix in (x, y) coordinates and E is local auto–correlation function.

In [20], it has been shown that corner response, R, can be obtained in each image point as:

$$R = Det(M) - KTr^2(M)$$
$$Det(M) = \alpha\beta = AB - C^2 \quad (8)$$

in which K is constant. The corner response in corner area is positive, in edge area is negative and in flat area, has a minimal absolute magnitude. To reduce the computation time, the existing mask in Harris algorithm is neglected. So, firstly A, B, C are obtained as:

$$A = (I_x)^2$$
$$B = (I_y)^2$$
$$C = (I_x I_y) \tag{9}$$

After computing these measures, R is obtained by Eq. (8). Experiments show two points with the highest value of R in an eye area are eye iris corners [20].

The last stage in eye tracking system is eye tracking stage. One most favored and applicable method useful for real time applications is particle filter [21,22].

Particle filter is a method numerical and computational to obtain the probability density function of a random process and the estimation of a specific parameter from on it and in the event are an extended chain Monte Carlo methods [21,22], since the predominant method has been using Kalman filter or extended Kalman filter. These methods usually assume for the case when the model is linear and Gaussian noise, to obtain the optimum solution problem. But in other cases, depending on the remoteness of these assumptions are far from optimum. In other words, since the particle filter is a numerical method for estimating signal, nonlinear model or non-Gaussian noise for it does not matter. In such cases, the particle filters with high performance in various fields such as communication, signal processing, tracking and many other fields are used.

The idea using of particle filters for tracking an object in a sequence of images initially were introduced separated by a few research groups on the field and among those who were working in the fields of computer vision and image processing are known to be dense algorithm. This method mainly is based on the edge detection methods; another powerful method for target tracking in particle filters is using color distribution model target or histogram model for target tracking.

Color histogram of an object creates for its tracking many benefits. Specifically, in particle filters a target is tracked based on comparison its color histogram with histogram of samples location and using Bhattacharyya distance.

Color distribution $p_y = \{p_y(u)\}_{u=1,2,\ldots,m}$ where y is calculated within the area around the target.

$$p_y^{(u)} = f \sum_{i=1}^{I} K\left(\frac{\| y - x_i \|}{a}\right) \sigma[h(x_i) - u] \tag{10}$$

Where u is discrete element in color and m is the number of elements, K is a function of the weight that to samples farther away of the target gives lower weights,

H_x and H_y respectively are the length of half axes the horizontal and vertical in the area around the target, $a = \sqrt{H_x^2 + H_y^2}$ used to adjust the size of the area and f is a normalization factor that ensures that the following relation is established.

$$\sum_{u=1}^{m} P_y^{(u)} = 1 \tag{11}$$

In the tracking stage, the state estimated at every time step, based on the new observation function is updated. So a function of observation or measurement that is based on color distribution is required. A popular measure between two distributions p(u) and q(u) is the Bhattacharyya coefficient.

$$\rho[p,q] = \sum_{u=1}^{m} \sqrt{p^u q^u} \tag{12}$$

What the higher is this coefficient; the two distributions are similar together, so that for two distributions, normalized coincident is $\rho = 1$, which is expression of full implementation of two distributions. The distance between two distributions is defined as follows:

$$d = \sqrt{1 - \rho[p,q]} \tag{13}$$

Color based tracking algorithm for updating calculated the previous density function by particle filter is used from Bhattacharyya distance. Each sample of density function represents the area around the target that is given as follows:

$$S = \left[x, \overset{\bullet}{x}, y, \overset{\bullet}{y}, H_x, H_y, \theta\right]^T \tag{14}$$

Where x and y specify the center of the area around the target, $x^{\bullet}$ and $y^{\bullet}$ represent the speed, H_x and H_y are the length of half axes in the area around the target and θ is the rotation angle of the ellipse.

Dynamic model or state of model that represents the samples changes over time is defined as follows:

$$S_k = AS_{k-1} + W_{k-1} \tag{15}$$

Where A for motion model with constant acceleration is as follows:

$$A = \begin{bmatrix} 1 & \Delta & \circ & \circ & \circ & \circ & \circ \\ \circ & 1 & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & 1 & \Delta & \circ & \circ & \circ \\ \circ & \circ & \circ & 1 & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & 1 & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ & 1 & \circ \\ \circ & \circ & \circ & \circ & \circ & \circ & 1 \end{bmatrix} \tag{16}$$

W is the vector of Gaussian white noise whose covariance matrix is calculated by the following relation:

$$R = \begin{bmatrix} Ry & O_{4\times 3} \\ O_{3\times 4} & Re \end{bmatrix} \tag{17}$$

Where Ry is covariance matrix of the position and the speed of target and computed as follows:

$$Ry = \sigma_y \begin{bmatrix} \frac{\Delta^3}{3} & \frac{\Delta^2}{2} & \circ & \circ \\ \frac{\Delta^2}{2} & \Delta & \circ & \circ \\ \circ & \circ & \frac{\Delta^3}{3} & \frac{\Delta^2}{2} \\ \circ & \circ & \frac{\Delta^2}{2} & \Delta \end{bmatrix} \tag{18}$$

And covariance matrix of the area around the target, Re, is as follows:

$$Re = \begin{bmatrix} \sigma^2_{H(x)} & \circ & \circ \\ \circ & \sigma^2_{H(x)} & \circ \\ \circ & \circ & \sigma^2_{\theta} \end{bmatrix} \tag{19}$$

To weight the samples, Bhattacharyya coefficient is calculated between the histogram of the target and the histogram of the area that each sample shows. The area of each sample is determined with its state vector, $S^{(n)}$, and since samples that their color distribution is the most similar color distribution of the target are required, more weight should be given to them with less Bhattacharyya distance. The weight of the samples, shown here with $\Pi^{(n)}$, is defined as follows:

$$\Pi^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(1-\rho[p,q])}{2\sigma^2}} \tag{20}$$

Therefore weights of samples are calculated from a Gaussian distribution with the variance, σ^2 as how to select the appropriate number for σ depends on the desired tracking problem. As a result of this weighting, samples are that more compatible with target, greater probability have that in several iterations of algorithm remain and samples with less compatible, greater probability to be removed.
Finally, the estimation of the target state vector is obtained as follows:

$$E[S_k] = \sum_{n=1}^{N} \Pi_k^{(n)} S_k^{(n)} \tag{21}$$

Where N is the number of samples which is selected based on the accuracy required and volume of calculations which can be performed.

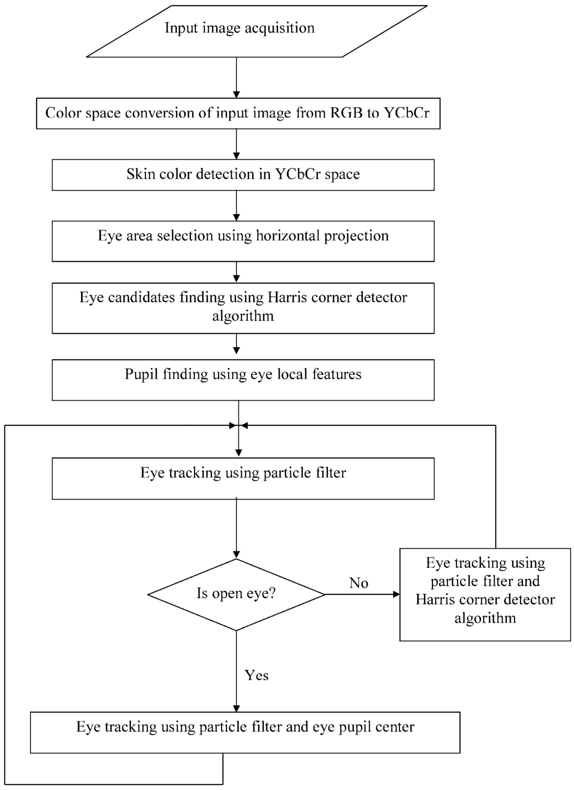


Fig.1 Proposed method algorithm for eye tracking

3. The Proposed Method

At the first, to make the intensity around ineffective in the proposed algorithm, the color space of input image is transformed from RGB to YCbCr by Eq. (1). Then a color model, according to Eq. (2) is defined for face and finally, the face area based on Gaussian model and Mahalanobis distance and using Eq. (3) is detected.
In the next stage, by using horizontal projection according to Eq. (3), the region including eyes is determined and divided into right and left areas, to precisely find the pupils' center and in each segmented area; the eye pupil's center and in each segmented area; the eye pupil's center is searched. This operation causes the searching area to be narrower. After that, by using vertical and horizontal projection via Eq. (3) and Eq. (4) the pupils' center is obtained. In the last stage, two particle filters track the eyes in parallel. Also, corners detections are used for eye tracking.
The proposed method includes four stages: face detection, detection of area including eyes, obtaining the eyes' pupils and corner detection, and eye tracking.

3.1 Face Detection

Since color is fixed when the head rotates and its processing is faster than other features of face, this feature is used for face detection. The most current color space is RGB which has many applications and is the standard space to present color images. But because of what is explained in section 2, the image is transformed into YCbCr space by Eq. (1). Fig.2 shows the differences between a person's images in color space RGB and YCbCr, Cb component and Cr component.

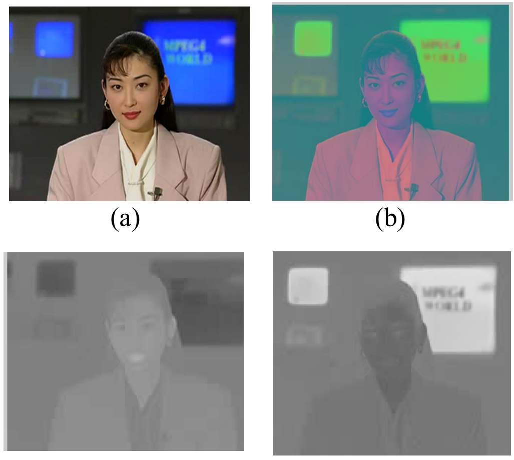


Fig.2 Differences between a person's images in various color spaces: (a) RGB space (b) YCbCr space (c) Cb component (d) Cr component

3.1.1 Finding Skin Color Model

To create skin color model, firstly a set of various skin images with different color and texture is prepared, then is transformed into YCbCr color space. After that by using Eq. (2), skin color model is obtained. In Fig 3 there are some samples of skin color.

3.1.2 Mahalanobis distance

After obtaining skin color model, by using a Gaussian model and Mahalanobis distance, every pixel's belong or non-belong to the model is determined. Figure (4) shows the result obtained by face detection.

Fig.3 Some samples of skin color

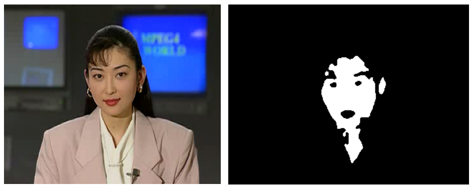


Fig.4 Result obtained by skin color detection

One of difficulties that there is in determining the skin color areas, similarity pixels color of image is with skin color model that belongs to any areas we are considering that this increases the computational complexity and in the proper functioning of the system occurs interference. For this propose, the largest area under skin color is selected as the original area, to do this work easily by calculating the area of each connection zone is possible. The region with largest area is considered as face and extracted from the image. When we have some region with the areas almost identical, the ratio of length of major to minor axis is used. In Fig. 5 can be seen the result of this process.

Fig.5 Result of face detection

3.2 Eye area extraction

After the face region detection, will turn to eye area extraction. This area includes two eyes, two eyebrows and distinctive features such as eyelids and the corners of the eyes. This area can be extracted from the horizontal projection of face area that can be calculated from Eq. (3). In Fig.6, horizontal projection of a face is observed.

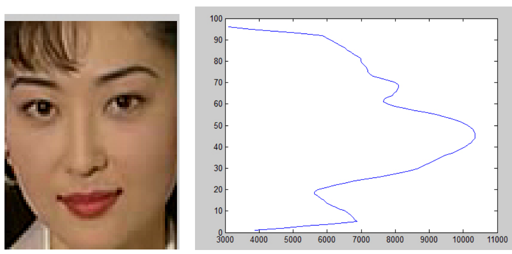


Fig.6 Horizontal projection of the face

As shown in Fig. 6, the first valley from above is related to place of pupil. Because of drastic changes in the eye area, the valley cannot easily be extracted. According to Fig. 6, because the forehead and below the eyes are brighter than pupil, instead of considering the minimum point on the graph, two local maximum point on the graph can be used to obtain the eye area.

From above face to down, it can be observed that two values of maximum horizontal projection, respectively, the one belonging to the forehead area and the other belongs to the area between eyes and mouth (the septum). According to the physiology of the face, this observation seems to be true because the flat area of the face, having the brightness level relatively high, is located in these regions. Figure 7 shows the result obtained from eye area extraction of face area.

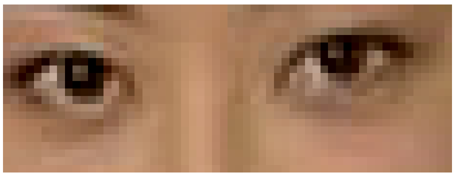


Fig.7 Eye area detection using horizontal projection

3.3 Finding Eye Candidates

To find eye candidates within the area obtained in the previous step is used Harris algorithm. Firstly, the image gradient is calculated and then, values of A, B, C from Eq. (9). By substituting A, B, C in Eq. (8), the value R can be calculated for each pixel of image. Points with the highest value of R can be introduced as candidates of iris. To remove unrelated candidates with iris, feature of color entropy in the area around iris is used. The area in size 50*50 pixels is considered for each of the candidates and entropy of color is calculated as follows:

$$H(S) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \qquad (22)$$

Where $p(x_i)$ and n are the probability of pixel x_i and the number of pixels per unit, respectively. Probability function, $p(x_i)$, for each pixel, is produced from counting the number of times that the pixel color intensity is repeated at selected portion to the total number of pixels. From candidates of eye, selected by Harris algorithm, two points are selected with the highest level of color entropy. Figure 8 shows that eyes have the highest degree of color entropy, as compared to the rest of the face components.

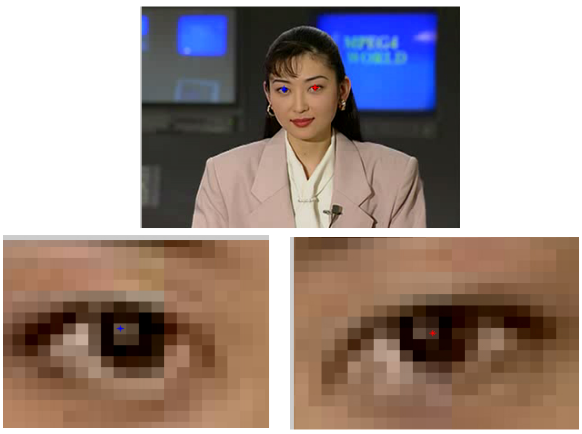


Fig.8 Degree of color entropy in Eye areas – from candidates selected by Harris algorithm, eyes that have the highest degree of color entropy are shown in the second row of images.

In Fig. 9, samples from detected corners of eyes are observed

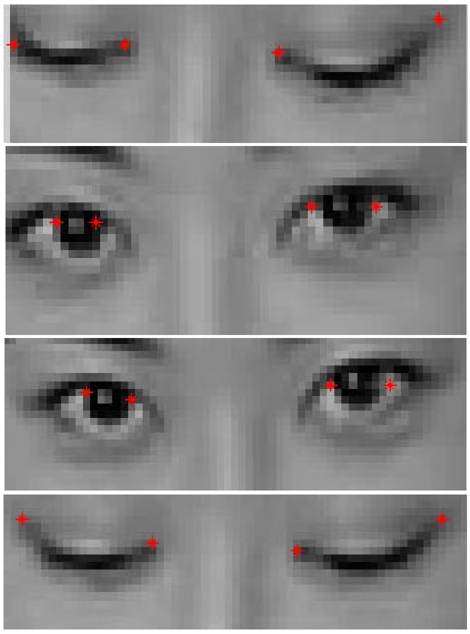


Fig.9 detected Corners of eyes

In Fig. 10, samples of pupil detection are observed.

Fig.10 Pupil detection

3.4 Finding the Pupil Center

By having a pixel from iris or pupil and having pupil circle, pupil center can be obtained. To find pupil center, lines intersection can be used. In this method, only one row and column of eye is processed instead of total eye area. So, a line is drown, along with x and y axis to the border edge of iris and sclera. As shown in Fig.11, if the intersection center of drown lines is considered as (x, y) and the end points of lines are considered as $(x_1, y'), (x_2, y'), (x', y_1), (x', y_2)$, the pupil center can be obtained as:

$$x_c = \frac{x_1 + x_2}{2},$$
$$y_c = \frac{y_1 + y_2}{2} \qquad (23)$$

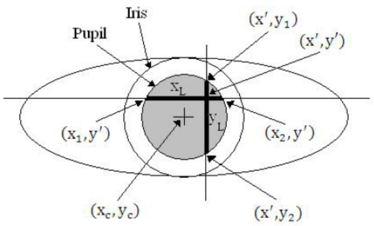


Fig.11 Using lines intersection to find pupil center

3.5 Eye tracking

In most applications, after eye detection, tracking of eye or face movements is considered. The same problems in eye detection state are in eye tracking, by a difference that in the tracking cases, the problems appears stronger and there is more destructive effect on tracking process. As an example, changing the environment light conditions cause loss of eye position on the image and consequently eye detection process is repeated.

For eye tracking, in addition to the estimation and location of eye positions in sequential frames, head and face movements should be considered and a method not sensitive to these changes and problems should be proposed. In addition to head, face components like eyelids have independent movement which should be considered.

To support head movements, the eye itself was not tracked and an area containing eye was tracked.

Among proposed methods, is used from particle filter. In the proposed method, two filters used in parallel, track the eyes separately because as explained in section 3-3, the eyes are separately recognized. In Fig. 12, results of eye tracking using particle filter and pupil detection in the different frames can be observed.

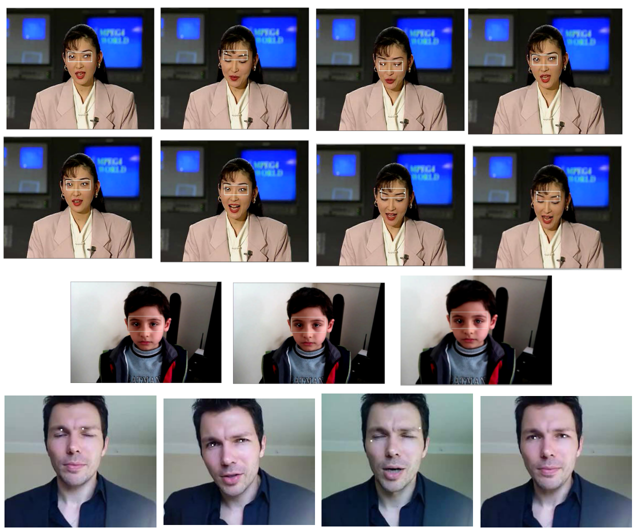


Fig.12 Eye tracking using particle filter

A problem here is that if both eyes are closed, Particle filter does not work. To overcome this problem, besides particle filter, the eye corners features are also used. To perform this, in each left and right area, two corners are recognized. To find the corners, Harris algorithm, explained in section 2 has been used. So, the proposed method for tracking, is a combination of particle filter and corner detection.

In proposed method, this problem is considered that in particle filter for each of the particles there is probability degree that much particles are farther from the area around the target, this probability is lower number; the average of probabilities of all particles are calculated when the eye is open, the average probabilities of particles is a high number but when the eyes are closed, the average probabilities of particles is a low number that this probability degree of a particle indicates the degree of similarity of each particle to the reference histogram. Threshold value is considered experimental. If in working frame, the average probabilities of the particles were higher than the threshold, eye is open and eye tracking does not fail in this case using particle filter. But if the average probabilities of particles were lower than the threshold, eye is close and in this case, eye tracking fail using particle filter. To overcome this problem, Harris corner detection algorithm is used and as mentioned before, corners of eyes are detected and used in eye close state for eye tracking.

Although in this method, the required hardware has been increased and for each eye, an additional filter has been used, but tracking with head and face rotation existence, make up existence and eyes closure is performed precisely.

4. Experimental results

Fig.13 shows the results of eye tracking and pupil detection using proposed method in different frames. In this figure, it can be observed that eye closure problem is resolved and the eye in closing state is also tracked. Fig. 14 and 15 are supported the results of eye tracking and pupil detection using proposed method in various states of head and face movements existence, make up existence and eye closure.

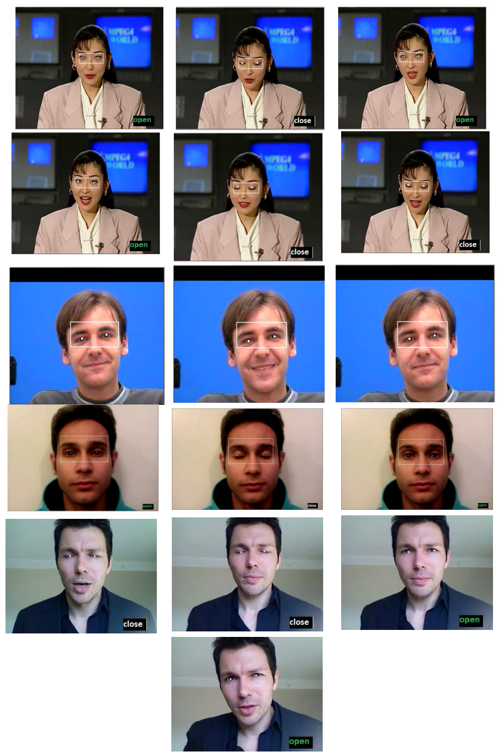
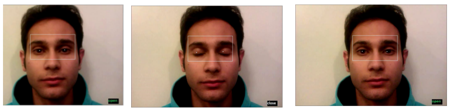
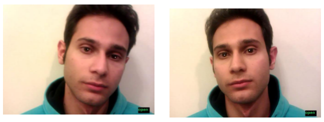


Fig.13 Eye tracking using proposed method

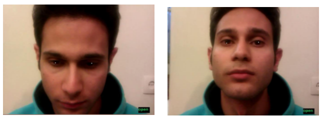
Fig.14 Results of the proposed method for eye tracking with make up



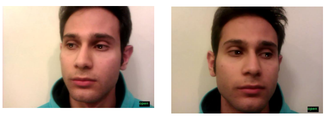
(a)



(b)



(c)



(d)

Fig.15 results of eye tracking in different frames using proposed method in:
a) Normal state b) Head rotation state horizontally
c) Head rotation into high and low d) Head rotation state vertically

Comparison between the proposed method and other ones is difficult because of their performing in different platforms and databases. So, the main parts of these methods are theoretically compared. In Table 1, the proposed method is compared with two similar methods.

Table 1: Comparison between the proposed and two different methods

	Used Method for detection and tracking	Detection in different light conditions, head rotation, make up and eye closure	Image with high resolution
Paper[5]	Entropy of eye and darkness of iris	Supports the different light conditions , head rotation	Need
Paper[12]	Eye form structure and corners	Supports head rotation	Need
Proposed Method	Combine color intensity around the eye, particle filter and corner detection	Supports head rotation to 35 degree, different light conditions, make up and eyes closure	No Need

Following results are concluded from Table1:

In [5], entropy of eye and darkness of iris is used. This method supports various light conditions, head rotation and in case of existing make up and eyes closure tracking is not possible. Also the images used by this method should be in high resolution.

In [12], the eye form, structure and corners are used for detection. This method only supports head rotation and in the case of different light conditions, make up and eyes closure, it cannot track. Also, the images used by this method should have high resolution.

In the proposed method, the combined color intensity around the eye, particle filter and corner detection are used for detection and tracking. The proposed method supports various light conditions, head rotation, make up and eye closure. Also, in this method, there is no need to the high resolution images.

5. Conclusions

In this paper, a new method for human eye tracking using particle filter has been presented. Considering the experiments in different conditions including images changing light conditions, head rotation, make up and eye closure, good results were obtained. Also, experimental results show the proposed method has low computational complexity and maximum stability.

References

[1] S. Yamamoto and V. G.Moshnyaga, "Algorithm Optimizations for Low-Complexity Eye Tracking", *IEEE International Conference on Systems, Man, and Cybernetics,pp.18-22,* 2009.

[2] K.N. Kim and R.S. Ramakrishna, "Vision-Based Eye-Gaze Tracking for Human Computer Interface", *IEEE International Conference on Systems, Man, and Cybernetics,* Vol. 2, pp. 324-329, 1999.

[3] R. Kothari and J.L. Mitchell, "Detection of Eye Locations in Unconstrained Visual Images", *IEEE International Conference on Image Processing,* Vol. 3, pp. 519-522, 1996.

[4] W. Liting, D. Xiaoqing and F. Chi, "Face Live Detection Method Based on Physiological Motion Analysis", *Tsinghua Science & Technology,* Vol. 14 , Iss. 6, pp. 685-690, 2009.

[5] M. Hassaballah and S. Ido, "Eye Detection Using Intensity and Appearance Information", *IAPR Conference on Machine Vision Applications,* pp. 346-349, 2009.

[6] Z. Zhu, K. Fujimura and Q. Ji. "Real-Time Eye Detection and Tracking under Various Light Conditions", *Eye Tracking Research & Applications Symposium,* pp. 139-144, 2002.

[7] A. Hyrskykari,"Utilizing Eye Movements: Overcoming Inaccuracy while Tracking the Focus of Attention during Reading", *Computers in Human Behavior,* Vol. 22, Iss. 4, pp. 657-671, 2006.

[8] K. Peng, L. Chen, S. Ruan and G. Kukharev, "A Robust Algorithm for Eye Detection on Gray Intensity Face without Spectacles", Journal of Computer Science & Technology, Vol. 5, No. 3, 2005.

[9] K.C.K. Lee, J. Schiffman, B. Zheng, W.C. Lee and H.V. Leong, "A System for Tracking Nearest Surrenders in Moving Object Environments", *IEEE international Conference on Pervasive Services,* Lyon, 2006.

[10] Z. Qian and D. Xu, "Automatic Eye Detection Using Intensity Filtering and K-means Clustering", *Pattern Recognition Letters,* Vol. 31, Iss. 12, pp. 1633-1640, 2010.
[11] B. Kroon, S. Maas, S. Boughorbel and A. Hanjalic, "Eye Localization in Low and Standard Definition Content with Application to Face Matching", *Computer Vision and Image Understanding,* Vol. 113, Iss. 8, pp. 921-933, 2009.
[12] Y. Tian, T. Kanade and J.F. Cohn, "Dual State Parametric Eye Tracking", *IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 110-115, 2000.
[13] F.H.C. Tivive and A. Bouzerdoum, "A Fast Neural-Based Eye Detection System", *iEEE International Symposium on Intelligent Signal Processing and Communication Systems,* pp. 641-644, 2005.
[14] Z. Niu, S. Shan, S. Yan, X. Chen and W. Gao, "2D Cascaded Adaboost for Eye Localization", *IEEE International Conference on Pattern Recognition,* Vol. 2, pp. 1216-1219, 2006.
[15] D. Comaniciu, V. Ramesh and P. Meer, "Kernel-Based Object Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 25, Iss. 5, pp. 564-577, 2003.
[16]Z. Zhu, Q. Ji, K. Fujimura and k. Lee "Combining Kalman Filtering and Mean Shift for Real Time Eye Tracking under Active IR Illumination", *IEEE International Conference on Pattern Recognition,* Vol.4, 2002.
[17] G. Veneri, P. Federighi, F. Rosini, A. Federico and A. Rufa, "Influences of Data Filtering on Human - Computer Interaction by Gaze-Contingent Display and Eye-Tracking Applications", *Computers in Human Behavior,* Vol. 26, Iss. 6, pp. 1555-1563, 2010.
[18] X. Long, O.K. Tonguz and A. Kiderman, "A High Speed Eye Tracking System with Robust Pupil Center Estimation Algorithm", *IEEE International Conference Engineering in Medicine and Biology Society,* pp. 3331-3334, 2007.
[19] Z. Zhu and Q. Ji, "Eye and Gaze Tracking for Interactive Graphic Display", *Machine Vision and Applications,* Vol. 15, Iss. 3, pp. 139-148, 2004.
[20] T. Liu, S. Zhu, "Eyes Detection and Tracking Based on Entropy in Particle Filter", *International Conference on Control and Automation,* Vol. 2, pp. 1002-1007, 2005.
[21] K. Nummiaro, E. Koller-Meier and L.Van Gool, "An Adaptive Color-Based Particle Filter", Image and Vision Computing, Vol. 21, Iss. 1, pp. 99-110, 2003.
[22]Y. Kuo, J. Lee and S. Kao, "Eye Tracking in Visible Environment", Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 114-117, 2009.

Elham Ghasemi-Dehkordi received the B.S. degree in computer engineering from Mobarakeh Branch, Islamic Azad University, Mobarakeh, Iran, in 2009. He is currently a graduate student in the Department of Computer Engineering at Dezfoul Branch, Islamic Dezfoul University, Dezfoul, Iran His research interests include image processing and pattern recognition.

Mahmoud Mahlouji received the B.S. degree in telecommunications engineering from Sharif University of Technology, Tehran, Iran, in 1990, the M.Sc. degree in electronics engineering from Sharif University of Technology, Tehran, Iran, in 1993, and the Ph.D. degree in telecommunications engineering from Science and Research Branch, Islamic Azad University, Tehran, Iran, in 2008. At present he is an assistant professor of the Electrical and Computer Engineering Department, Kashan Branch, Islamic Azad University, Kashan, Iran. His current interests are in image processing, pattern recognition, neural networks, computer vision, and iris recognition.

Hossein Ebrahimpour-Komleh is currently an assistant professor at the Department of Electrical and Computer Engineering at the University of Kashan, Kashan, Iran. His main area of research includes Computer Vision, Image Processing, Pattern Recognition, Biometrics, and Robotics, Fractals, chaos theory and applications of Artificial Intelligence in Engineering. He received his Ph.D. degree in Computer engineering from Queensland University of technology, Brisbane, Australia in 2006. His Ph.D. research work is on the Fractal Techniques for face recognition. From 2005 to 2007 and prior to joining the University of Kashan, he was working as a Post-doc researcher in the University of Newcastle, NSW, Australia and as a visiting scientist in CSRIO Sydney. Hossein Ebrahimpour-Komleh has B.Sc. and M.Sc. degrees both in computer engineering from Isfahan University of Technology (Isfahan, Iran) and Amirkabir University of Technology (Tehran, Iran.) respectively. He has served as the editorial board member and reviewer of several journals and international and national conferences.

Neighborhood Crossover Operator： A new operator in Gravitational Search Algorithm

Zhongping shang [1]

- [1] School of Continuing Education, Yanshan University , Qinhuangdao Hebei 066004, China

Abstract

In order to improve exploration and exploitation of the Gravitational Search Algorithm (GSA) for solving more complicated problems, Neighborhood Crossover Operator (NCO) is applied to GSA. In GSA, the gravitational force guides the masses. As the force absorbs the masses into each other, if premature convergence happens, there will not be any recovery for the algorithm, the NCO help the GSA recover from premature convergence and improve the local search ability. The improve GSA has been evaluate on 23 functions, compared with the GSA, the obtained results confirm the high performance of the proposed method in solving various nonlinear functions.

Keyword: *Optimization; Gravitational search algorithm; Neighborhood crossover operator; Heuristic search algorithm; nonlinear functions.*

1. Introduction

Optimization is an old problem, the pursuit of the optimal target has been a human ideal, Some scholars have proposed a lot of feasible effective optimization methods on the problem of the exploration and exploitation, the exploration is the ability of expanding search space and investigating the search space for finding new and better solutions, the exploitation is the ability of finding the optima around a good one. In most heuristic algorithm, the abilities of exploration and exploitation are applied with special operator, the special operator can improve the local search ability.

Heuristic algorithms simulate physical or biological processes，such as, Genetic Algorithms [1-4], Simulated Annealing Algorithm [5-7], Artificial Immune Algorithm [8, 9], Ant Colony Algorithm [10-12], Particle Swarm Optimization [13-16], Gravitational Search Algorithm[17,18]. Those methods have made great successful.

Xun et al. [1] indicates the importance of the two new genetic operators is designed to overcome the defect of genetic algorithm in local searching, which combines with uniform crossover. New operator has turned for other heuristic algorithm. For example, Wu et al. [7] add mutation operator to a hybrid simulated annealing algorithm solving the manufacturing cell formation problem. An improved artificial immune algorithm with a dynamic threshold is presented; the calculation for the affinity function in the real-valued coding artificial immune algorithm is modified through considering the antibody's fitness and setting the dynamic threshold value [8]. Two new efficient and robust ant colony algorithms are proposed [10]. It is two new and reasonable local updating rules that make them more robust and efficient. While going forward from start point to end point of a tour, the ants' freedom to make local changes on links is gradually restricted. Chen et al. [13] used a local search to improve the Particle swarm optimization. For increasing the diversity of particles, Jiang et al. [14] utilized a mutation operator. Groenwold et al. [15] divided the population to sub-divisions, applies particle swarm optimization to them separately and then combines the results of the sub-divisions to transfer the information. In the improved particle swarm optimization [16], a new velocity strategy equation with a scaling factor is proposed, and the Constriction Factor Approach (CFA) utilizes the value analysis to control the system behavior.

GSA [17] is the newest algorithm introduced by Rashedi et al in 2009. It is inspired by the law of gravity and mass interactions. In this algorithm, the gravitational force guides the masses. As this force absorbs the masses into each other, if premature convergence happens, the algorithm loses its ability to explore and then becomes inactive. Therefore, the Neighborhood Crossover Operator should be added to GSA in order to increase its flexibility for solving more complicated problems.

This paper is organized as follows. In the first section, some Heuristic optimums are introduced, In the second section, “Gravitational Search Algorithm” provides a brief review. In the third section, Neighborhood Crossover Operator is described. A comparative study is presented in “Experimental Results” and finally in the last section, the paper is concluded.

2. Gravitational search algorithm

The GSA is a novel meta-heuristic stochastic optimization algorithm introduced by Rashedi et al. in 2009. It bases on the metaphor of gravitational interaction between masses and is inspired by the Newton theory. Every particle attracts every other particle with a gravitational force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them in the universe. The heavy masses are good solutions of the problem. In other words, each mass correspond to a solution,and the heuristic algorithm is navigated by properly adjusting the masses and gravitational. With the passage of time, the masses will be attracted by the heaviest mass which it corresponds to an optimum solution in the search space, the heaviest mass which it represents an optimum solution in the search space.

In GSA, consider a system with N agents (masses) in which the position of the agent i is defined by:

$$X_i = \left(x_i^1, x_i^2 \cdots\cdots x_i^n \right), \ i = 1, \cdots\cdots, N \quad (1)$$

Where x_i^n presents the position of agent i in dimension. n is the search space dimension.

After evaluating the current population fitness, the mass of agent is calculated for a minimization m, as follows:

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \quad (2)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)} \quad (3)$$

Where $fit_i(t)$ defined the fitness value of agent i at time t, $best(t)$ and $worst(t)$ are the best and worst fitness of all agents

$$best(t) = \min_{j \in \{1, \cdots N\}} fit_j(t) \quad (4)$$

$$worst(t) = \max_{j \in \{1, \cdots N\}} fit_j(t) \quad (5)$$

To evaluate the acceleration of an agent i at time t in direct d th, the next velocity of an agent is considered as a fraction of its current velocity added to its acceleration, velocity and position of the agent i at time t .Therefore, $a_i^t(t)$, $v_i^d(t+1)$, $x_i^d(t+1)$ is given as follows:

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} = \sum_{j=k\,best,\, j \neq i}^{n} rand_j G(t) \times \frac{M_j(t)}{\sqrt{\sum_{d=1}^{n} \left(X_i^d - X_j^d \right)^2 + \varepsilon}} \left(x_j^d(t) - x_i^d(t) \right) \quad (7)$$

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \quad (8)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (9)$$

Where $rand_i$ and $rand_j$ are two uniformly distributed random numbers in the range of $[0,\ 1]$, ε is a small value to avoid division by zero, n is the dimension of the search space, The set of first K agents with the best fitness value and biggest mass is K_{best}. K_{best} is a function of time, initialized to K_0 at the beginning and decreasing with time. K_0 is set to N (total number of agents) and is linearly decreased to1. G is a decreasing function of time that is set to 1 at the beginning and decreases linearly towards zero with lapse of time. It is noted that $X_i = \left(x_i^1, \ldots, x_i^d, \ldots x_i^n\right)$ indicates the position of agent i in the search space, which is a candidate solution. The different steps of the proposed algorithm are given by Figure 1.

We compare IGSA with GSA, in all cases, population size is set to $N = 50$. The dimension is $N = 30$ and maximum iteration ($t_{\max}$) is 1000 for functions of the Tables 1- 3.

In both forms of IGSA with GSA, G is set using Eq. (10), where G_0 is set to 100, α is set 20 and T is the total number of iterations

$$G = G_0\left(-\alpha \frac{t}{T}\right) \quad (10)$$

Furthermore, K_0 is set to N (total number of agents) and is decreased linearly to 1.

3. Neighborhood crossover operator

In GSA, as the gravitational force absorbs the masses into each other, if premature convergence happens, the algorithm loses its ability to explore and is inactive. So a new operator is added to GSA in order to improve its flexibility to solve complex problems.

$$X_i = rand_i \times X_i + U(-1,1) \times \left(rand_i \times X_i - X_i\right), i = 1,2,3 \ldots \quad (10)$$

X_i is the position of ith agent, $U(-1,1)$ is a random number in the interval $(-1,1)$. $rand_i$ is a random number in the interval [0, 1],

We take into account the global search ability of the gravitational search algorithm and the local search ability of the neighborhood crossover operator. To achieve both the advantages of complementary, we introduce a factor,

$$w = w_{\max} - t \times \left(w_{\max} - w_{\min}\right) / t_{\max} \quad (11)$$

$w_{\max}$ and $w_{\min}$ are the maximum and the minimum of the scale factor respectively, t is the current number of iterations and $t_{\max}$ is the maximum number of iterations. w is a scale factor, r is a random number in the interval [0,1], where $r < w$, the gravitational search algorithm is used to search the space , where $r \geq w$, the neighborhood crossover operator is used to generate some individuals. In the early stages of searching, considering the search efficiency of the solution space, global search ability of the gravitational search algorithm should be fully utilized, the gravitation optimization algorithm guide the neighborhood crossover operator searches near the front end, with the depth of searching, the algorithm should be gradually change into the depth from breadth, to ensure that the solutions converge to the front.

The different steps of the algorithm are the followings:

(a) Search space identification, $t = 0$;
(b) Randomized initialization
(c) $X(t)$ for $i = 1, \ldots N$;
(d) Fitness evaluation of agents;
(e) Update , $Best(t)$, $worst(t)$
(f) and $M_i(t)$ for $i = 1, \ldots N$;
(g) Calculation of acceleration and velocity;
(h) Updating agents' position to yield

(i) $X_i(t+1)$ and $i=1,\ldots N, t=t+1$;
(j) Neighborhood Crossover Operator is
(k) on the $X_i(t)$;
(l) Repeat steps c to g until the
(m) stop criteria is reached;
(n) End;

Fig.1 :Pseudo code of the IGSA

4. Experimental results

To evaluate the performance of the IGSA, we apply it to 23 standard benchmark functions [17]. The standard functions are presented in the nest section.

Table1:Unimodal test functions.

Test function	s
$f_1(\vec{x}) = \sum_{i=1}^{n} x_i^2$	$[-100,100]$
$f_2(\vec{x}) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	$[-10,10]^n$
$f_3(\vec{x}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2$	$[-100,100]$
$f_4(\vec{x}) = \max_i \{ \lvert x_i \rvert, 1 \le i \le n \}$	$[-100,100]$
$f_5(\vec{x}) = \sum_{i=1}^{n-1} \left[100\left(x_{i+1} - x_i^2\right)^2 + (x_i - 1)^2 \right]$	$[-30,30]^n$
$f_6(\vec{x}) = \sum_{i=1}^{n} \left([x_i + 0.5] \right)^2$	$[-100,100]$
$f_7(\vec{x}) = \sum_{i=1}^{n} i x_i^4 + random[0,1)$	$[-1.28,1.28$

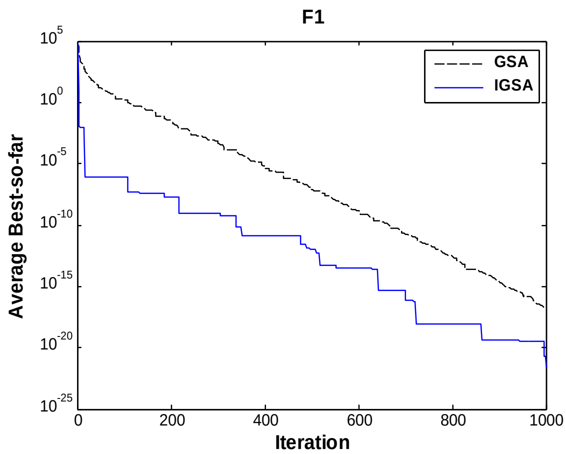
Table2:Multimodal test functions.

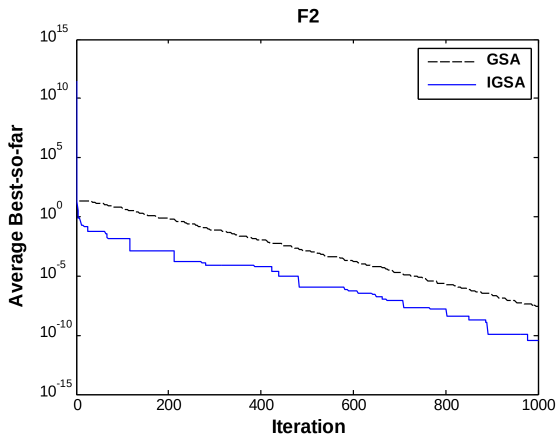
Test function	s
$f_8(\vec{x}) = \sum_{i=1}^{n} -x_i \sin\left(\sqrt{\lvert x_i \rvert}\right)$	$[-500,500]'$
$f_9(\vec{x}) = \sum_{i=1}^{n} \left(x_i^2 - 10\cos(2\pi x_i) + 10 \right)$	$[-5.12,5.12$
$f_{10}(\vec{x}) = -20exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$	$[-32,32]^n$
$f_{11}(\vec{x}) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right)$	$\left[-5.12, 5.12\right]^n$
$f_{12}(\vec{x}) = \frac{\pi}{n}\left\{10\sin(\pi y_1) + \sum_{i=1}^{n-1}(y_i - 1)^2 \times \left[1 + 10\sin^2(\pi y_{i+1})\right] + (y_n - \right.$ $+ \sum_{i=1}^{n} u(x_i, 10, 100, 4)$; $y_i = 1 + \frac{x_i + 1}{4}$; $u(x_i,a,k,m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a < x_i < \\ k(-x_i - a)^m & x_i < -a \end{cases}$	$[-50,50]^n$
$f_{13}(\vec{x}) = 0.1\left\{\sin^2(3\pi x_1) + \sum_{i=1}^{n}(x_i - 1)^2 \times \left[1 + \sin^2(3\pi x_i + 1)\right] + (x_n - 1)^2\right\}$ $+ \sum_{i=1}^{n} u(x_i, 5, 100, 4)$	$[-50,50]^n$

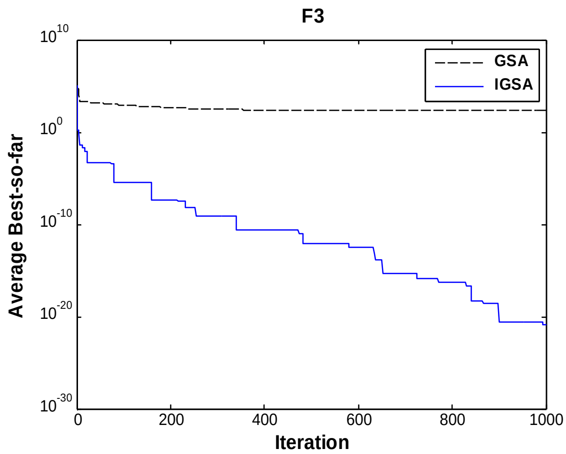
Table3:Mutimodal test functions with fix dimension

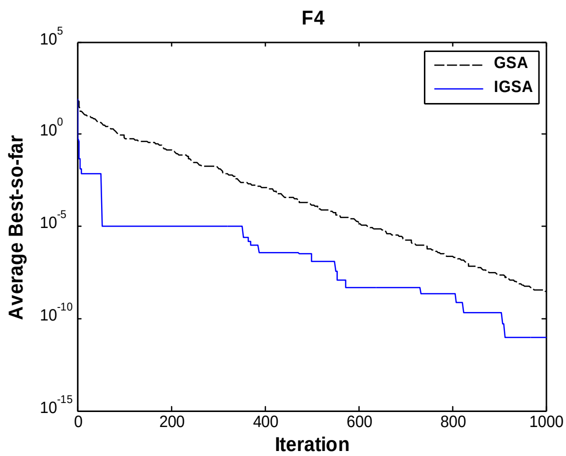
Test function	s

$f_{14}(\vec{x}) = \left(\frac{1}{500} + \sum_{j=1}^{25} \frac{1}{\sum_{i=1}^{2}(x_i - a_j)^6}\right)^{-1}$	$[-65.53, 65.53]^2$
$f_{15}(\vec{x}) = \sum_{i=1}^{11} \left(a_i - \frac{x_1(b_i^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4}\right)^2$	$[-5,5]^4$
$f_{16}(\vec{x}) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1 x_2 - 4x_2^2 + 4x_2^4$	$[-5,5]^2$
$f_{17}(\vec{x}) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos x_1 + 10$	$[-5,10] \times [0,15]$
$f_{18}(\vec{x}) = \left[1 + (x_1 + x_2 + 1)^2 \times (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)\right] \times 30 + (2x_1 - 3x_2)^2 \times \left(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2\right)$	$[-5,5]$
$f_{19}(\vec{x}) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{3} a_{ij}(x_j - p_{ij})^2\right)$	$[0,1]^3$
$f_{20}(\vec{x}) = \sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6} a_{ij}(x_j - p_{ij})^2\right)$	$[0,1]^6$
$f_{21}(\vec{x}) = \sum_{i=1}^{5}\left[(\vec{x} - a_i)(\vec{x} - a_i)^T + c_i\right]^{-1}$	$[0,10]^4$
$f_{22}(\vec{x}) = \sum_{i=1}^{7}\left[(\vec{x} - a_i)(\vec{x} - a_i)\right.$	$[0,10]^4$
$f_{23}(\vec{x}) = \sum_{i=1}^{10}\left[(\vec{x} - a_i)(\vec{x} - a_i)\right.$	$[0,10]^4$









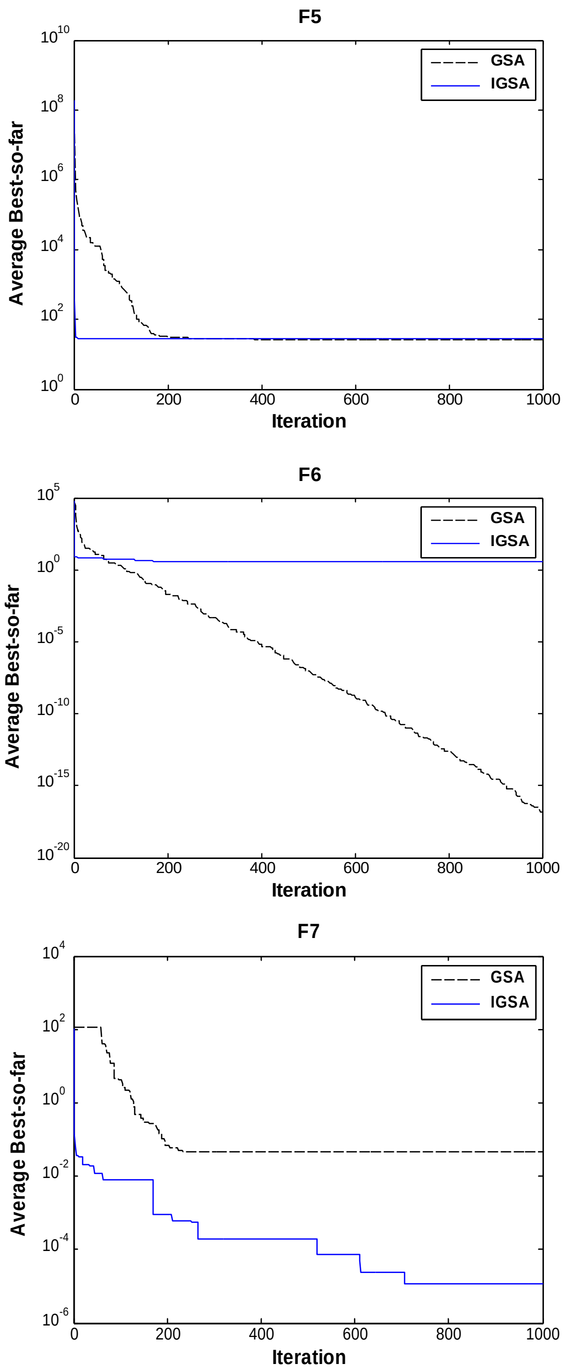
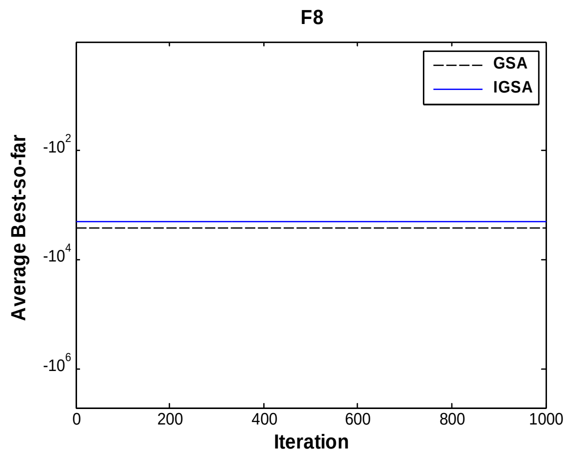
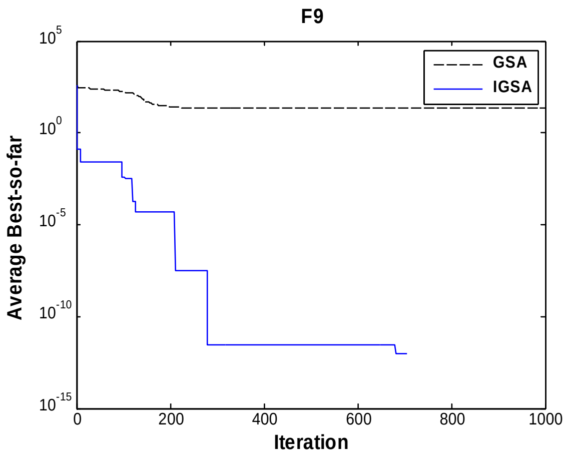


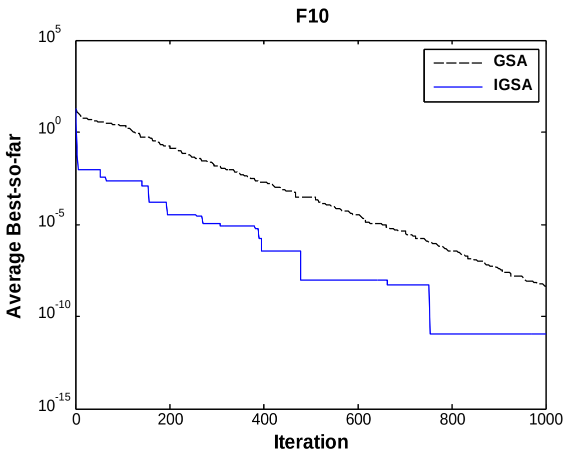
Fig.2.Comparison of performance of IGSA and GSA for minimization with n = 30

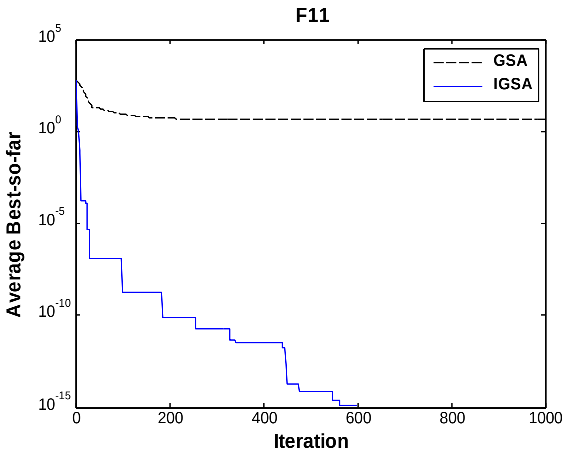
Table4: Minimization result of benchmark functions, with n=30, t_{max}=1000.	
Average Best-so-far	
IGSA	GSA
F1 7.3275e-022	2.0527e-017
F2 1.1859e-010	2.3129e-008
F3 3.0936e-021	261.7802
F4 1.0986e-011	3.3463e-009
F5 28.7738	30.1717
F6 4.3930	2.0821e-017
F7 3.3941e-005	0.0234
Median Best-so-far	
IGSA	GSA
2.5578e-022	2.0172e-017
7.0418e-011	2.2862e-008
3.7860e-022	252.2580
5.0181e-012	3.2572e-009
28.7873	25.9890
4.3602	2.0705e-017
1.9619e-005	0.0220

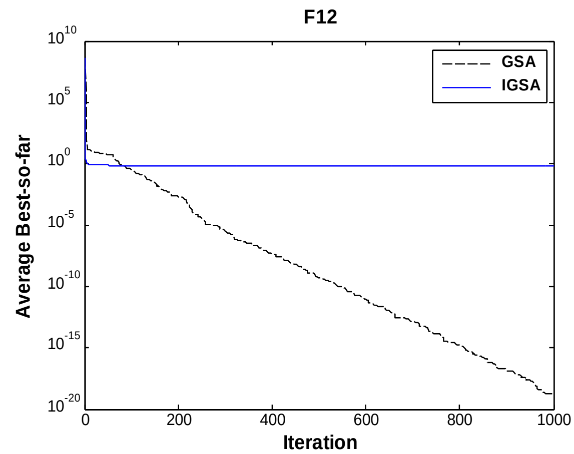
Functions of the table1 are unimodal functions. In this case, the convergence rate of the search algorithm is more important than the final results for functions F1 to F7, because there are other methods particularly designed to optimize F1 to F7 functions. The results are averaged over 30 runs under different random seeds , the average best-so-far solutions and median of the best solutions are reported for unimodal functions in Table 4. In Functions 1, 2, 3， 4 and 7, IGSA has a very powerful ability to explore and exploit the search space and also has a high convergence rate. So, these characteristics significantly cause good results. In Function5, both algorithms could find the optimum, IGSA is a little better than GSA in exploiting. In Table 4, the progress of the average best-so-far solution of IGSA and GSA over 30 runs, for F1， F3, F4, F6 and F7. The good convergence rate of GSA could be concluded from Fig.2. According to these figures, IGSA tends to find the global optimum faster than GSA and hence has a higher convergence rate.











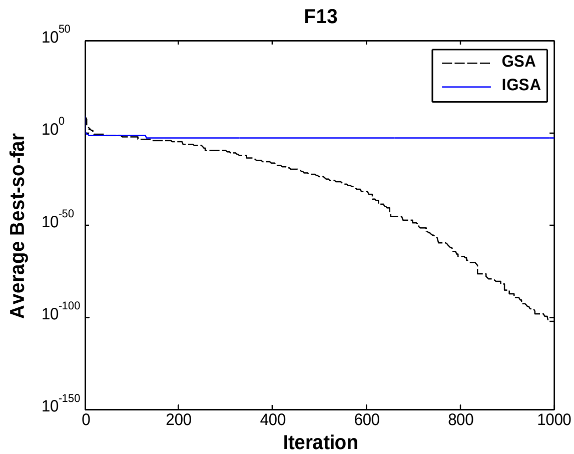
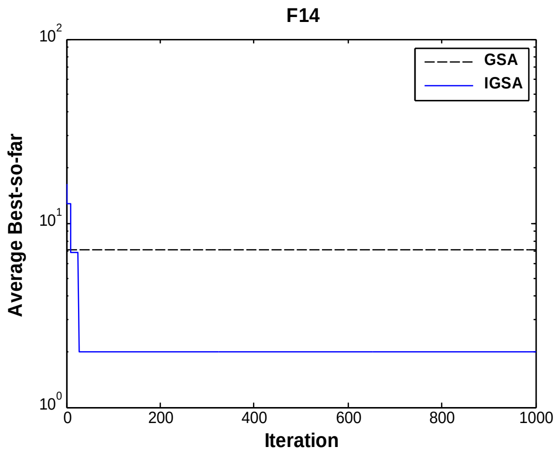
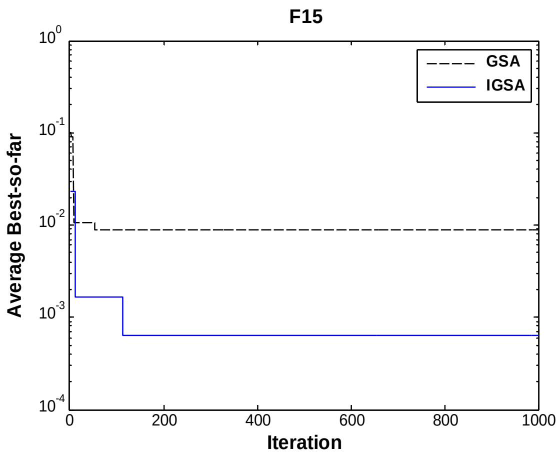


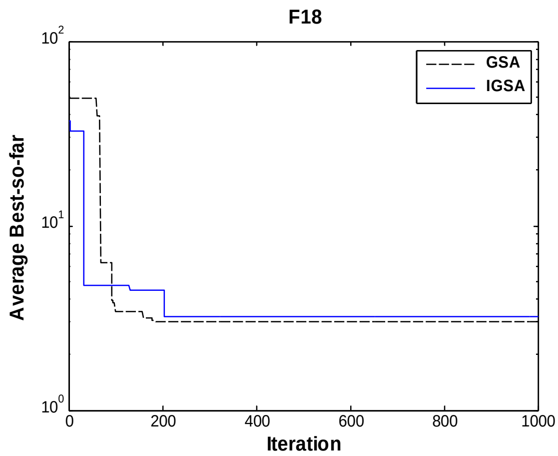
Fig.3.Comparison of performance of IGSA and GSA for minimization with n = 30.

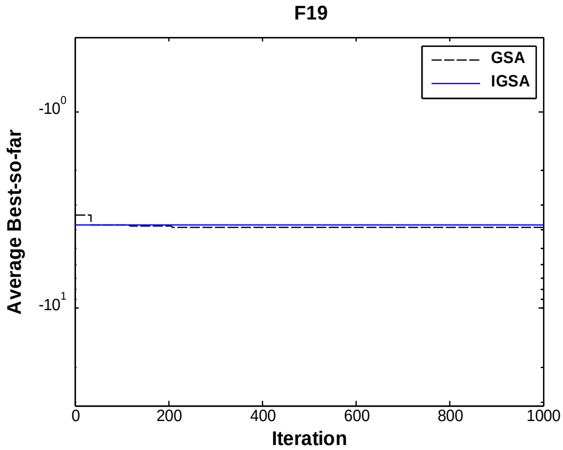
Table5: Minimization result of benchmark functions, with n=30, t_{max}=1000.	
Average Best-so-far	
IGSA	GSA
F8 -2.5712e+003	-2.8416e+003
F9 0	17.4781
F10 2.1798e-011	3.7009e-009
F11 0	4.0144
F12 0.5753	0.0362
F13 0.0199	2.1797e-032
Median Best-so-far	
IGSA	GSA
-2.4472e+003	-2.8452e+003
0	17.9093
1.4309e-011	3.7478e-009
0	4.1021
0.5978	1.4923e-019
0.0173	1.3498e-032

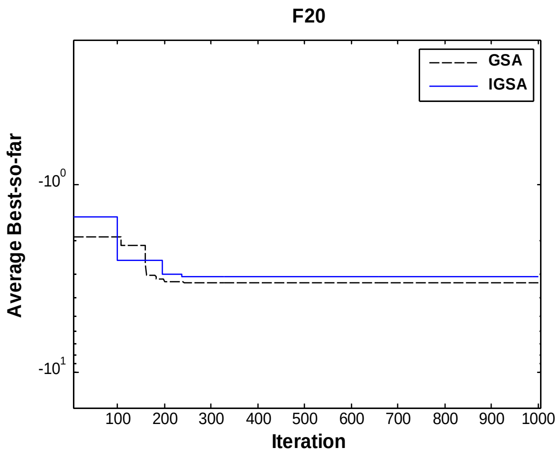
Multimodal functions have many more local minima and are almost too difficult to optimize. For multimodal functions, the final results are important, because they reflect the ability of the algorithm to escape from poor local optima and locate a near global optimum. Experiments of table2 functions are carried out, the results are averaged over 30 different runs and the average best-so-far solutions and median of the best solutions are reported for these functions in Table 5. The largest difference in performance between IGSA and GSA occurs with these multimodal functions for the robust power of the proposed algorithm to explore and exploit. In Functions 9, 10 and 11, IGSA performs significantly better than GSA in exploring and exploiting, and it exactly finds the optimum. In F13 and F12, GSA is better than IGSA, in F8 GSA is a little better than IGSA in exploiting.The results of the average best-so-far solution over 30 runs are shown in Figs. 3 and table 5.

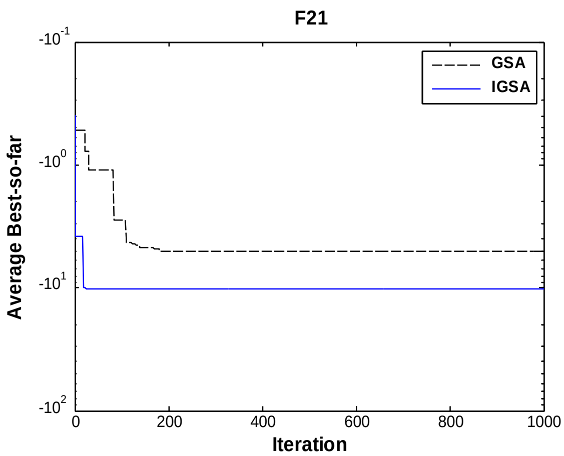












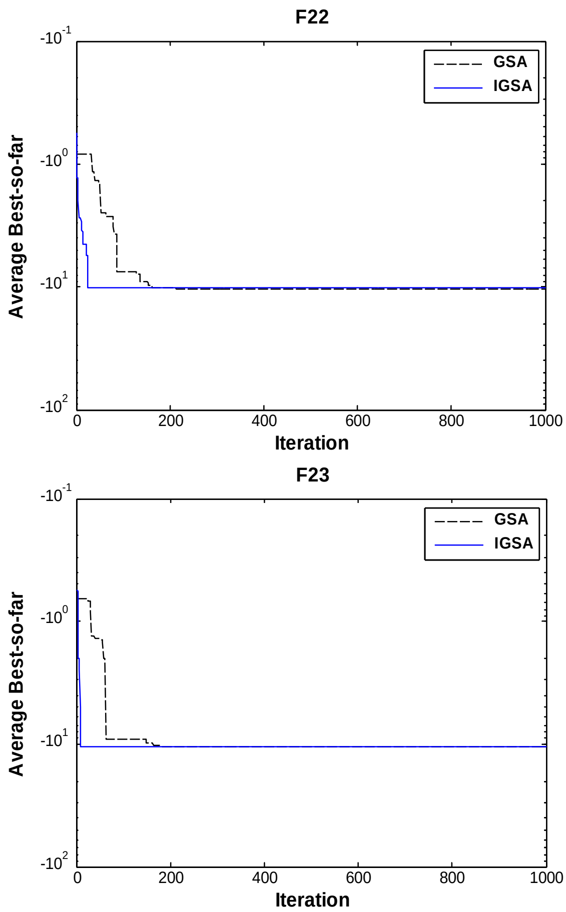


Fig.4.Comparison of performance of IGSA and GSA for minimization

Multimodal Low-Dimensional Functions. Table 6 shows a comparison between GSA and IGSA on the multimodal low dimensional each mark functions of F14-F23. The dimension of these functions is set according to Table 6, and the maximum number of iterations for both GSA and IGSA is set to 1000. The results are averaged over 30 different runs and the average best-so-far solutions and median of best solutions. Table 6 contains multimodal low-dimensional functions in which exploitation is more effective than exploration. So, IGSA work slightly better than GSA, except in F18, F19 and F20. The exploration ability of neighborhood crossover operator in IGSA does not let it exploit as well as GSA. According to the results, it is concluded that the GSA has the ability to explore and exploit, while IGSA has improved the exploration and exploitation of GSA for high-dimensional unimodal and multimodal optimization functions. The results of the average best-so-far solution over 30 runs are shown in Figs. 4 and table 6.

Table6: Minimization result of benchmark functions, t_{max}=1000.

Average Best-so-far	
IGSA	GSA
F14 3.2665 n=2	4.0796
F15 6.9793e-004 n=4	0.0042
F16 -1.0306 n=2	-1.0306
F17 0.4848 n=2	0.3979
F18 3.2015 n=2	3.0000
F19 -3.7966 n=3	-3.6630
F20 -2.5261 n=6	-2.0358
F21 -9.4916 n=4	-5.2251
F22 -10.1491 n=4	-7.3910
F23 -10.3273 n=4	-10.5364
Median Best-so-far	
IGSA	GSA
2.9826	3.0186
6.8627e-004	0.0035
-1.0311	-1.0306
0.4527	0.3979
3.1668	3.0000
-3.8142	-3.7651
-2.9007	-1.8714
-9.9794	-5.0552
-10.3823	-5.0877
-10.4523	-10.5364

The benchmark functions are taken form [17]. Tables1-3 is the benchmark functions used in the experimental study. In the tables, n is the dimension of function f_{opt} is the optimum value of the function, S is a subset of the R^n. The functions of table1 are unimodal and f_{opt} are zero, the functions of table2 are multimodal having many local minima, the minimum values are zero except for F_8 which has a f_{opt} of $-420\times n$. Table3 is multimodal functions have a few local minima, A detailed description of these functions can be found in the appendix of [17, 18].

5. Conclusion

GSA is a powerful global searcher, but it is not effective enough for more complicated problems. The overall goal of this paper was to increase the exploration and exploitation abilities of GSA, therefore the neighborhood crossover operator was applied to the GSA, The operator is used to enhance the gravitational search algorithm for the local search capacity, and scale factor line adjust the proper balance between the GSA and the neighborhood crossover operator to obtain good results. The experiment and simulation results show the IGSA is an effective optimization algorithm, avoiding premature convergence in cases where standard GSA failed.

Acknowledgments

Project supported by the National Natural Science Foundation of China (Grant No. 60774028) and Natural Science Foundation of Hebei Province, China (Grant No. F2010001318).

Reference

[1] B. S. Xun, X. G. Zhou, A genetic algorithm based on combination operators, Procedia Environmental Sciences 11(2011), 346–350.

[2] R. Gábor, E. Anikó, Genetic algorithms in computer aided design.Computer Aided Design, 35(2003), 709-726.

[3] H.H.K Timo, S. Jukka, Accelerating genetic algorithm computation in tree shaped parallel computer ,Journal of Systems Architecture, 42 (1996), 19-36.

[4] Y. H. Chang, Adopting co-evolution and constraint-satisfaction concept on genetic algorithms to solve supply chain network design problems, Expert Systems with Applications, 37(2010), 6919-6930.

[5] E. B. Donald, L. H. Christopher, A practical application of simulated annealing to clustering, Pattern Recognition, 25(1992), 401-412.

[6] A. Vasan, S. R. Komaragiri, Comparative analysis of simulated Annealing, simulated quenching and genetic algorithms for optimal reservoir operation. Applied Soft Computing, 9,(2009), 274-281.

[7] T. H. Wu, S. H. Chung, Hybrid simulated annealing algorithm with mutation operator to the cell formation problem with alternative process routings. Expert Systems with Applications, 36(2009), 3652-3661.

[8] Q. Zhang, An Improved artificial immune algorithm with a dynamic threshold, Journal of Bionic Engineering, 11 (2006), 93-97.

[9] A. Kalinlia, N. Karabogab, Artificial immune algorithm for IIR filter design, Engineering Applications of Artificial Intelligence 18 (2005) 919–929.

[10] M. N. Hossein, T. Nima, New robust and efficient ant colony algorithms: using new interpretation of local updating process, Expert Systems with Applications, 36(2009), 481-488.

[11] C. García-Martínez, O. Cordón, F. Herrera, A taxonomy and an empirical analysis of multiple objective ant colony optimization algorithms for the bi-criteria TSP, European Journal of Operational Research, Volume 180(2007), 116-148

[12] B. J. Zhao, S. Y. Li, , Ant colony optimization algorithm and its application to Neuro-Fuzzy controller design, Journal of Systems Engineering and Electronics, 18(2007), 603-610.

[13] T.Y. Chen, T.M. Chi, On the improvements of the particle swarm optimization Algorithm, Advances in Engineering Software, 41(2010), 229–239.

[14] Y. Jiang, T. Hu, C. Huang, X. Wu, An improved particle swarm optimization algorithm, Applied Mathematics and Computation, 193(2007), 231–239.

[15] P.C. Fourie, A.A. Groenwold, The particle swarm

optimization algorithm in size and shape optimization, Structural and Multidisciplinary Optimization, 23(2002), 259–267.

[16] G.. Baskar, M.R. Mohan, Contingency constrained economic load dispatch using improved particle swarm optimization for security enhancement, Electric Power Systems Research, 79(2009), 615–621.

[17] E. Rashedi, Gravitational search algorithm, Electrical Engineering Department, Shahid Bahonar University of Kerman, Iran (2007) (in Farsi).

[18] S. Sarafrazi, H. Nezamabadi-pour, S. Saryazdi, Disruption: A new operator in gravitational search algorithm, Scientia Iranica, 18 (2011), 539-548.

Probabilistic Latent Semantic Analysis for Unsupervised Word Sense Disambiguation

Gaurav S Tomar, Manmeet Singh, Shishir Rai, Atul Kumar, Ratna Sanyal, Sudip Sanyal

Indian Institute of Information Technology
Allahabad, Uttar Pradesh, India

Abstract

This paper presents an unsupervised approach for disambiguating between various senses of a word to select the most appropriate sense, based on the context in the text. We have defined a Probabilistic Latent Semantic Analysis (PLSA) based Word Sense Disambiguation (WSD) system in which sense tagged annotated data is not required for training and the system is language independent giving 83% and 74% accuracy for English and Hindi languages respectively. Also, through word sense disambiguation experiments, we have shown that byapplying Word net in this algorithm, performance of our system can be further enhanced.

Keywords: *Word Sense Disambiguation, Probabilistic Latent Semantic Analysis, Word net, Algorithm*

1. Introduction

In the field of natural language processing, "Word Sense Disambiguation (WSD) is defined as the problem of computationally determining which "sense" of a word is activated by use of the word in a particular context" [1]. Sense disambiguation is used in NLP applications like Document Summarization, Document Categorization Information Retrieval,Information extraction and Text mining, Lexicography, Machine Translation, Document Similarity measurement, Document Classification, Question Answering systems and Cross Language Information Retrieval [1]. The paper provides a new technique for performing WSD that governs the process of identifying which sense of a word is used in a sentence, when the word can have multiple meanings (polysemy) [34]. For example, word "bank" can have up to 10 senses as per Word Net (the semantic lexicon for English Language). Consider the following two sentences: "He went to the bank to deposit money" and "The temple is situated on the bank of Ganga". Here we have used the word "bank" in both the sentences but the sense in which the word "bank" comes is entirely different in the two sentences. A WSD system expects a sentence and it has to suggest the sense of each ambiguous word based on the context in which that word appears.

Various approaches have been proposed for WSD which can be categorized as Knowledge Based, Supervised, Semi supervised and Unsupervised. Knowledge based approaches depend on resources like dictionaries, ontologies and collocations. In this paper we present the induction of word sense using an approach based on word clustering which clusters semantically close words using a purely statistical method, Probabilistic Latent Semantic Analysis (PLSA) and second order co-occurrence which generates rich and informative clusters. PLSA, using tempered Expectation Maximization (EM), is then used to generate *'k'* clusters (containing semantically similar words), each representing a certain concept/topic. These clusters are further expanded by enriching them with more semantically related words like synonyms, hypernyms and homonyms using Word Net which is a lexical database. The correct sense of a polysemic word, present in the test corpus, is then deduced by calculating the similarity between the test corpus having the target polysemic word and clusters generated earlier. The cluster with highest similarity score is attributed to be the most appropriate cluster representing the sense of the polysemic word.

Unsupervised approaches are highly robust, portable and do not require resources like concept hierarchies, dictionaries and hand crafted knowledge resources. Classical unsupervised approach for Word Sense disambiguation, Lesk Algorithm, was based on the hypothesis that words in a given neighbourhood will tend to share a common theme [35]. It compares the set of words enclosed in the dictionary definition and examples of a polysemic word with the words present in its neighborhood. Unfortunately, Lesk's approach is perceptive to the exact wordings of the definitions in the dictionary. So, the existence / nonexistence of a certain word can drastically change the results. Further, the algorithm calculates overlaps only among the glosses and examples of the senses being measured [35]. This is a significant limitation of the dictionary that glosses tend to be fairly small and do not provide sufficient vocabulary (glossary) to communicate fine-grained sense distinctions [36]. That is why Lesk's approach gives less accuracy due to its dependence on dictionary definitions and examples to build the set of words with which it compares the set of words present in the neighborhood of the ambiguous word.

Given the above limitations of the Lesk approach, one can think of three possible directions for improving it. These are:

(a) Improve the glosses

(b) Improve the method used for measuring similarity / distance

(c) Improve the context of the ambiguous word whose correct sense has to be identified.

The unsupervised approach proposed here considers the first possibility i.e. it tries to find a better set of words with which it measures the similarity of the words appearing in the context of the given ambiguous word. This different mechanism is based on a purely statistical method of Probabilistic Latent Semantic Analysis explained later in this paper which gives better set of words (clusters) to decide sense of a word in a given text. Set of words (clusters) obtained from PLSA are more rich and informative than set of words obtained using Lesk's approach.

In the next section we shall give a brief review of the existing approaches. The third section contains the details of our approach. In particular, this section contains a brief overview of the PLSA and the way it is used for obtaining word clusters that can be used for performing WSD. The fourth section describes our results. Since the method adopted is fairly generic, we have tested it on two different languages, English and Hindi. The fifth section contains the concluding remarks and suggestions for further improvement.

2. Related Work

Most of the unsupervised methods for Word Sense Disambiguation are based on similarity methods and graph based methods. Graph based methods have two steps [16][17]. In the first step, a graph is constructed from the lexical knowledge based on possible hidden meaning representation of all possible compilations of the word whose sense is being disambiguated. Basically, graph nodes represent possible senses and edges of graph correspond to relations between senses. This graph structure is then used to find the value of each node in the graph [19] [20]. In this way the graph is used for finding the most appropriate sense of the word. In a graph based approach proposed by Veronis [3], firstly a co-occurrence graph is formed in which the nodes are words appearing in the paragraph of the text corpora in which target word exists. An edge between a pair of words is added to the graph if they reappear in same paragraph. Finally, a minimum spanning tree is used to disambiguate specific examples of the target word. A graph based method that uses the degree of centrality for WSD has also been explored [6].

Recently Navigli [19][20] proposed a graph based algorithm for large scale WSD which does not require sense annotated data for the training but has to investigate the graph structure in its whole depth. This method [14] aims to capture word sequence dependency information in the given corpora. Similarity based methods use clustering which can be further categorized in two types of clustering - word clustering and context clustering. Similarity based approach assigns a best suited sense to a word with the help of its surrounding words. In the similarity based algorithm, sense is calculated for each word individually whereas in graph based approaches, the sense of a word is found along with all its neighboring words with the help of dependencies across senses. An approach based on context clustering depends on the concept of word space [26]. Word space method derives a vector for each word in the corpora using a co-occurrence matrix. But, this word space has a large dimension and so we need to apply Singular Value Decomposition for reducing the dimensionality of the word space. Many unsupervised approaches depend on context clustering.

Recent years have popularized unsupervised methods based on word clustering like Probabilistic Latent Semantic Analysis [9] and Latent Dirichlet Allocation [2] etc. for information retrieval, document similarity [11], Keyword spotting [32] and document categorization [7]. The proposed approach in this paper also uses word clustering. Agglomerative clustering of words [22] considers each word as a single cluster initially and then proceeds by including similar words into the same cluster until a predefined threshold is reached. This approach has been successfully used in biomedical domain [25]. But, the above approach needs a large amount of unlabeled training data for the construction of context vectors. A LDA based approach has been used for discovering concepts, which has subsequently been used for WSD and showed improvement over conventional methods [24]. However, the authors have considered domain specific documents for LDA as most words tend to have same sense in a specific domain. Similar idea was also explored for medical documents [28].

Clustering By Committee (CBC) takes a word type as input and finds clusters of words that represents each sense of the word [13]. Hyperspace Analogue to Language (HAL) is based on word by word co-occurrence statistics [4]. HAL does not include large units of context and it captures co-occurrence data for words by considering a window of 10 words. There is a fully unsupervised method that is to cope with WSD [12] using a mono-lingual corpus but uses a different clustering approach. Traupman and Wilensky performed experiments [29] to improve discrimination accuracy of Yarowsky's Classifier using an iterative approach to re-train the classifier, using part of speech knowledge and training using weighing of senses distributed to dictionary order. Yet another method [5] for unsupervised WSD combines multiple information sources, including semantic relations, large unlabeled corpora and cross lingual distributional statistics. There exists an unsupervised learning method using Expectation Maximization algorithm [21] [27] which perform an optimal number of iterations of CV EM and CV EM2. In this unsupervised approach only a dictionary and an un-annotated text are required as input. This proposed method overcomes problem of brittleness present in many existing methods.

3. Proposed Method

3.1 Probabilistic Latent Semantic Analysis

PLSA depends on latent class model (or aspect model) [8][9] which consists of latent class variables. These latent class variables represent aspects/topics (senses in our case) and the model uses Expectation Maximization (EM) Algorithm for maximum likelihood estimation. Tempered EM can be used for better results. Probabilistic Latent Semantic Analysis relates the latent variables $\{z_1, z_2, \ldots z_k\}$ with the documents $\{d_1, d_2 \ldots d_i\}$ in addition to the terms $\{w_1, w_2, \ldots w_j\}$. There are three variables related with this approach as defined below:

- Select a document d_i with probability $P(d_i)$.
- Pick a latent class or concept z_k with probability $P(z_k \mid d_i)$.
- Generate a word w_j with probability $P(w_j \mid z_k)$.

The distribution P(w,d) can be written as

$$P(d_i, w_j) = P(d_i)\, P(w_j \mid d_i) \quad (1)$$

Where

$$P(w_j \mid d_i) = \sum (P(w_j \mid z_k)\, P(z_k \mid d_i)) \quad (2)$$

Or, using Bayes' rule, we get

$$P(d_i, w_j) = \sum_k (P(z_k)\, P(w_j \mid z_k)\, P(d_i \mid z_k)) \quad (3)$$

The above distributions can be found by using Expectation Maximization Algorithm which consists of two steps namely the expectation step and the maximization step. In the expectation step the posterior probabilities for latent parameters are calculated based on the current estimates of the parameters. For the Expectation:

$$P(z_k \mid d_i, w_j) = \frac{P(z_k)(d_i \mid z_k)P(w_j \mid z_k)}{\sum_{r=1}^{K} P(z_r)(d_i \mid z_r)\, P(w_j \mid z_r)} \quad (4)$$

In the maximization step we update the parameters based on the maximized log likelihood probability, based on the values calculated in the expectation step. We get the following set of equations:

$$P(w_j \mid z_k) = \frac{\sum_{i=1}^{N} n(d_j, w_j) P(z_k \mid d_i, w_j)}{\sum_{s=1}^{M} \sum_{i=1}^{N} n(d_j, w_s) P(z_k \mid d_i, w_s)} \quad (5)$$

$$P(d_i \mid z_k) = \frac{\sum_{j=0}^{M} P(d_i \mid w_j) P(z_k \mid d_i, w_j)}{\sum_{q=1}^{N} \sum_{j=1}^{M} n(d_q, w_j) P(z_k \mid d_q, w_j)} \quad (6)$$

and

$$P(z_k) = \frac{1}{\sum_{i=1}^{N} \sum_{j=1}^{M} n(d_i \mid w_j)} \sum_{i=1}^{N} \sum_{j=1}^{M} P(d_i \mid w_j)\, P(z_k \mid d_i, w_j) \quad (7)$$

In the above expressions, $P(w_j \mid z_k)$ represents the probability of observing a particular term or word in a given concept. $P(z_k \mid d_i)$represents the probability of a topic in a given document, $P(z_k)$ is the probability of a topic and $n(d_i, w_j)$ is number of times a term or word occurs in a particular document d. Now we iterate through these steps again and again till convergence. We have used the parameter $P(w_j \mid z_k)$ to create word cluster in PLSA because they essentially represent the probability of finding the word w_j in the cluster z_k. If the number of concepts is K, then we will have K clusters. The set of words belonging to a cluster will be those words that have a probability above a threshold value in that cluster. In this approach a word may belong to more than one cluster.

3.2The WSD System

In this paper we propose a WSD system based on word clusters obtained using PLSA. The system consists of two phases: the training phase and the testing phase. The training phase creates the word clusters that will be used for disambiguating and the testing phase performs the actual disambiguation. These phases are described below.

3.2.1The WSD System

In the training phase, the words that relate to a particular context are grouped together and each such grouping (cluster) represents one sense. In Figure1, we present the architecture of the training phase of the proposed system.

Step 1- Removal of stop words from training data:

Stop words are the high frequency words that have very low semantic value. These words comprise 30% of the whole training data and hence must be removed. Usually one maintains a list of such words. Any word in this list is not passed to the subsequent steps. For English language, {is, am, on, off, the, a, an, about} are some of the stop words.

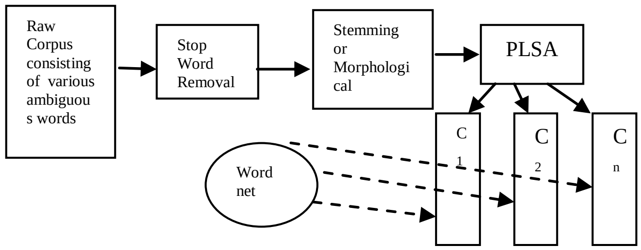


Fig. 1 Training Phase Architecture for the WSD System.

Step 2- Reduction of inflectional and derivational variants to their root form (Stemming):

To reduce the size of the training corpus to be processed further, we use a statistical stemmer for reducing inflectional or derived words to a reduced form that may or may not be the morphological root of the words. It is not necessary that the stemmed words should give the morphological root of the word. It is sufficient that similar words map to the same stem. For instance, the words "call", "caller", "calls" map to same stem "call" after stemming.

Step 3- Word clustering using PLSA**:**

PLSA helps in clustering similar words or the words related to a particular topic together by giving $P(w_j|z_k)$ as output for each term 'w' and topic 'z'. We take the list of probabilities $P(w_j|z_k)$ for each topic 'z_k' and sort the list so that among $(w_1, w_2,\ldots w_j,..)$**,** only the words which represent a topic/sense strongly come together. After sorting, for each topic words having low $P(w_j|z_k)$ value are removed as they do not represent the topic so strongly and perhaps belong to some other topic/sense. This step finally results in clusters/topics, each having words closely related to each other.

Step 4 - Expansion of clusters using a lexical database:

Although the clusters obtained from the previous step contain rich and informative words which strongly represent the topics, but the number of words in each cluster is quite small. Hence, disambiguation done with this set may not give best possible results. The list of words in each cluster, which we got as output of PLSA in previous step, are further expanded by including more semantically related words with the help of ontology in Word Net which is a lexical database consisting of the semantic networks of words, their synsets, homonyms, hypernyms and hyponyms. After the expansion of word list in each cluster, we have a significant number of words which define a particular sense in that cluster.

3.2.2The WSD System

After getting clusters of similar words representing topics (senses) from training phase, an ambiguous word present in test corpus can be classified into one of these clusters obtained from training corpus by computing the similarity score of test corpus containing the target ambiguous word with each cluster. The cluster with which the highest similarity score is obtained is attributed to be the most appropriate sense of the queried ambiguous word. The idea here is that word lists (clusters obtained from training phase) and expanded with the help of WordNet as explained in step 4 of training phase consists of all the related words which make up a sense. Overlap of these extended lists with test corpus can be calculated and then the cluster/word list having maximum overlap with the test corpus gives the best representation of the sense for the ambiguous word in the test corpus. The proposed WSD system uses cosine coefficient as a similarity score measure. Our experiments show that using other similarity score measures like Dice Coefficient, Jaccard coefficient etc. give similar results.

4. Result and Analysis

Presently we have evaluated our algorithm on two languages, English and Hindi. We have collected various categories of ambiguous words. In the first category, some words have two senses and some even have more than two senses. In the second category, the part of speech of the ambiguous word may be different for different senses. In the third category, the origin of different senses of the ambiguous word is same. For example, the word "bank" has originated from the word "safe". Thus, a financial bank means it is safe for money and a river bank means it saves from flowing water. We created a corpus of more than 150 documents containing instances of all the three categories. Within this corpus we earmarked 50 words which appeared in multiple senses. We also evaluated our WSD algorithm performance with, as well as without, using WordNet to expand the clusters. Our results for some words are summarized in Table 1. The first column shows the ambiguous word and its sense. The second column shows the number of occurrences of the ambiguous word in the test corpus. The third column gives the number of correct matches and the last column shows the accuracy of correct sense identification for a given ambiguous word. In Table 1 we have presented the results when the word clusters are expanded using WordNet.

An examination of Table1 shows that we get an average accuracy of 83.17%. As discussed earlier, we can put ambiguous words in three categories. We will now analyze the accuracy of disambiguation for each category separately. The first ambiguous word, "well" have senses belonging to different part of speech. The first sense of "well" is a noun while the second sense of "well" is an adverb. Our proposed approach has given similar accuracy (88%) for both types. We now consider the word "cricket" which can mean a sport or an insect. In this case both senses have the same part of speech - noun. In this case our proposed approach has given 100% accuracy.

Table 1: Accuracy of disambiguating several English ambiguous words

Ambiguous Word and Sense	**Number of Occurrencesin Corpus**	**Number of Correct matches**	**Accuracy percentage**
Well (Deep Hole)	16	14	88
Well (Good)	8	7	88
Fair (Gathering)	42	34	81
Fair (Good)	13	13	100
Fair (Travelling)	40	40	100
Book (Booking)	8	3	38
Book (Notebook)	15	15	100
………………….	……………..	…………	……….
………………….	……………..	…………	………..
Cricket (Game)	13	13	100
Cricket (Insect)	9	9	100

Average Accuracy = 83.17%

Now let us consider the word “fair”. In this corpus, word “fair” has three possible senses. Two of the senses are nouns while the third sense is an adverb. The results show that we get different accuracy for different senses. For example, in the training corpus, word “fair” as an adverb is present with word “fair” as a noun in the sense of gathering. Thus, the clusters corresponding to these two senses have considerable overlap and hence lead to a drop in accuracy. A similar effect occurs in the case of the ambiguous word “book”. One sense of “book” is a Verb (as in – booking a ticket) while the second one is a noun (as in – to read a book). We again see a drop in accuracy. This happens because in the training corpus we have sentences like “someone has left this book on this seat” and “this seat has been booked by someone”. As we can observe, these sentences have the same collocation words leading to an overlap in the corresponding word clusters. We also observed that WordNet has an important role in increasing the accuracy. On an average there is an increase of 2 to 5% in accuracy when we expand the clusters using WordNet.

Analysis of proposed approach shows that unsupervised approach discussed here is quite superior compared to the Lesk algorithm. Moreover, the technique is language independent since it is a purely statistical approach. It merely requires a suitable, untagged corpus to build the clusters. Availability of a Word Net type of resource would enhance the accuracy but it is not mandatory. We are presenting some examples of the results obtained with our WSD tool.

Example 1

Word: Date
Possible senses
Sense 1: Fruit (Noun)
Related cluster snapshot: date, fruit, desert, eat, market
Sense 2: related to Calendar (Noun)
Related cluster snapshot: date, month, year, birthday, current, time, number
Sense 2: related to Love (verb)
Related cluster snapshot: date, love, girl, escort, travel, single, email, internet, personal

Example 2

Word: Kite

Possible senses
Sense 1: Bird
Related cluster snapshot: kite, nest, bird, snail, raptor, wings, water, tree
Sense 2: Paper Toy
Related cluster snapshot: kite, fly, festival, people, art, competition, dor, cut, sky, makar, rajasthan, sakranti

The results obtained for Hindi are summarized in Table 2.

Example 3

Word: फल

Possible senses
Sense 1: Result
Related cluster snapshot: सफलता [success],द्वीप [island],फल [result], परिणाम [result],असफलता [failure],प्रतिफल [failure]
Sense 2: Fruit
Related cluster snapshot: आम [mango],फल [fruit],भारत [India], खेल [game], मोटर [automobile]

Example 4

Word: सोना

Possible senses
Sense 1: Gold

Related cluster snapshot: सोना [gold], पैदल [foot], धातु [metal], रंग [color], सिक्का [coin]
Sense 2: Sleep
Related cluster snapshot: सोना [sleep],नीद [sleep], सो [to sleep], रात [night], दौरान [during]

Table 2: Accuracy of disambiguating several Hindi ambiguous words

Ambiguous Word and Sense	**Number of Occurrencesin Corpus**	**Number of Correct matches**	**Accuracy percentage**
फल (Result)	8	6	75
फल (Fruit)	8	7	87
सोना (Gold)	4	5	80
सोना (Sleep)	4	7	57
कलम (Pen)	4	5	80
कलम(Kill)	4	6	66
आम (fruit)	34	30	88
आम (common)	5	4	80
…………………………	…………………	………………	………………
…………………………	…………………	………………	………………
लाल(red color)	24	21	88

Average Accuracy = 74.12%

Now when a test corpus having ambiguous word फल or सोना is given as input to test the system, our experiments showed that the cluster representing the related sense had the maximum overlap with the test corpus and hence was returned as output**.** For instance, if सोना is used in sense of 'sleep', it tends to have words like रात [night], सो [to sleep] etc. as its neighbouring words and hence, will have maximum overlap with cluster having these words thereby helping to identify correct sense.
In the above samples we see that the related cluster snapshots contain many words that are very relevant. Thus, we can say that PLSA, with the enhancements proposed in earlier sections, leads to very good clustering of words and thereby increasing the accuracy of the disambiguation process compared to the Lesk algorithm. Analysis of proposed approach also shows that unsupervised approach discussed here is language independent and contrary to standard supervised approach do not utilize manually tagged data in any way.

4. Conclusion and Future work

In this paper we employed word clustering based on Probabilistic Latent Semantic Analysis for developing an unsupervised and a relatively generic WSD algorithm. Our experiment shows that the proposed approach is language independent and obtained state of art performance on well managed evaluation data sets giving 83% and 74% accuracy for English and Hindi languages respectively. Adopting WordNet enriched clusters further improve the accuracy in the range of 2 to 5%. This shows that cluster based WSD algorithms perform better with more sense inventories as we get more clusters and more words in them. WordNet like reference lexicon exist for several languages. It is really an interesting future direction to establish to see how well our WSD algorithm performs with other such needful resources. Performance of the system proposed here can linearly increase according to size of training data. Our results focused basically on cluster measures and their improvement using word net sense inventories. More research problems need to be assessed to see whether our results can be extended to other NLP problems, other than WSD.

References

[1] E. Agirre, and P. Edmonds, Word Sense Disambiguation: Algorithms and Applications Springer, New York, NY, 2006.
[2] D. Blei, M. Jordan, and I. Michael, "Latent Dirichlet allocation". Journal of Machine Learning Research Vol.**3**: pp. 993-1022, 2003.
[3] R. Bruce, and J. Wiebe, "Word Sense Disambiguation using Decomposable model," In Proceedings of the 32nd Annual meeting of the association for Computational Linguistics (ACL, Las Cruces, NM).pp.139-145, 1994.
[4] B. Curt, and K. Lund, "Modeling parsing constraints with high dimensional context space", Language and Cognitive Processes, Vol.12 No.(2-3) pp. 177-210 ,1997.
[5] R. Florian, and R. Wincentowski, "Unsupervised Italian Word Sense Disambiguation using Wordnets and Unlabeled Corpora," In Proceedings of SigLEX'02, pp. 67-73,2002.
[6] E. Hessami, F. Mahmoudi, and A.H. Jadidinezad, "Unsupervised Graph based Word Sense Disambiguation using lexical relation of Wordnet", In International Journal of Computer Science Issues, Vol.8, Issue 6, No. 3, pp. 225-230, 2011.
[7] T. Hoffmann, "Tutorial on Matrix Decomposition Technique in Machine Learning and Information Retrieval" In: 5th Max-Planck Advanced Course on the foundation of computer science, Sept 6-10, 2004 Soarbriicken.
[8] T. Hofmann, "Probabilistic Latent Semantic Indexing" In Proceedings of the 22nd annual international ACM SIGIR conference on Research and Development in information Retrieval. Berkeley, California, United States: ACM, 1999, pp. 50 – 57, 1999.
[9] T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis", In Journal of Machine Learning, Vol. 42, Numbers 1 – 2, pp. 177 – 196, 2001.

[10]V.N. Huynh, J.T. Ngugen, and C. Annie, "Adaptovely Entropy based Weighting classifiers in Combination using Dempster – Shaper Theory for Word Sense Disambiguation," In Journal of Computer Speech and Language Elsevier Volume 24 pp. 461 -473, 2010.

[11]A. Kumar, and S. Sanyal, "Effect of Pronoun Resolution on Document Similarity, in International Journal of Computer Applications" Volume 1, No. 16, pp. 60-64, 2010.

[12]D. Levinson, "Corpus-Based Method for Unsupervised Word Sense Disambiguation", www.stanford.edu/ dmitryle/acai99w1.ps, 1999

[13]D. Lin, and P. Pantel, "Concept discovery from text", In Proceedings of the 19th International conference on Computational Linguistics (COLING) pp.577-583, March 2002 Taipei, Taiwan.

[14]S. Lin, and K. Verspoor, "A Semantics-Enhanced Language Model for Unsupervised Word Sense Disambiguation" In CICLing 2008, LNCS 4919, pp. 287–298, 2008.

[15]R. Mihalcea, "Bootstrapping large sense tagged corpora," In Proceedings of the 3rd International Conference on Language Resources and Evaluations (LREC, Las Palmas, Spain) 2002a .

[16] R. Mihalcea, "Unsupervised Large-Vocabulary Word Sense Disambiguation with Graph-Based Algorithms for Sequence Data Labeling," Proc. Human Language Technology and Empirical Methods in Natural Language Processing, pp. 411-418, 2005.

[17]R. Navigli, and P. Velardi, "Structural Semantic Interconnections: A Knowledge- Based Approach to Word Sense Disambiguation," IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 27, no. 7, pp. 1075-1088, July 2005.

[18]R. Navigli, and M. Lapata, "An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation" IEEE Transactions on Pattern Analysis and Machine Intelligence Vol 32, pp. 678-692, 2010.

[19]R. Navigli, and M. Lapata, , "Graph Connectivity Measures for Unsupervised Word Sense Disambiguation," In Proceedings of 20th International Joint Conference Artificial Intelligence, pp. 1683-1688, 2007.

[20]R. Navigli, "Word Sense Disambiguation: A Survey" In Journal Computing Surveys, Vol. 41, N0.2, pp. 1-69, 2009.

[21]K. Nigam, A. McCallum, S. Thrun, and T. Mitchell, "Text Classification from Labeled and Unlabeled Documents using EM" In Machine Learning, 39(2/3): pp.103–134, 2000.

[22]T. Pedersen, and R. Bruce, "Distinguishing word senses in untagged text" In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP, Providence, RI) pp.197 – 207, 1997.

[23]T. Pedersen, "Learning Probabilistic Models for Word Sense Disambiguation" Ph.D. dissertation Southern Methodist University, Dallas, TX , 1998.

[24]J. Preiss, and M. Stevenson, "Unsupervised Domain tuning to improve Word Sense Disambiguation," In Proceedings of NAACL-HLT 2013, pages 680-684 Atlanta, Georgia, 9-14 June 2013.

[24] J. Preiss, and M.Stevenson, "Unsupervised Domain tuning to improve Word Sense Disambiguation," In Proceedings of NAACL-HLT 2013, pages 680-684 Atlanta, Georgia, 9-14 June 2013.

[25]G. Savova, T. Pedersen, A. Purandare, and Kulkarni, "A. Resolving ambiguities in biomedical text with unsupervised clustering approaches" Res. rep. UMSI 2005/80. University of Minnesota Supercomputing Institute, Minneapolis, MN, pp/ 494-503, 2005.

[26]H. Schutze, "Dimensions of meaning" In Proceedings of the 1992 ACM/IEEE Conference on Supercomputing. IEEE Computer Society Press, Los Alamitos, CA. pp.787-796 1992.

[27]H. Shinnou, and M. Sasaki, "Unsupervised learning of word sense disambiguation rules by estimating an optimum iteration number in the EM algorithm" In Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003.

[28]M. Stevenson, E. Agirre, and A. Soroa, "Exploiting domain information for Word Sense Disambiguation of Medical documents", In Journal of the American Medical Informatics Association, Mar-Apr Vol. 19 Issue 2 pp. 235-240, 2012.

[29]J. Traupman, and R. Wilesky, "Experiments in Improving Unsupervised Word Sense Disambiguation" In Report in UCB /CSD -03-1227 Berkley California, pp. 1-39, 2003.

[30]G. Tsatsaronis, M. Vazirgiannis, and I. Androoutsopoulos, "Word Sense Disambiguation with Spreading Activation Networks from Thesauri," In Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI) pp. 1725-1730, 2005.

[31]J. Veronis, "Hyperlex: Lexical cartography for information retrieval" In Journal Computer Speech Language Vol. 18, No. 3, pp. 223–252, 2004.

[32]N. Yadav, Y. Gupta, M. Kumar, and R. Sanyal, "Semantic Document Classification and Keyword Spotting in Digital Repositories" In The International ACM Conference on Management of Emergent Digital EcoSystems (MEDES 09), Lyon/France, pp. 157-161, 2009.

[33]D. Yaroswsky, "Unsupervised Word Senses Disambiguation Rivaling Supervised Methods," In Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (Cambridge, MA) pp. 189-196, 1995.

[34]M. Breithaupt, "A fuzzy Word Sense Disambiguation Algorithm using Wordnet" M S Thesis, Georgia State University GA US, 2010.

[35]M. Lesk, "Automatic Sense Disambiguation using Machine Readable Dictionaries: How to Tell a Pine Cone from An Ice Cream Cone" In Proceedings of the 5th annual international conference on Systems documentation (SIGDOC '86) New York, NY, USA pp 24-26, 1986.

[36]A. Kilgarriff, and J. Rosenzweig, "English SENSEVAL: Report and Results" In Proceedings of the 2nd International Conference on Language Resources and Evaluation, LREC, Athens, Greece, 2000.

The Academic use of Social Networks Among University Students in Jordan

Atika Al Doghmi[1], Hasan Al-Shalabi [2], Jwaifell Mustafa Odeh[3], Swidan Andraws[4] Arafat Awajan[5] and Adnan I. Alrabea[6]

[1] Faculty of Information Technology, Al-Hussein Bin Talal University, Jordan

[2] Faculty of Engineering, Al-Hussein Bin Talal University, Jordan

[3] Department of Curriculum and Teaching, Al-Hussein Bin Talal University, Jordan

[4] Computer engineering department, Faculty of Engineering and Technology, University of Jordan, Jordan

[5] Computer Science Department, Princess Sumaya University for Technology, Jordan

[6] Faculty of Information Technology, Al Balqa Applied University, Jordan

Abstract

Most of the Arab world youths are using Social Networks for many different uses, while academic uses are still not in a formal perspective. This study aimed at exploring the academic uses of Social Networks among the students of several Jordanian Universities. The sample was analyzed according to university, faculty, gender, and year level. The study took place in four Jordanian universities in the first semester / 2012, by exploring (727) of those students through a questionnaire to collect data about SNS uses. Results of T-test, ANOVA, and Scheffe Post Hoc Tests, revealed a great deal of uses of Social Networks in three domains:1) Academic, 2) Intensity, and 3) Group uses, but not in a formal perspective. The researchers' recommendations are to make better use of those Social Networks by integrating them in universities' learning management systems.

Keywords: *Social Networks Sites (SNSs), Higher Education, Academic Use, academic relations, University students, Jordanian University.*

1. Introduction

Over the past decade, the use of social network sites (SNSs) has been increasing. SNSs is considered an important research field for scholars who area interested in online technologies and the social and cultural impact on them, it is shown in recent scholar work in the area (Andraws et al., 2013; Boyd & Ellison, 2007; Donath, 2007; Ellison, Steinfield & Lampe, 2007; Golder, Wilkinson & Huberman, 2007; Valkenburg, Peter & Schouten, 2006, Lampe, Ellison & Steinfield, 2007). SNS is considered as the "web-based services" which allow individuals to (1) create a public or semi-public profile within a certain bounded system, (2) connect with other users with whom they share a connection, and (3) view and visit their list of connections and other connections made by others within the system" (Boyd & Ellison, 2007). In 1997 first social network site was launched and now there are hundreds of SNSs worldwide, supporting many kinds of practices, users and interests (Boyd & Ellison, 2007).

Facebook is one of the biggest social network sites among the Jordanian students. Referring to statistical date by the Arab Social Media Report (ASMR), there were 2.244 million Facebook users in Jordan by till October 2012, and more than 217,000 new accounts were created between January and June 2012. Jordan has been ranked the 7th among the Arab states in number of Twitter users and 2nd in volume of Facebook users across the region. 75 percent of Facebook users in Jordan are between the age 15 and 29, 56 per cent are male, and 58 percent prefer to use Arabic language in their activities son these sites.

Facebook became also very popular among undergraduates, with usage rates up to 90% at most American campuses (Stutzman, 2006; Lampe, Ellison & Steinfield, 2006. Many recent researches has been inducted on various aspects of SNSs usage, such as the use of Facebook in academic settings and the demographic predictors of Facebook use (Hewitt & Forte, 2006). Many of the research aspects focus on the outcomes of Facebook

usage in terms effects among young adults and the relationships between peers which is significant both for generating o benefits offline, commonly known as social capital.
Social capital is used to describe the benefits person receives from one's relationships with other people (Nan, 1999). Moreover, (Ellison, Steinfield & Lampe, 2007) as for the psychosocial development it is found evidence that self-esteem may work as a moderator of the relationship between social network site use and social capital. In other words the young people with low self-esteem were found to benefit more from their use of Facebook than those with higher self-esteem. However, it was not possible for Ellison, Steinfield & Lampe (2007) to establish any time order to the relationships among Facebook use, self-esteem, and social capital with data at only one point in time.

This paper presents the main findings of a study conducted in Jordanian universities to evaluate the use of social networking sites by universities' students. It investigates the use and the academic impacts of social networking sites among the students. It also tries to depict the relationship between the intensity of use and the field of the higher education of the users. A particular focus the use of social networking sites and academic impact and extent to which user's self-esteem moderates relationship. We were motivated in this study by the continuous increasing of the number of students joining the SNS. For example the percentage of students who joined the social sites on the campus of the university in Jordan is estimated to be between 85% and 95% (ASMR).

The changing concepts of learning and communication among the academic community requires more researchers to overcome and face-to-face interaction to examine the Internet in order to understand the relationships and academic students to develop more of the blocks on the SNS for the exchange of educational resources, townships relationships and support. As such, Riva and Galimberti (1998) confirms this, where the residence halls, student unions and classrooms once flourished as the collective community centers, virtual reality has diminished the need for these geographic locations to form a community. It is suggested that Facebook provides "easy to use" democracy and environment for student participation in communities of practice in ways that traditional learning spaces such as learning management systems can least afford to do.

2. Background of the study

Nowadays, most universities in the world are using social media in order to advertise and pass on specific information, and maintain a high level of communication with their students whereas their academic use still very limited. The power of social media as a common means to build networks and replace traditional community building has rapidly recognized by universities. They have moved from just having a website to be engaged in the land of new social media platform. Their use is wide to cover a lot of tasks such as maintaining connections to alumni, hunting for prospective students, and promoting the advances of their research departments.

There are many benefits educational and personal for the use of social networking sites for both teachers and students. Social networking sites like Facebook, Multiply, Myspace, and Twitter are currently used by people of all ages around the world to connect with each other. Some of them are using these new media for educational purposes.

The interest of academic researchers has been increased by the rising popularity of social networking sites. Researchers in developed countries confirm that social networking sites have done lots of impact on the students learning in a positive different ways. Now social networks focuses on retaining and developing relationships among teachers and their students. They are used for professional development, as well as learning, and sharing content or information. In the survey conducted by the National School Boards Association (NSBA), it was found that 50% of teenagers would use social media to talk to their friends regarding school homework online or through text messaging. 60% of the surveyed teenagers show that they discuss topics that are linked to education through their account.

Various studies are conducted by Boyd & Ellison (2007) in which researchers examined and studied a the impact of the online professor existence on the student-professor relationships, the correlation and connection between libraries and social networking sites, and also the connection between students' social networking and the technical ability and notice their level of skills. Students can get educational benefits like going beyond than the classroom to follow interests, as well as creativeness and self expression, contacting varied viewpoints, and online networking to make decisions correlated to higher education such as which college to focus on and or which programs to chase, as per Greenhow & Robelia (2009).

Students replied to a question about what they study from social networking site, they listed technology skills at the top, then creativity, also being open to new or different views, and communication skills. All the students did not know that internet and networking would provide lots of academic and professional networking opportunities.

Bookmarks of the desired and most used website can be given by teachers and among students for lessons and or problems that are for practice by using social networking and social bookmarking. Also posting in all kinds of blogs or other forms of writing students can also be given out assignment using social networking. Teachers are rely now more and more on social networking sites to help their students and fellow teachers in receiving important information details for schoolwork or other events going on.

In general the social networking sites focuses intensively on the field of Learning, the social network sites also handle big problems raised by educators in the recent period. Like the problem of lacking of humanitarian aspect which is described as teaching lacks spirit. However, Social networks have been helping to resolve some of these problems, through adding interactive way with humans, by making participation of the human element in the educational process something important. Which can lead to an increase of attraction people toward e-learning and some of researcher in the field of social sciences carrying out studies to examine this phenomenon and to clarify the reason behind the attraction educated about social network sites such as Ractham & Firpo (2011) and Jiang & Tang (2010).

Ractham & Firpo (2011) has identified and studied the possibility of using social networking technology in a graduate-level introductory MIS course through Facebook in order to enhance learning and boost that up. The main finding was that Social networking technologies such as Facebook is able to allow members participating in a learning environment where the learning process can occur interchangeably from both inside and outside of the classroom. They underlined the benefits of a successful use of social networking in education for both students and instructors. The benefits for students are some degree of informal learning through informal communication, in addition to support for collaboration, as well as feedback on thoughts and collaboration that is independent of space and time. For instructors, it was found out that gaining feedback from students and constant communication with them and producing effective instructional technology for their customers as the most significant benefits. As a result, networking technology could be used effectively to promote a culture of learning, for both students and teachers. Furthermore, in the future, there is a significant potential to expand teaching and learning out of classroom. As the social-networking sites will have the best potential.

Jiang & Tang (2010) outlined the way to use the social networking to support education for building pedagogical model through using social networking and that study tried to find out the balance between developing individual creativity along with group activity of collaboration. Based on a traditional instructional model, this model was created and gained the characteristic of design education. The social networking sites have become one of the main means of communication between students and teachers. Nevertheless, this model has been predicted to enhance learning efficiency.

Looking at the previous studies, we can conclude that the social-networking sites will have a major role in improving and developing education on the level of the students as well as teachers.

3. Problem statement of the study

The Arab world, especially the youth are very much fond of social networks, such as Facebook, twitter .and the Arab spring is a clear example for that. Although, university students are using Social Networks in different ways, so the academic may had to do a part of this use. Therefore, the current study focused on surveying Social Networks uses among the students of Jordanian universities.

4. Objective and questions of the study

The primary objective of this study is to describe the academic participation, intensity and group uses of social networks among the students of Jordanian Universities beside the uses differences according university, faculty, gender and year level. To achieve this goal, the two research questions were made:

Question 1: What are the academic participation, intensity and group uses of social networks among the students of Jordanian Universities?

Question 2: What are the academic, intensity and group uses of social networks according to University, Faculty, Gender and Year Level among the students of Jordanian Universities?

5. Method

By the nature of the subject, the data of which was collected by a questionnaire, determined the descriptive quantitative research design for this study. Hopkins (2008) considered that quantitative research is quantifying the relations between variables.

5.1 Sample of the study

The sample of the study consisted of (727) students drawn randomly out of (4) Jordanian universities, resembling the study variables: 1) university, 2) faculty, 3) gender, and 4) year level, as shown below:

Table 1: The breakdown of sample:

		Frequency	*Percent*
University	University of Jordan	280	38.5
	Hussein	155	21.3
	Balqa	135	18.6
	Sumaya	157	21.6
Total		727	100
Faculty	Science	531	73
	Arts	196	27
Total		727	100
Gender	Male	319	43.9
	Female	408	56.1
Total		727	100
Level	First Year	163	22.4
	Second Year	154	21.2
	Third Year	177	24.3
	Fourth Year	233	32.0
Total		727	100

It was hard for the researchers to collect data about the whole population, so the research results are limited to the sample as participants of this study.

5.2 Measurement tools

The data collected though out a questionnaire prepared to have evidence about each: Academic, Intensity, Group Use, and overall uses. The questionnaire items were collected through literature reviewing and social networks nature.
The validity of the questionnaire where confirmed through (4) referees to judge the items with the relation of its domains. The questionnaire's items have been recoded as shown below:

Table 2: Questionnaire Scale:

Scale	*strongly agree*	*agree*	*undecide d*	*Disagree*	*strongly disagree*
Rate	5	4	3	2	1

The score 3 considered as a crucial point to determine the students' agreements at the questionnaire scale
The reliability calculated by repeating the questionnaire to the same (30) participants drown randomly out of the population but not included in the sample of the study, where Pearson Coefficients were calculated as shown below:

Table 3: Pearson Correlation:

Domain	*Correlation*
Academic Use	.980**
Intensity of Social Networks Use	.992**
Group Use	.972**
Overall	.990**

All the correlations were statistically significant at the level $\alpha \leq .001$

6. Results and discussions

To achieve the study objective, the questionnaire distributed to the four university students, then data where analyzed by using SPSS to calculate the means and standard deviations as show below:

Table 3: Means and standards deviations of participant's responses:

Domain	*Item*	*Mean*	*Std*
Academic Use	Communicating with my instructor at the university	3.158 2	1.1741 1
	A requirement of the course I took	3.155 4	1.2034 4
	I would add my instructor to my group if he had an account at the social networks	3.576 3	1.2034 8
	I get benefits of the published sheets and documents about my course at the social networks	3.609 4	1.1226 8
	I prefer that every course has its	3.643	1.1126

	own webpage at a social network site	7	2
	I prefer having my own portfolio at learning management system at my university	3.5227	1.11425
	My achievement became better after communicating with my friends at the university, through the social networks	3.2779	1.20939
	Total	**3.4205**	**0.76939**
Intensity of Social Networks Use	Social Networks are part of my everyday activity	3.7964	1.05488
	I am proud to tell people I am on Social Networks	2.8735	1.13203
	I would be sorry if Social Networks Blocked at the campus	3.5103	1.29402
	Looking at or posting photos	3.0880	1.12715
	Entertainment	3.6039	1.08198
	Finding out about or planning events	3.6052	1.02295
	Total	3.4129	0.71890
Group Use	I feel I am part of the Social Networks community at the campus	3.4897	1.08944
	Communicating with friends on campus	3.7070	1.05405
	Communicating with friends not on campus	3.9959	0.97841
	Communicating with friends seen rarely	4.0454	0.98158
	Sending or receiving messages	3.5282	1.05312
	Reading comments or discussion board	3.4017	1.04494
	Total	3.6946	0.69641
Overall		**3.5047**	**0.60528**

Out of the table above:
1. The highest item where in the domain of Group Use: Communicating with friends seen rarely (4.0454), while the lowest item where in the Intensity of Social Networks Use: I am proud to tell people I am on Social Networks (2.8735).
2. the most approval item among the students at the Academic Domain, where: I get benefits of the published sheets and documents about my course at the social networks (3.6094) while the Item: A requirement of the course I took, took the lowest improvement.
3. the most approval item among the students at the Intensity of Social Networks Use, where: Social Networks are part of my everyday activity (3.7964)
2. The Item: Reading comments or discussion board took the lowest approval.

6.1 Results of Question 1

What are the academic, intensity and group uses of social networks among the students of Jordanian Universities?

One sample T test used to determine the academic, intensity and group uses of social networks uses at the crucial point 3 among the students of Jordanian Universities, as shown in the table below:

Table 4: T-test results:

Uses	*Mean*	*Std*	*DF*	*T*	*Sig*
Academic	3.4205	0.76939	726	14.737	0.000
Intensity	3.4129	0.71890		15.486	0.000
Group	3.6946	0.69410		26.894	0.000
Total	3.5047	0.60528		22.481	0.000

All the means where above the crucial score 3 significantly, which shows that all students' responses ranged between agree and strongly agree, while Group uses where the highest mean as shown in the following chart:

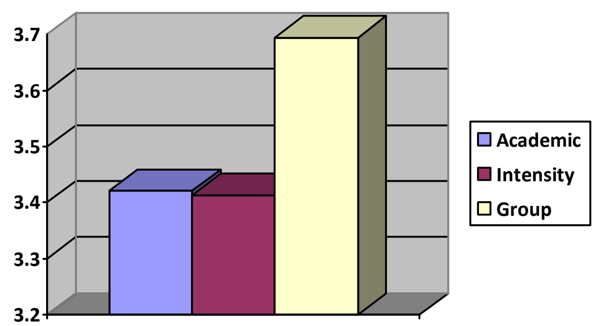


Figure 1: Students responses according to Social Networks Uses

Not surprisingly, most uses of Social Networks were the group use, which can be referred to nature of Arab youths; this nature refers to social communicating needs that Arabs always seeking, as a part of religion and family raise. One of the comments that students add to the questionnaire was: "it is our obligation to have one good build community"

6.2 Results of Question 2

What are the academic, intensity and group uses of social networks according to University, Faculty, Gender and Year Level among the students of Jordanian Universities?

The students' response according to each: Social Networks domains and the study variables (university, faculty, gender, and year level) where calculated as shown below:

Table 5: Participant's responses according to the study variables:

Variables		Academic		Intensity		Group		Total	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std
University	Jordan	3.5398	.70341	3.5012	.64988	3.8381	.60132	3.6218	.54965
	Hussein	3.6488	.65672	3.4312	.67280	3.6194	.70526	3.5708	.56467
	Balqa	3.2360	.79296	3.1568	.76695	3.4901	.75825	3.2912	.62034
	Sumaya	3.1410	.84778	3.4575	.79032	3.6890	.73938	3.4140	.66537
Total		3.4205	.76939	3.4129	.71890	3.6946	0.6941	3.5047	.60528
Faculty	Sciences	3.4498	.75932	3.4749	.69890	3.7546	.66935	3.5540	.59574
	Arts	3.3411	.79260	3.2449	.74677	3.5323	.74274	3.3711	.61225
Total		3.4205	.76939	3.4129	.71890	3.6946	0.69641	3.5047	.60528
Gender	Male	3.4048	.79492	3.3584	.73096	3.6839	.69675	3.4783	.61203
	Female	3.4328	.74959	3.4555	.70729	3.7030	.69688	3.5253	.59990
Total		3.4205	.76939	3.4129	.71890	3.6946	0.69641	3.5047	.60528
Year Level	Year1	3.5004	.74491	3.4622	.69668	3.6656	.66137	3.5405	.56274
	Year2	3.3924	.78664	3.3929	.72377	3.6526	.72309	3.4747	.63977
	Year3	3.4399	.75676	3.3917	.79299	3.7090	.72055	3.5097	.63842
	Year4+	3.3685	.78385	3.4077	.67342	3.7318	.68574	3.4956	.58704
Total		**3.4205**	**.76939**	**3.4129**	**.71890**	**3.6946**	**0.69641**	**3.5047**	**.60528**

One-way AONVA used to examine the means differences of social networks uses among the students of Jordanian Universities according to University. The table below is showing the ANOVA results:

Table 6: Summary of F-test results according university:

Source		Sum of Squares	DF	Mean Square	F	Sig
Academic	Between Groups	28.925	3	9.642	17.391	.000
	Within Groups	400.844	723	.554		
	Total	429.769	726			
Intensity	Between Groups	11.402	3	3.801	7.553	.000
	Within Groups	363.803	723	.503		
	Total	375.205	726			
Group	Between Groups	11.960	3	4.097	8.718	.000
	Within Groups	254.023	723	.470		
	Total	**265.983**	**726**			

The table of ANOVA shows significant differences of all social networks uses among the students of Jordanian Universities according to the University. Scheffe Post Hoc Test where used to determine which means (University) differ. The multiple comparisons are shown in the following table:

Table 7: Summary of Scheffe Post Hoc Test:

Academic		**Hussein**	**Balqa**	**Sumaya**
	Jordan	-.10905	.30382*	.39876*
	Hussein	-	.41287*	.50781*
	Balqa			.09494
Intensity		Hussein	Balqa	Sumaya
	Jordan	.07001	.34440*	.04365
	Hussein	-	.27439*	-.02635
	Balqa	-	-	-.30075*
Group		Hussein	Balqa	Sumaya
	Jordan	.21874*	.34797*	.14914
	Hussein	-	.12923	-.06960
	Balqa	-	-	-.19884

The comparisons show:

1. The Academic uses of social networks are significant at the level α≤0.05 between both of the students of Jordan University and Hussein University and both of Balqa and Sumaya Universities students, where the academic uses among the students of Jordan University and Hussein University's students, where more than Balqa and Sumaya Universities' students.
2. The intensity of using social networks, where significant at the level α≤0.05 between both of the students of Jordan and Hussein Universities versus Balqa University, where the intensity of using social networks is more higher among them than the students of Balqa University, While the students of Sumaya University have more intensity of using social network than the students of Balqa.
3. The Group uses of social networks are significant at the level α≤0.05 between the students of Jordan University and both Hussein and Balqa Universities' students, where the group uses of social networks are higher among the students of Jordan University.

The following chart showing the means according to the social networks uses and the universities:

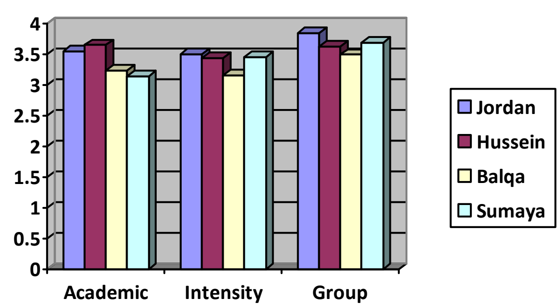


Figure 2: Social Networks uses according to universities

To examine the uses of social networks among the Jordanian Universities students according to Faculty and Gender, Independent Samples T Test where used:

Table 8: a. According to Faculty:

Variables		Mean	Std	T	df	Sig
Academic	Sciences	3.4498	.75932	1.693	725	.091

	Arts	3.3411	.79260			
Intensity	Sciences	3.4749	.69890	3.864		.000
	Arts	3.2449	.74677			
Group	Sciences	3.7546	.66935	3.855		.000
	Arts	3.5323	.74274			

The table shows that the Academic uses were not significant among the students according to Faculty, but the Intensity and Group uses are significant for the Sciences faculty students.

Table 9: b. According to Gender:

Variables		**Mean**	**Std**	**T**	**df**	**Sig**
Academic	Male	3.4048	.79492	.486	725	.627
	Female	3.4328	.74959			
Intensity	Male	3.3584	.73096	1.809		.071
	Female	3.4555	.70729			
Group	Male	3.6839	.69675	.367		.714
	Female	3.7030	.69688			

The table shows that the Academic uses were not significant among the students according to Gender at the level $\alpha \leq 0.05$.

One-way AONVA used to examine the means differences of social networks uses among the students of Jordanian Universities according to Year Level. The table below is showing the ANOVA results:

Table 10: Summary of F-test results according to year level:

Source		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig**
Academic	Between Groups	1.860	3	.620	1.048	.371
	Within Groups	427.909	723	.592		
	Total	429.769	726			
Intensity	Between Groups	.543	3	.181	.349	.790
	Within Groups	374.662	723	.518		
	Total	375.205	726			
Group	Between Groups	.767	3	.256	.526	.664
	Within Groups	351.331	723	.486		
	Total	352.098	726			

The table of ANOVA shows no significant differences of all social networks uses among the students of Jordanian Universities according to the Year Level.

7. Conclusion

After examining the results discussed above, it appears that Jordanian universities students are using Social Networks in a very high degree at all levels, especially group use, while the academic uses are existed, beside intensity. It appears that students are willing to use Social Networks as a communication tool for the benefits of academic uses. It is up to educators who can take advantages of Social Networks to be integrated within the academic role, as Social Networks are part of students' every day activity, and they are ready to communicate between each others in order to deal with coerces, even if those sheets or memos are not confirmed by their teachers or in a formal frame.

The researchers recommend the policy makers in higher education in Jordan and Universities, to integrate web 2.0 beside web 1.0 to learning management systems as a tool of communication, which may help greatly both students and educators.

References

[1] Andraws, S.; Al-Shalabi, H.; Jwaifell, M.; Awajan, A.; & Alrabea, A. (2013).The Intensity and the factors affecting the use of Social Network Sites among the students of Jordanian Universities. IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 3, 493-498.

[2] www.arabsocialmediareport.com

[3] Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication, 13(1), 210-230.

[4] Donath, J. (2007). Signals in social supernets. Journal of Computer-Mediated Communication, *13*(1), article 12.

[5] Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication, *12*(4), article 1.

[6] Golder, S. A., Wilkinson, D., & Huberman, B. A. (2007, June). Rhythms of social interaction: Messaging within a massive online network. In C.

[7] Greenhow**, C**. & Robelia, E. (2009). Old communication, new literacies: Social network sites as social learning resources. *Journal of Computer-mediated Communication*, *14*(4). 1130-1161.

[8] Hao Jiang; MingXi Tang;(2010) , "Web-Based Learning Platforms Integrating Social Networking for Design Education at High Schools in China," Computational Intelligence and Software Engineering

(CiSE), 2010 International Conference on , vol., no., pp.1-3, 10-12 Dec. 2010.

[9] Hargittai, E. (2007). Whose space? Differences among users and non-users of social network sites. Journal of Computer-Mediated Communication, *13*(1), article 14.
Lin, Nan, 1999, "Building a Network Theory of Social Capital", Connections, 22, 28-51.

[10] Hewitt, A., & Forte, A. (2006, November). Crossing boundaries: Identity management and student/faculty relationships on the Facebook. Paper presented at CSCW, Banff, Alberta, Canada.

[11] Hopkins, W. G. (2008). Quantitative Research Design. Retrieved from: http://www.sportsci.org/jour/0001/wghdesign.html.

[12] Lampe, C., Ellison, N., & Steinfield, C., (2006). A Face(book) in the crowd: Social searching vs. social browsing. Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (pp. 167-170). New York: ACM Press.

[13] Lampe, C., Ellison, N., & Steinfield, C. (2007). A familiar Face(book): Profile elements as signals in an online social network. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 435-444). New York: ACM Press.

[14] Ractham, P.; Firpo, D.;(2011) , "Using Social Networking Technology to Enhance Learning in Higher Education: A Case Study Using Facebook," System Sciences (HICSS), 2011 44th Hawaii International Conference on , vol., no., pp.1-10, 4-7 Jan. 1-10.

[15] Riva, G. and Galimberti, C. 1998. Computer-mediated communication: Identity and social interaction in an electronic environment. Genetic, Social and General Psychology Monographs, 124, 434-464.

[16] Steinfield, B. Pentland, M. Ackerman, & N. Contractor (Eds.), Communities and Technologies 2007: Proceedings of the Third International Conference on Communities and Technologies (pp. 41-66). London: Springer.

[17] Stutzman, F. (2006, April). An evaluation of identity-sharing behavior in social network communities. Paper presented at the iDMAa and IMS Code Conference, Oxford, Ohio.

[18] Valkenburg, P. M., Peter, J., & Schouten, A. P. (2006). Friend networking sites and their relationship to adolescents' well being and social self-esteem. CyberPsychology and Behavior, 9, 584−590.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Quality Issues in Infrastructure as a Service

Nikita Yadav[1] and V.B Singh[2]

[1] **Computer Science, Singhania University**
Pacheri Bari, Rajasthan, India

[2] **Delhi College of Arts & Commerce, University of Delhi**
New Delhi, India

Abstract

With rapid development in IT world, technologies are getting dynamic and very advance. With this advancement in IT world, online services have also been proliferated a lot. Now a day, computer resources are made available to users online on temporarily basis. This technology is cloud computing. In cloud computing environment, different services are provided to users online. One kind of such service is IaaS in which users are provided with infrastructures for certain amount of time. These infrastructures are computer resources such as storage, virtual server etc. storage infrastructure is an important resource provided by clouds as IT works on information and this information needs storage.
In this paper, we have described the need of cloud of good quality and also the quality issue in cloudcomputing environment. We have also devoted a part of work to highlight that storage is important infrastructure of clouds and Why SAN with Machine Learning is needed. Some change in SAN model has also been proposed here.
Keywords: *Cloud computing, quality, SAN, middleware, machine learning*

1. Introduction

As we know today's is the world of technology. Technology can be mechanical, chemical or IT field. When we talk about technology in IT field, computers play a vital role. Computers can be used to work alone inside a home without internet or with internet inside an organization. With internet we can use power of computers at a completely different location. Here's the clouds come into the focus.

Clouds in IT mean storage and computing means online services. Online availability of computer resources over a network for a temporary amount of time is known as cloud computing. These services are not provided to consumers permanently Cloud computing is the latest and superb technology for storing, sharing and controlling data online in an efficient manner [5].

The name cloud computing was inspired by the cloud symbol that's often used to represent the Internet in flowcharts and diagrams

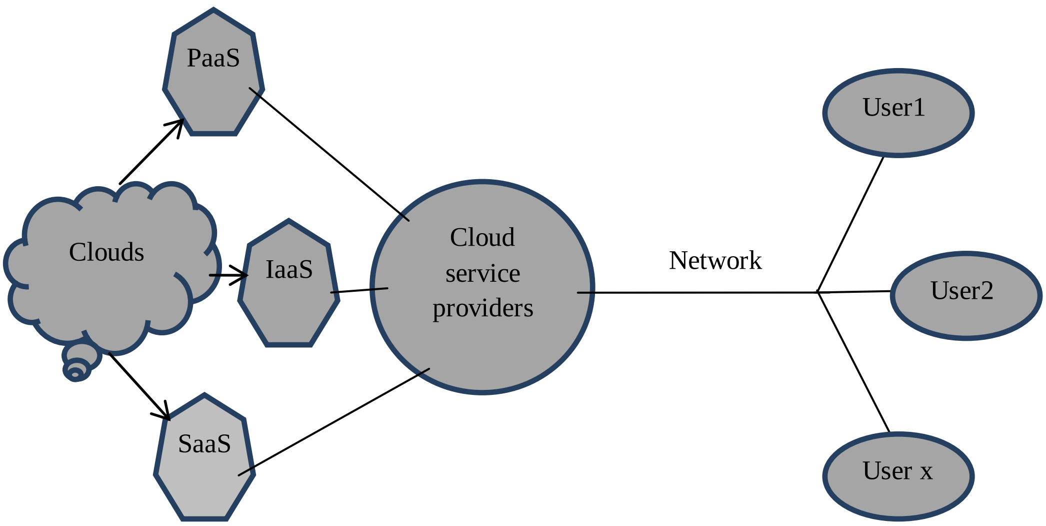


Figure1: communication between cloud and users

In above figure, communication between clouds and users has been shown. Users contact to cloud providers over a network which in turn makes an access to clouds. Clouds provide three types of services and out of these services which are of users interest is provided to them by cloud providers.

1.1 Elements of clouds

To make cloud working there are number of elements. These elements are

- consumers
- cloud providers
- services of clouds

1.1.1 Consumers: -consumers are the person who consumes cloud services. Consumer can consume these services for a temporary amount of time as these services are not permanent. With clouds users can save their money as there is no need to buy these services.

1.1.2 Cloud providers: - cloud providers are the agents who provide cloud services to consumers. Cloud provider offers services at different prices and performance levels with different set of features.

1.1.3 Services of clouds: - Basically, three types of services are provided by clouds [3]

1. IaaS:-In this service, computer infrastructure – typically a platform virtualization environment – as a service, along with raw storage and networking are delivered.
2. Paas: -In this service computing platform and a solution stack are provided as a service. In the classic layered model of cloud computing, the PaaS layer lies between the SaaS and the IaaS layers
3. SaaS: -This service is sometimes referred as "on-demand software”. In this software and its associated data are hosted centrally typically in the cloud and are accessed by users using a thin client normally a web browser over the Internet.

2. Why We Need Quality in Clouds

In cloud computing environment, Satisfying customer with good quality and low cost is very challenging task. As we know, in this competitive and fast growing technology world Quality counts a lot. But the question is how the quality count is and why we need it. A good quality count is that in which customer get’s what it wants. In fact in this competitive world, quality count’s with customer desires and satisfaction of those desires above their expectation level. Requirement changes with time and so the quality. But to get the best quality with low cost is a really challenging job.

2.1 Need of Quality for Cloud Providers:-

- In vogue: - today people want to be trendy with technology. With time they want change in cloud usage and its feature. They want something new in cloud services which are far much better then services provide by other cloud providers.
- Value for money: - value for money is another reason that customer want that much of quality in cloud services that much they have spent on clouds.
- Convenience: -The quality of being suitable to one's comfort is all that the customer wants. Customer wants quality so that cloud will suit to their comfort.

2.2 Need of quality for cloud providers:-

- Reputation issue: - In this competitive world, quality pays a lot to cloud provider reputation in IT sector.
- Minimize risk: - good quality cloud minimizes the risk of operational failure and hence control the error rate.
- Reusability factor: - a good quality cloud can adapt to any hardware and contribute to reusability and increases profit.
- Quality is cost effective.

3. Quality Factors for Clouds

There are many factors on which the quality of clouds depends. These are:-

1. Transaction Speed: - transaction speed is the amount of time in which user can have access to data on clouds. A good cloud service provides access to its data base in minimum amount of time. Transaction speed should be good enough so that users can have access to cloud database in minimum amount of time. There are several protocols which are used to accelerate transaction speed.
2. Storage: - Space where database of clouds is stored is known as storage of clouds. There are many factors in storage which are responsible for up gradation or degradation of quality of clouds. SAN (storage area network) is the new technology to improve the quality of storage. Use of AI (artificial intelligence) in storage also adds to the quality of storage.
3. Network security: - network security is the security that should be provided over the network during the transmission of data from server to clients. During data transmission data hacking, data manipulation should be avoided and for this cryptography and proper network protocols should be used.
4. Error free service: -transaction from server to clients should be error free and fault tolerable. If a fault occurs during the transaction then that session should resume as frequently as possible and for this error free protocols should be implemented.
5. Performance: - performance refers to overall quality of cloud that depends on all above factors.

Above factors can be improved by:-

- Speed: - by using speed transaction protocols
- Storage: - SAN, middleware and machine learning
- Network security: - security control model [3]
- Error free services: -Graph model,Markov model

Out of above quality issues, we have taken storage factor and describes that how can we improve storage quality of clouds and how this factor will contribute to increase in cloud quality.

3.1 Storage: -Space where database of clouds is stored is known as storage of clouds. There are many factors in storage which are responsible for up gradation or degradation of quality of clouds. For upgrading cloud quality we have discussed some technology here.

1. SAN (Storage Area Network)
2. Middleware
3. Machine learning

3.1.1 SAN: -SAN (Storage Area Network) is a block-based storage. As we know storage is the heart of cloud as all the services and data need storage. Main challenge for cloud providers is to provide scalable storage at the peak demand.

<u>Benefit of SAN</u>:-

- Scalability
- Manageability

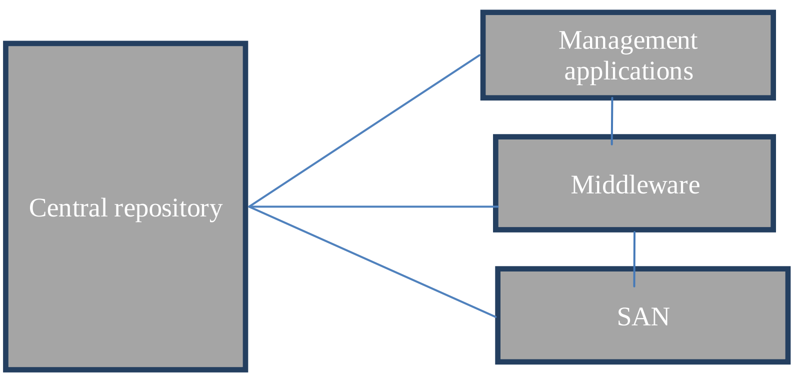


Figure2: middleware in SAN configuration

3.1.2 Middleware: -Middleware [1] is a module that lays between SAN, central repository and management applications. It checks whether proposed changes to configuration violate the best practice rules. If it does then it report to cloud providers through management applications.

Benefit of middleware: - Main benefit of middleware is to reduce heterogeneity of storage which is the main factor for degradation of cloud quality and also a main challenge for cloud providers.

3.1.3 Machine Learning: -Machine learning is an Artificial Intelligence, in which machines are made more smart and intelligent.

Here, smartness means machines are capable to analyze hidden data and intelligence stand for those machines which can learn and understand without being explicitly programmed.

Benefit of Machine Learning: -benefits of machine learning lies in the generation of best practice repository. Best practice repository contains the best validation rules which should be imply within the clouds. Generating these rules manually requires lots of time and is very costly. So with machine learning we can have best rules which are

- Error free
- Inexpensive
- Require less time

In SAN configuration model [1] as shown in following figure, reconfiguration request from management application module is sent to central repository through middleware. This access to middleware has a drawback as it increases the avenue time because here two accesses take place, one from management application to middleware and another from middleware to central repository. Because of this, an avenue time increase which also increases service response time and due to this quality of cloud suffers.

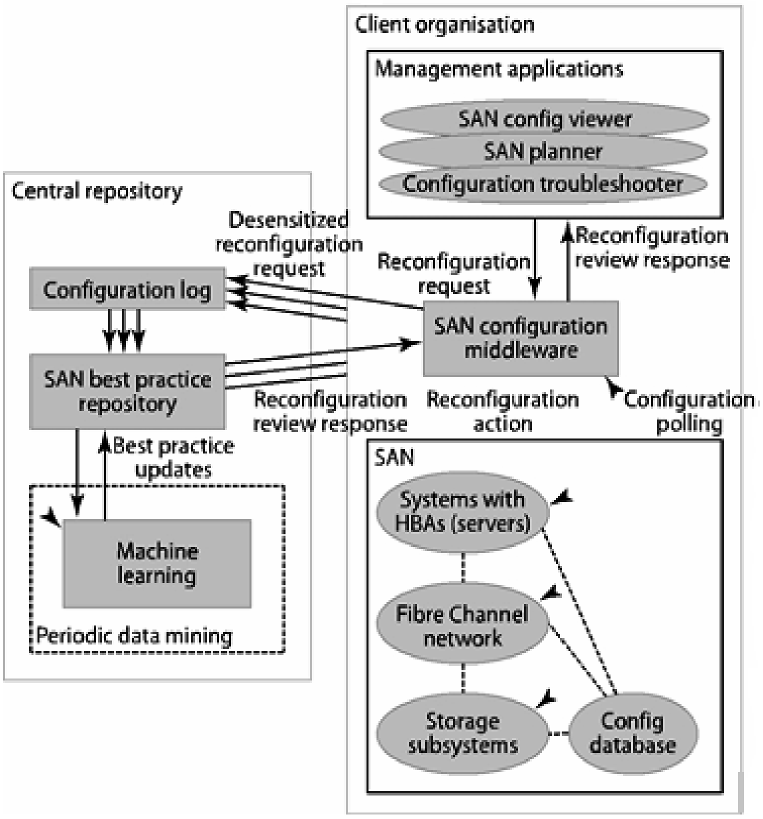


Figure 3:- SAN Configuration middleware

4. Proposed Plan of Work

In this, we have proposed some changes to the SAN configuration middleware [1]. In earlier model, reconfiguration request from management application module is sent to central repository through middleware. This access increases the avenue time as two access take place here one from management application to middleware and another from middleware to central repository. To decrease this avenue time which in turn increase service response time, we have proposed reconfiguration request direct to central repository from management application in spite of reconfiguration request from management applications to middleware and then to central repository.

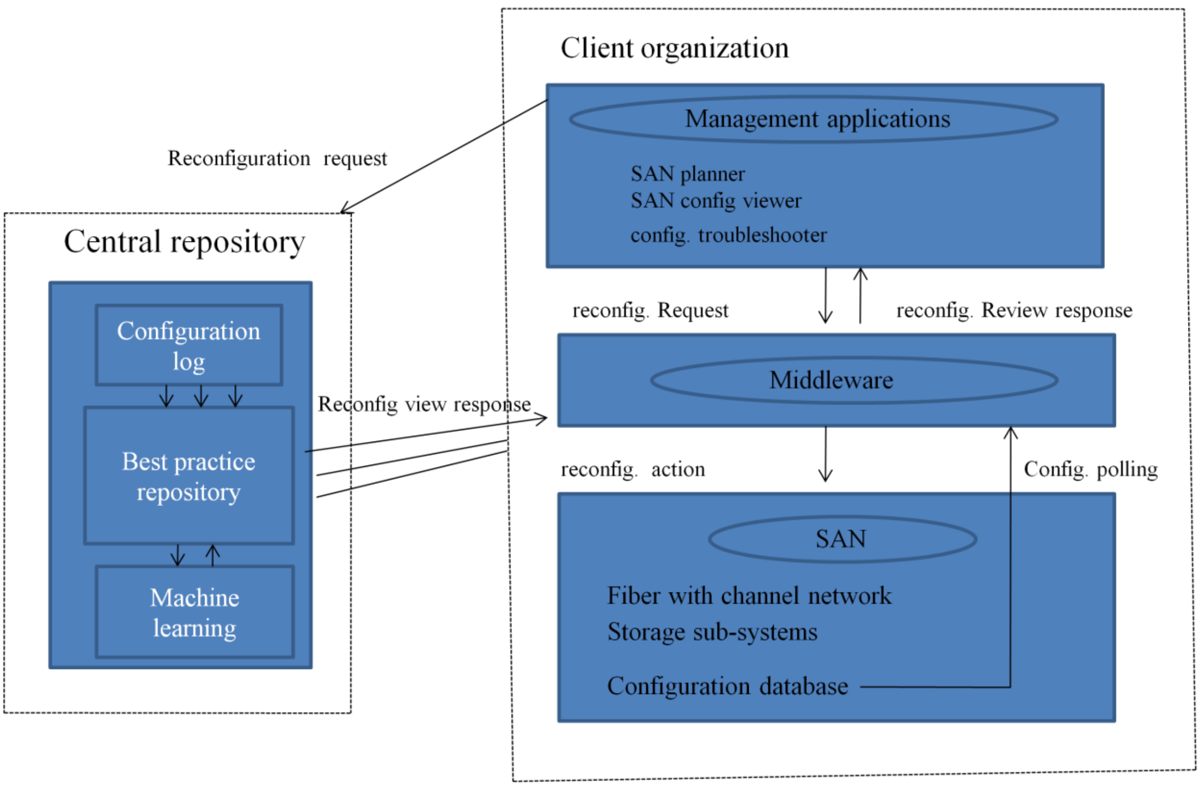


Figure4: Proposed SAN configuration middleware

In this model, machine learning play's a very important role. Best validation rules that are required to update the database are generated by machine learning. Machine learning is a technique in which machines are made more smart and intelligent. Intelligence of machines can be of supervised learning or unsupervised learning. In supervised learning, functions are mapped from input to output while in unsupervised learning

Now any request that will be done by management application will be directly sent to central repository in spite of being sent to middleware. Best validation and compatible rules generated by machine learning will be sent to best practise repository. Reconfiguration view response by best practise repository then will be send to middleware and reconfiguration action will take place on SAN sub-systems. Through configuration database, configuration polling will be send to middleware which in turn send this reconfiguration response back to management application. If management find this reconfiguration appropriate and suits their interest then this configuration changes is implemented to SAN otherwise again reconfiguration request is send to central repository and above cycle continues till management finds it appropriate.

4.1 Advantage of proposed SAN middleware

- As transaction is direct from management application to central repository, avenue time got decrease.
- Due to this decrease in avenue time, service response time will also decrease and hence quality of cloud also increases.
- Above model increases sustainability as reconfiguration within SAN module can be checked by best practice repository.
- With above model, adaptability of cloud services increases and hence suitability to customer requirement also increases.

5. Some Cloud Providers

Following are common cloud providers with their respective services [6].

	Providers	**Service**
1.	Amazon	Infrastructure
2.	Lunacloud	Infrastructure
3.	Lanlogig	Cloud Management
4.	VMware	Infrastructure
5.	Qrimp	Platform
6.	Thinkgrid	Infrastructure

6. Conclusion

In this paper work, we describe cloud, its utilities and its elements. In clouds, not only kind of services that it provides matter but also quality matters a lot. In today's rapid development world customer wants to be updated with latest technology and trend. They want best quality whatever services are provided to them. Quality is meant not only to customers but also to cloud providers. Cloud quality depends on several factors which have been discussed here along with storage factor. Manual planning of storage infrastructure is time consuming and error prone process. To gain scalability, SAN is employed by cloud providers but with increase in storage demand and network heterogeneity, SAN is not only the solution. The solution lies within Machine Learning and Middleware. At last we have proposed some changes in configuration in SAN model [1] and advantage we get with this change have also been discussed here. Some cloud providers with their respective services have also been discussed. In future, we will work towards improving the quality of the services of clouds

7. References

[1] David M.Eyers, RamaniRoutray, Rui Zhang, Douglas Willcocks and Peter Pietzuch, Configuration large-scale storage using a middleware with machine learning, Wiley Online library.DOI:10.1002/cpe.1716.

[2] Jae Yoo Lee, Jung Woo Lee, Du Wan Cheun, and SooDong Kim , A Quality Model For Evaluating Software-as-a-Service in Cloud Computing, Department of Computer Science, 2009 Seventh ACIS International Conference on Software Engineering Research, Management and Applications

[3] Yuan-Shun Dai, Bo Yang, Jack Dongarra, Gewei Zhang, Cloud Service Reliability: Modellingand Analysis.Weiss, A. "computing in the cloud," networker, Vol.11, No. 4, pp.16-26, 2007

[4] Turner, M., Budgen, D., and Brereton, P., "Turning Software into a Service," IEEE Computer, vol.36, No.10, pp.38-44, 2003

[5] Benefit of cloud computing with Amazon web services at: http://aws.amazon.com/

[6] http://cloud-computing.findthebest.com/

Security Issues and Challenges – Cloud Computing

Daniyal M. Alghazzawi[1] Syed Hamid Hasan[2]

[1,2] Information Security Research Group
Faculty of Computing and Information Technology, Department of Information Systems
King Abdulaziz University, Kingdom of Saudi Arabia

Abstract

Cloud computing is in its preliminary stages when we respect to its implementation & usage. The reason for this is that the technology is heavily dependent on the high-tech resources that the academicians and researchers find it practically impossible to analyze & test the technology. The basis of the cloud computing process is a desire to have another layer added while information is processed. Cloud computing is generally understood as computing in grids and as a utility where the software is provided as a service and the storage is provided through the cloud via virtualization, which means that the client remotely uses a service provided by a provider commonly termed to be "in the cloud". It is being debated if the aforementioned components can be separated and can be individually dealt with, however there is a general agreement that they all come under the cloud computing collectively. There have been many discussions related to the Cloud computing from the actual providers of the Cloud computing, as there has been a lack of published work from the academicians. However, the researchers are also making their presence felt by addressing many of the issues related to cloud computing.

Keywords: *Cloud Computing, grid computing, Security Issues, Secure Architecture Model*

1. Introduction

The concept of Cloud computing should not be termed as an invention, but it is has its origins in the needs to provide an infrastructure, as a service, that is aimed for requirements of mega computing or big storage spaces. From the viewpoint of the provider it is a mechanism by which peripheral infrastructure is provided so as to mitigate any downtime in the network without the users being aware of it, in addition to the fact that the users' data is transferrable in-between clouds. Balding at [1] suggest that the front-end software, the middle ware and the backend be designed in the form of layers making it possible for each layer to be independently dealt with in implementation, testing and functionality. The following paper attempts to introduce cloud computing in its present state, along with the research efforts from the industry and the academicians and the challenges faces in its development. It goes on to describe the security issues and the benefits in cloud computing along with presenting a secure architecture model that can be used for implementing cloud computing.

2. Present Viewpoint

The Critics of cloud computing contend that it is not a secure mechanism as the data is no longer present within the safe confines of the Company's LAN. The security of the data is dependent on the level of commitment the vendor has in enforcing security policies and verification done by third party. The verification is being offered to the clients as an on-demand service by vendors such as Google, Salesforce and Amazon etc. Statistics at [3] show that 1/3 of the security breaches occur because of the unauthorized access to machines, whereas information being stolen by employees amounts to 16% of the breaches. The aforementioned breaches can be completely avoided if the data is stored in the cloud. Additionally, since vendors have high availability of resources they can update security patches for applications, Operating Systems and middleware much faster. As per the vendors offering cloud service, most cases of theft occurs when the data is not protected properly by users that have authorization to access data. In case of a session in a cloud we can configure the browser to delete

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

all sessions data and to maintain a log in the cloud that describes what data was access by which user. This approach is far secure than storing the data on the client side. For certain application the best option is cloud computing, like the NY Times that uses the cloud service provided by Amazon. It uses the service to convert its archive of articles into PDF format. If the servers at NY Times were put to use they would have taken about 14 years in completing the task, however the same task was achieved by the cloud in a matter of just a day at the expense of a few hundred dollars. [4]

Despite the advantages the technology is still only being used by small companies that want to reduce cost, or individual developers that want to offer their application as services on the cloud. The big names are only testing the cloud computing with small time applications. However, we cannot blame the fraternity for this cautious approach and we will discuss below the Continual Development, the research efforts of the academicians and the industry along with the challenges of cloud computing.

2.1 Development

New developments are continuously being added into the cloud computing technology with the aim of enhancing security. One such product is QualysGuard, which is meant for discovering the loopholes in a network. It has acquired considerable reputation after being used by almost 200 of the Forbes Global 2000 companies. The product functions by placing an apparatus behind the Firewalls where it monitors the network for any breaches or threats. All data is encrypted by the device and it has no access to the client's data, with just access to only specified IPs and the administrator for modification of script & credentials. The product provides a new mode of security where any possible threats or attacks are monitored and dealt with by a 3rd party. In case of an unauthorized access attempt or an attack on any of the services rendered by the cloud, the solution cuts off access to the service from the source of the threat and prevents it from affecting the service availability.

2.2 Research efforts by Academicians & Industry

A few IT companies and universities have partnered to further research in the field of cloud computing. They aim to explore methods to reduce the cost of computing and storage in the clouds. The group is named Open Cloud Consortium (OCC). It also aims to establish the standards of communication amongst various providers. The OCC was formed in the middle of 2008 making it a fairly new entity, and showing that the entire discipline is fairly young. Following are some of the aims of the OCC:

1. To develop the cloud computing standard & framework to be used for cloud interoperations
2. Development of benchmarks
3. Implementation of support references for cloud computing, if possible open source references.
4. Management of a cloud computing testbed called "Open Cloud Testbed".
5. Sponsoring workshop and event linked to cloud computing.

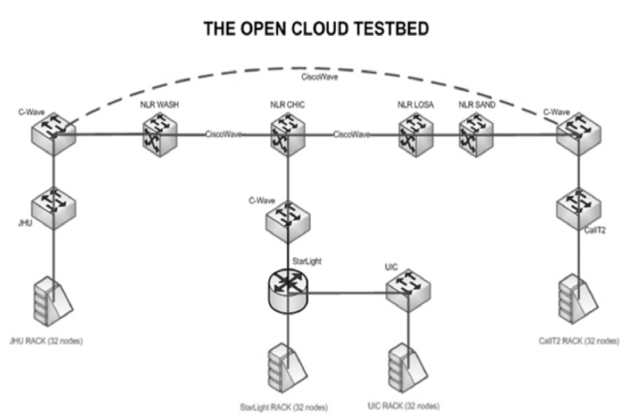


Fig. 1 OCC Network Layout

Figure-1 shows the network layout of the Open Cloud Consortium. The network is established between the servers at Universities of Illinois, John Hopkins, Calit2 and StarLight. Using the aforementioned architecture and the published network, the OCC has been working to implement a design and protocol for high traffic flow amongst various location[5].

2.3 Challenges

There are certain primary challenges that are faces by cloud computing in being deployed on a large scale by the enterprises, some the challenges are:

a. High Availability – Whenever there is disruption of service by failure of Network, Storage or application the vendor must ensure that the service does not get affected, through a backup that must run at all times. The switch should also be seamless so as not to have a major impact on the users.
b. Service Level Agreements – The network is governed by the service level agreement that necessitate multiple instances of the application on different servers so if it is required a low priority application may have to be shut down or minimized.
c. Multi-Tenancy – There is a possibly a conflict of interests arising of the fact that the same application and hardware is being used by many customers at the same time.
d. Linear scalability – The cloud should be able to cope with the requirement of increasing the data processing capability linearly as the if we have x number of users requiring the resource y then the time for completing the request for resource should remain Y*1 rather than X*Y so the cloud should be capable of adding X number of devices.
e. Management of Data – When we have a large number of clients using the same resources like data storage then to distribute, partition, secure and synchronize the data is a challenge.

3. Security Challenges

The small companies, most of the times, do not have suitable protection available against the attacks as they lack hardware as well as software resources. Loopholes in the programing make the firewall ports vulnerable. This challenge compels the small companies to switch to cloud computing rather than option for maintaining their own hardware. Yet, cloud computing also has its share of drawbacks. For example, if the cloud fails than it becomes responsible for the failure of multiple clients connected to the cloud. Additionally, an optimal approach to Low power transmissions & high availability needs to be found despite the fact that major players like AT&T believes that distributed cloud architecture is the optimal mode of implementing the cloud. [2]. It is believed that due to the lack of an effective security policies would soon cause most of the major clients to shift away from the cloud computing approach. One more issue is the different methods and approaches in storing of data by the various cloud providers thus making is more difficult to establish the distributed cloud architecture between them.

3.1 Data Security

The term security encompasses various factors like confidentiality, integrity & availability. This a major concern for the providers. With Confidentiality we mean to decide who has the onus of storing of encryption keys used to encrypt the data of a company, The encrypted data stored with the vendor must be kept secured from the employees of the vendor. A possibility is that the client stores the encryption key.

With Integrity we mean absence of a common policy for approved data exchange; different protocols are being used by vendors to transfer various software image or job across the industry. We can work around this, for maintaining the security on the client's side, by using thin clients requiring the minimal possible resources and not storing any data on the client side so that there is no possibility of the data being stolen. The method is resistant to the attacks that aim on acquiring this data on the client side. Some, companies contend that by implementing a system with unpublished APIs we can provide better security, but it also have the danger of being reverse engineered. Additionally upgrading of firmware via DHCP & FTP is considered to be insecure for quite some time. However, the very same features being implemented through thin client products from Wyse are considered to be the safest available. [1]

Last but not the least is the most crucial problem of availability, as many of the companies that use the concept of cloud computing have suffered from downtime. E.g. a "denial of service" attack was experienced by the servers of Amazon.com. One more thing to consider is that contract policy exist between the client and the vendor implying that the data would belong only to the client and third party may not get involved at any manner at any point of time. Authentication should also be proved by a combination of hardware and software means like a flash card or a finger print reader along with a password. This provides us with an advantage that the security of the client software may not be as stringent as earlier.

3.2 Cloud Computing Security Issues

A number of issues must be tackled before making it possible for the large companies to adopt the cloud computing architecture. The major ones are:

a. User access Privileges – Transmission of on the internet is prone to be attacked and is susceptible to risk owing to the issue of ownership of data. Thus companies should have ample knowledge of the policies of the provider. They may even test them by using a small application first.
b. Compliance to Regulations – The security of the solution and its accountability finally rests with the client as they have the option to choose between vendors who agree to be audited by 3rd parties for security and those who do not.
c. Location of Data – Some of the companies do not even have the knowledge of the location of their data, based on the contract type.
d. Segregation of Data – Mechanism for segregation of data need to be in place for as encrypted data for different clients is stored on the same storage.
e. Recovery – There should be a disaster recovery mechanism deployed by the vendor for protecting the user's data.
f. Investigative support – The Companies should have recourse to different legal means of investigation, if it is suspicious of any dubious activities by the vendor.
g. Viability for a Long term – The companies should have the option to cancel the contract and retrieve all the data if the vendor company is taken over by some other company.

Even though we understand that not all of the aforementioned points are relevant for all applications however, it is still critical that the concept is subjected to some sort of standardization if it needs to be accepted by the large enterprises.

3.3 Benefits

We know that we have a number of issues related to security that need to addressed when it comes to security, even then the concept brings forth a number of benefits related to security of Data. We now discuss some of such concepts like Incident Response, Centralized storage of data and logging.

The approach of Centralized Data is similar to putting all the eggs in a single place. Even though it presents a danger of being dependent on one resource, however it makes easier to monitor the data. The cloud data storage does away with many issues like misplacing the devices containing data, which is single largest reason of data loss by large companies or governments. In the cloud architecture only a small amount of data related to the thin client would be stored on device and even for connection the authorization is done at the cloud. Additionally when an access device is reported to be stolen the administrators can block all access attempts originating from the MAC address of the device. It is also easier to encrypt the data on the cloud then having to encrypt the data on each individual disk or backup tape.

The concept of Incident Response means the clients ability to have its hands on any required resources whenever required. This resource could be a processing capability, storage server or even a test environment. The concept circumvents the traditional and tedious process of securing resources in the traditional corporate environment. Additionally, in case of downtime of a server for maintenance the environment may be easily recreated by the client on an alternative machine with ease causing an improvement in the acquisition time. Hashes and checksum are provided for each file stored on the cloud by the vendors to evade the requirement of encryption of the files on the client side. However, it does not mean that the data is not encrypted before being sent on the internet, it simple means that the service is rendered from the cloud side.

Another feature available is the Password Assurance Testing which utilizes the processing capabilities available in the cloud to track and monitor the attempts made to break into the system of the company by use of password breaking/guessing., in turn it helps in minimizing the required resources and time from the client. The concept of Logging is benefited by the fact that the client side does not need to worry about maintaining the logs of the transactions happening and the cloud searching facilities yield the results much faster. Moreover, the accessing of any resource by the user at any point of time can be easily identified.

4. Secured Architecture Model

OSA (Open Security Architecture) provide a free framework capable of being integrated easily into any application. The framework has its basis on plans that depict the flow of information for specific implementations, along with the policy being implemented in every phase for security reasons. In the following figure we propose the architecture of a cloud to enable envisioning of the constituents of a cloud architecture alongside description of the element that makes the architecture secure. End users, System Architects, developers, third party auditors along with the cloud are the important entities that are involved in the flow of data. We now look at the attributes and the mechanism existing for them.

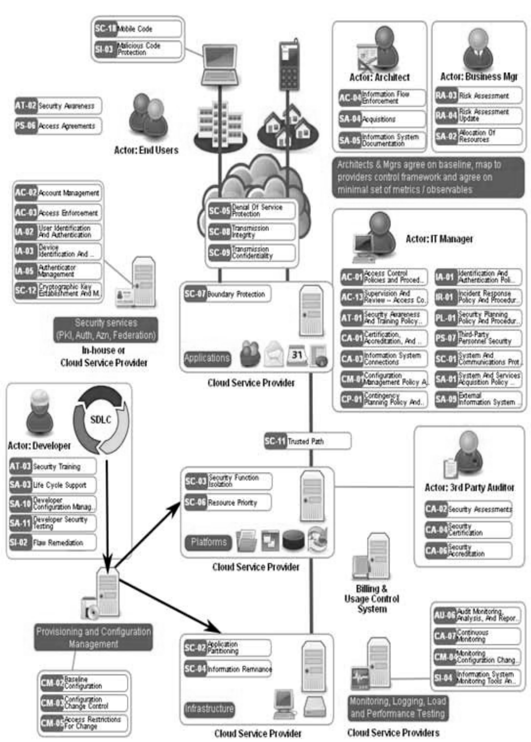


Fig. 2 Open Secure Architecture- Cloud Computing Model - [6]

4.1 End User

Access to certain resource, is required by end user, in the cloud and hence he must be conscious of the access agreement like the acceptable method of use / conflict of interests. Here, the commitment to the policy may be confirmed through the signature of the end user. All code and protocol on the client side devices such as the Server, Firewall or Mobile device, susceptible to attack must be detected and a patch uploaded to secure the local machine immediately upon detection. This ensures similar security for the cloud and the client. However, in case of the cloud there is a need of additional security from a user who tries to access the cloud with ulterior motives. Thus the denial of service (DOS) protection should be included in the cloud. One method of implementing denial of service protection is through improvement of infrastructure via increased bandwidths and higher processing capabilities which are in abundance in the cloud. From the traditional perspective, it involve filtration of specific packets that have analogous source IP addresses/server request. Once the security is catered to at the cloud level then we come across the issue of integrity of transmission between the cloud and the client. SSL/TSL (Secure Socket Layer/ Transport Layer Security) protocols can be utilized as one of the measured to ensure integrity of the transmission and preventing any attacks in the middle. IPsec (Secure Internet Protocol) can also alternatively be used to secure the network at the lower level. The last concern is to ensure that there is no eavesdropping on the session and the above mentioned measures can also be used to present unauthorized barging into the session.

4.2 System Architects

After the end users the system architect plays another role in flow of data. They write the policy pertaining to the configuration and installation of the HW component like firewall, server, router, & SWs like the OS and the thin client etc. System architects are also responsible for designating the control protocol that is to be used in directing the flow of data within the cloud e.g routers update/queuing protocol, proxy server's configuration or encryption of tunnels.

4.3 Developers

Next in line are the Developers who need access to the infrastructure of the development environment to build the applications in the cloud. Access to configuration servers is also required for testing of the applications designed from different viewpoints. Elasticity and availability of the resources offered by the cloud computing environment can be utilized to scale the SW environment and in turn help in the development E.g. if storage space is required by a developer he can get it instantly on demand from the cloud instead of going to the lengthy process to place a workorder & wait for long for the required permission. Additional virtual machiens may be prepared by the developers that can be used to obtain test data and for data analysis purposes all of which are time taking tasks. Additionally, the extra computing capabilities available at hand from the cloud can considerable cut down on the development time. The developers can also use the cloud for creating multiple evaluation environment to be used for the application, this will help to bypass the requirement of incorporating added security in the applications and burdening the cloud. However, the cloud computing environment is currently limited to Intel(x86) based architecture. Which proves to a be hurdle that the developers and the experts on the cloud computing need to be overcome. It is expected that the alternative architecture may also be available in the future. Call for server request from the API can be monitored to monitor Software. Thus Data centralization architecture makes is possible for focusing in a single direction making monitoring easier, even though it depends upon the clients and the developer about the amount of effort they want to put in this regards. The “Software as service” approach makes it easier for updating of the security patches on the cloud rather implementing the patches on each individual machines that have the SW on them.

4.4 Third Party Auditors

The security level of the cloud implementation need to be audited by a 3rd party auditor and this practice is supported both by the client companies as well as the Cloud vendors. Accreditation obtained by the vendor provides him the competitive edge and proves his level of commitment towards security. The process of accreditation is required to be carried out once in a three year cycle. So some companies may adopt an approach of continuous monitoring to lower the dependability on the cloud provider.

4.5 Overview

DMZ (demilitarized zone) approach of boundary protection is implemented by the cloud, like data centers, to ensure secure interaction among the various constituent entities like firewall, router gateways, storage & proxy server. Information that is most critical is stored on the other side of the demilitarized zone. Application partitioning and resource priorities are the other policies utilized to. Application partitioning uses a single server or storage for a number of client that have different forms of encryption used to encrypt data. The cloud must be able to distribute the view of the users of an application from the information stored on the backend. This can be achieved through may virtualization, multiple network adaptors or processors. Resource priority means allowing the processes or hardware requests to be queued in a queue on the basis of priority [6].

5. Conclusion

We can conclude that the concept of Cloud computing is yet to mature in to a full blown option that can be considered to be implemented by large conglomerates. And the challenge of making it so is being lead by the IT experts with the academicians nudging behind slowly. Formation of many groups like CSA and OCC have been done to find the offerings of cloud computing. They also aim at establishing a protocol for communication between various providers. In the present situation recognition is being gained by cloud computing however it faces many hurdles in doing so. However we need to carefully analyze the pros & cons, advantages & disadvantages of the concept in order to pass a judgment on the feasibility of Cloud computing being a major option of being implemented on a mass basis. Yet things look promising for the sphere as more and more researchers are option to pursue the task of analyzing and improving cloud computing.

References

[1] Craig-Balding, "ITG2008 World Cloud Computing Summit", 2008, http://cloudsecurity.org/

[2] Alistair-Croll,"Why Cloud Computing Needs Security", 2008, http://gigaom.com/2008/06/10/the-amazon-outage-fortresses-in-the-clouds/

[3] Elinor-Mills, "Cloud Computing Security Forecast:-Clear Skies", 2009, http://news.zdnet.com/2100-9595_22-264312.html

[4] Neil-Wienberg, "Cloudy picture for cloud computing", 2008, http://www.networkworld.com/news/2008/043008-interop-cloud-computing.html? p1 =rcb

[5] The Open Cloud Consortium, 2008, http://www.opencloudconsortium.org/index.html

[6] Open Security Architecture, 2009, http://www.opensecurityarchitecture.org/cms/

[7] Paul Jaeger, Jimmy Lin, and Justin Grimes, "Cloud Computing and Information Policy", March-2008

[8] John Rittinghouse, "Cloud Computing: Implementation, Management, and Security", 2009.

[9] Michael Armbrust, and Armando Fox, "Above the Clouds: A Berkley View of Cloud Computing", February-10, 2009.

[10] George-Reese, Cloud Application Architectures, O'Reilly-Media, April-2009.

[11] Jonothan Erickson, "Best Practices for Protecting Data in the Cloud", 2008, http://www.ddj.com/security/210602698

[12] Jon Brodkin, "Seven Cloud Computing Security Risks", 2008, -http://www.networkworld.com /news/2008/070208-cloud.html

[13] Geva Perry, "How Cloud & Utility Computing Are Different", 2008, http://gigaom.com /2008/02/28/how-cloud-utility-computing-are different/

[14] Ephraim Schwartz, "Hybrid model brings security to the cloud", 2008, http://www.infoworld.com/d/cloud-computing/hybrid-model-brings-security-cloud-364

Daniyal M. Alghazzawi has completed his PhD in Computer Science from University of Kansas in 2007, Master of Science in Teaching & Leadership in 2004 and Master of Science in Computer Science in 2003 from University of Kansas. He has worked as Web Programmer at ALTec (Advanced Learning Technologies). Dr. Daniyal is currently working as Associate Professor in the Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, KSA. He has more than 20 publications to his credit. His research interest includes e-Security and Cryptography and Educational Technology. Dr. Daniyal is a member of IEEE (Education Transaction) and ACM-SIGCSE (Special Interest Group in Computer Science Education).

Syed Hamid Hasan has completed his PhD in Computer Science from JMI, India, MSc in Statistics and PGDCS from AMU, India. Dr Hamid has a teaching and research experience of more than 30 years and is currently working as a Professor at Information Systems department, faculty of Computing and Information Technology, King Abdulaziz University, Kingdom of Saudi Arabia. Prof. Hamid has worked as the Head of Computer Science department at AMU, India and also Head of IT department at the Musana College of Technology, Sultanate of Oman. He was Reviewer for NDT 2009, Ostrava, Czech Republic, 2009 Co Sponsored by IEEE Communications Society. He is included in the Panel of referees of "The Indian journal of community health", and was Chief Coordinator of the National Conference on "Vocationalization of Computer Education" held on 28-29 September-1996 at A.M.U. Aligarh, India. He is a life Member of Indian Society for Industrial and Applicable Mathematics (ISIAM), Computer Society of India (LMCSI), Fellow National Association of Computer Educators & Trainers (FNACET), India and a member of the IEEE. He has more than 25 research articles in conferences & journals to his credit. His research interest includes e-Security and Cryptography, administrative computing and software engineering.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Security in Electronic Business

Dr. Abdullah Saad AL-Malaise AL-Ghamdi[1]

[1] Department of Information Systems
Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah-21589, PO Box – 80221
Kingdom of Saudi Arabia

Abstract

A crucial area of electronic transactions is the domain of electronic commerce. Yet, a large number of people do not want to transact online as they are not sure of the level of security that the transaction would be provided by the site and the technology used by the sites. According to surveys, one of the factors affecting the spread of ecommerce is the (lack of) security measures that assure both businesses and their customers that the business relationship and transactions will be carried out in secure manner. This paper describes the security requirements for electronic Business application and attempts to discuss the method and ways to be used to meet them.

Keywords: *e-commerce, e-business security, security issues, security technologies.*

1. Introduction

The purpose of the Internet was to transfer information between computers, to enable easy remote access to systems, and its use for commercial purposes has grown immensely since the advent of the Web. Transfer of information through electronic means, availability of the internet and other advantages of technology are altering our economy, and our business methodologies. These technologies have a great impact on almost every part of the business and social life. Business are continuously transacting amongst themselves and even the customers; people regularly take the help of internet to identify vendors, and to assess their product or service, to even compare prices, and to take the advantage of the market. But, e-commerce is a lot more than only purchasing and distributing good through the internet or electronic means: It means interaction within the company/business and/or outside network through the means of network and telecommunication technology. It would also include a commercial transaction done over any computer-supported link that has the ability to transfers itself or can help transfer of “value” for items and/or service offered this may include rights of property and ownership, or the rights to utilize the items and/or service.

The domain of e-commerce is considered to be a vital side of digital business. The rate at which e-commerce is currently growing is a good example of the fact that the full potential that this technology can offer is yet to be reached. During the early 2000 and the last part of the 1990’s there was a lot of speculation about what shape e-commerce would take and how it would develop in future. It was predicted, “a volume of US$ 184 billion of online retail sales in 2004 in United States alone” Forrester Research of 1999 [1] whereas the value of approx. US$ 69 billion, actual sales [2], represent huge gap. At first glance, it appeared we have tremendous opportunity for global commerce: a global communication backbone that is very conducive for low cost information exchange and a global economy that is trending to become highly information-base. However, one crucial cause of the difference between the two values has been proposed by the researchers’ fraternity and supported by the results of a number of studies is the scarcity of Confidentiality, Reliability and Security in electronic transactions. A substantial portion of people who are involved in business do not want to transact online as they are not sure of the level of security that the transaction would be provided by the site and the technology used by the sites.

The figures above refer to B2C interactions only, so it was projected that the condition of B2B, and between business & government, interactions is better, which is true. However the is still not what can be called as being optimal. In today’s world and economy, B2B transactions are making the biggest impact on the e-commerce industry. In industries, like automobile and mining, electronic business has at present made a major impact on the relationship between the vendor and the manufacturer. The reason for this fact is that communication between parties is carried out through dedicated line i.e. secure and private virtual networks. This promotes the development of trust, because these relations are long lasting and are also based on physical business from the outside world. However, for even the above mentioned industries the complete potential of ecommerce cannot be said to have been reached.

Thus for e-commerce to achieve its complete potential, either the involved companies need to increase the confidence level, and the confidentiality provided to their customers or the technology needs up gradation, so as to have strong features for protecting the privacy and security of the e-commerce transaction.

As the topic goes beyond one single function or arena of electronic commerce, we are required to have a global perspective of things. We are going to discuss the key factors and areas related to Confidentiality, Reliability and security of ecommerce transactions, in this paper. We will begin discussing the Issues related to Trust and confidentiality, then we will discuss the broad meaning of Confidentiality and technologies used for enforcing privacy. In the end we will complete by discussing the major factors related to providing security for electronic business and to its underlying infrastructures.

2. Confidentiality

Privacy and confidentiality is a vast topic. Maintaining of privacy is being discussed in many spheres like politics, society, law and in the recent times computer science. The electronic business arena links Confidentiality to using the information of the customer. Business Transactions usually include exchange of great amount data related to customers and businesses, which may be necessitated by electronic transaction itself (e.g. information related to credit card, bank accounts, and delivery location of personal address etc.) or it may be a requirement of the Business involved in the transaction. Customers practically have no idea about the uses that their information can be applied to, and so have little knowledge of the violation of privacy that might be happening. In the context of e-Business, we may define Privacy as the right of an individual to control, the collection of information about him during the e-commerce transaction; and to be informed the initiation of any process where his personal data would be gathered; and the right to choose details of sharing their personal information with others.

Initially the two point of views, of promoting the interest of the business and thus making world economy much stronger, or the one supportive of the views of the individual seems distinct. Yet, we presented with a need to achieve a consensus between the two, thus enabling us to find a solution which is beneficial to both the business and the individual. Such a consensus has been named *consumer-centric privacy*: under which individuals tends to attain the most possible confidentiality while business earns economic benefit by providing increased privacy to its customers. Economic benefit is derived out of improved public image of the business (means more and loyal customers) or improved brand recognition, which would naturally lead to an increased amount of e-commerce transactions by individuals who are now more comfortable in transacting online.

2.1 Concerns of the Consumer

In today’s world both the Consumer and the vendor are transact across the globe with almost anyone willing to do business. Earlier, the situation was completely different where acquiring of a good or service was the result of individual’s contact directly with the vendor who was located in the same locality, area or country. However, the current scenario presents us with certain aspect of these transactions that give rise to most of concern related to privacy and confidentiality. There is not direct contact between the consumer and the vendor, and no interaction between them, there is a gap between the buying, delivering and paying processes, from the timing perspective; during the business transaction information about the consumer can be collected effortlessly and very quietly without incurring too much cost as well, and effective regulation are either not present or are not enforced ineffectively; all above mentioned points contribute to these concerns. The latter is particularly true if transactions occur across borders and different set of laws are applicable. Several of the concerns of the consumer about privacy are discussed in ([3] - [6]). Following are certain examples that are noteworthy:

<u>Cookies</u>: Cookies are alternatively used for storing the behavior of the customer on his own computer which can be later matched with database of the customers and used for tracking him, once identified to have a certain repetitive behavior the customer might be forced to pay higher prices as his buying habits are being tracked by the vendors, and it is likely that the customer will even purchase at a high cost based on his profiles.

<u>Site spoofing</u>: Customers are being diverted to other sites where incorrect information is provided to them and based on this information their purchase decisions are affected. Or they are being lead to specific web-sites where the privacy policy of the web site where they came from does not apply, and they are unaware of this fact.

<u>E-mailing</u>: Spam mails being sent to consumers offering services and/or products.

<u>Data gathering</u>: Unauthorized use of the customer’s personal data for purposes of data quarrying or marketing, or even selling.

Lack of regulations: Different Privacy laws in different countries and no effective means to enforce implementation.

Privacy statements: The declared Privacy statements on the websites are not updated, not correct or in some cases they are left un-applied completely.

2.2 How to Preserve Privacy

Privacy can be achieved through various means like legislative steps, organizational steps, or technological means. A combination of different techniques may also work.

Legislative Steps

Consumers have established legislations in many countries to protect privacy. Following are some examples of legislations in different countries.

In Japan the governmental and commercial usage of an individual's information is regulated by the Act of 2003 called "Personal Data Protection" which is an enhancement of a previous act laid out in 1998 which regulated the use and storage of an individual's information by government & the administration. China also has several laws relevant to protection of Data.

The Personal Information and Electronic Documents Act 2004, in Canada determines for businesses the procedure for Customer data collection, usage and disclosure. United States even though does not have a regulations dedicated to privacy; however there are a number of laws that cater to different issues related to privacy.

Sweden and UK have imposed legal restrictions on possession of any individual's personal information in absence of explicit approval of the information owner, and if an organization or business need to store this information then the business has have this storage registered with the governmental agencies. The German law additionally requires the collection of data to be kept at minimum so as to perform a certain function and prohibits use of the data other than the specified and declared purpose. Data Protection Directive of the European Union declared privacy as a fundamental right in 1998.

However, the attempt to solely protect privacy via legal means is not as fruitful as it used to be. Legislations are lagging in the race with the latest technological developments and the legal system is not equipped to react at an appropriate pace to the changing technology. On top of it, laws are specific to a country.

Organizational Steps

Users and vendors have available organizational means that can be used in protecting the privacy of the individual during the ecommerce transaction. E.g. physical separation of consumer data into two parts one that is personally identifiable while the other is under the non-identifiable category. Data which refers to specific kinds of services or which tell about the types of the products that is individual is purchasing termed as is non-identifiable and should not be, combined with traits that are termed as personally identifiable, like the banking details, credit card information, date of birth, address or name of the individual. Under the effort to protect privacy we are concerned with only data that is personally identifiable, so Non-identifiable info can be put under analysis in a suitable manner.

Involvement of a transaction service provided by a third party is another means of achieving privacy and confidentiality. This third party may be required to act as an intermediary for guarantying the trustworthiness of the transactions. Another means of increasing trust and privacy may be the "online privacy seal" i.e. TRUSTe, or Platform for Privacy Preferences (P3P). These technologies give customer the ability judge if the Business' privacy policy is in accordance to what he prefers. Yet both the above approach rely on constructing awareness amongst the business about the privacy of its customers and cannot be an assurer of actual execution as per policy. This leads us to the same point where a customer has to basically rely and trust the vendor to keep its promises and not violate his privacy.

Technological Steps

We can use Privacy enhancing technologies (PET) to enhance consumer Privacy. These technologies provide the option to delink an individual and their personal data. We come across different modes of anonymity being described in literature, we can have full anonymity or pseudo-anonymity (where an individual's identity is normally un-know but if required it can be disclosed), or pseudonymity (where the users creates multiple virtual identities and uses them for different purposes). Anonimyzation can also be accomplished by one of the following three techniques: anonymizing the medium of transport, statistical databases, or anonymous access.

Through anonymizing the medium of transport, we strive to hide the identity of the user in an un-revealable manner.

Setting up a free email account with an email provider, that is trusted for not tracking information like IP addresses, is probably the simplest possible manner to accomplish this task. This approach, however, is not a very feasible approach as most of the email providers are legally required to keep details of the communication that are made through them for a specified time period and every transactions that exceeds a specific value must also be recorded and maintained. Another method to have anonymous browsing is through anonymizing servers. Under this method all the traffic originating from any user is passed through the anonymizing server, this makes the IP address or user's identity unidentifiable. However, this technique relies on the trust of the third-party for the information on not being disclosed.

In an attempt to discard these third-parties., the "Crowds" system was created by Reiter and Rubin [7]. This system clubs the users into large groups called the crowds. Whenever a user requests information from the internet the request is routed through this Crowd instead of being passed onto the recipient directly. Before being submitted to the website the request is bounced between a random numbers of hosts that are part of the Crowd. Thus the originator of the request, remains unidentified by the recipient. Another privacy enhancing technology uses encryption. Chaum Mixes is prominent in such technology, utilizing public key cryptography [8]. The entire message is divided into chunks of equal size that are cryptographically changed before being delivered to the recipient. The order of the messages is also changed. This process is aimed at making it difficult for linking the original messages to their senders and even the outgoing message is unlikable this prevents the possibility of performing any analysis of the traffic to establish the identity of the user. The original concept of Chaum Mixes was built upon in to arrive at various other methods. E.g. "Onion" protocols under which a dynamic mix of protocols is being used by the user to submit the request as a data structure. The request is structured like the peeling skins of an onion. The layers are designed in such a manner that at any point only the specific layer can be decrypted only to reveal the next point in the chain. Onion routing, even though is available on the internet for anonymizing users, is not free and incurs cost.

Another method available for privacy enhancing privacy is anonymous access. In such a system an individual remains unknown to the business organization except through his pseudo-name (awarded credential). A user has multiple unrelated pseudonyms. Certification authorities issue credentials to a user, he just have to prove that he possess these credential to an organization and he is not required to reveal his identity. However, this system has a weakness is that even though a user may be legitimate he can still pass on his pseudonym to another user. This can be regulated through linking together the certificate and the private key issued to the user, thus maintaining his privacy as well.

Anonymous access gave rise to the authentication and authorization infrastructure (AAI), which is based on the principal that it isn't required to know who is the user, it is sufficient that the particular user has authorization for the specified actions. What it means is that it enables customers to buy items online by hiding their identity but proving possession of the required authorizations, which means that they have authorized access to a bank account etc. or that they have already made the payment. This essentially infers that the AAI is entrusted to an organization that relies on these services. The uses and the type of AAIs are discussed at [9].

Statistical databases are utilized in another approach to privacy. A statistical database is a collection of data, e.g. data of the customers and the item that have been purchased by these customer but it does not reveal the info that can be used to identify any individual uniquely. The values stored in this database are not the data itself, but the statistical information. This gives rise to methods for keeping the statistics of the data useful while keeping the individual's data hidden. Few of the methods are :

a. Only allow queries that retrain privacy (Query restriction),
b. Change the data about the individual in such a manner that data is useless but the statistics remain unaffected (Data perturbation), or
c. Altering the result of a query if it appears to be a threat to privacy (Output Restriction).

However, the above described methods have a drawback that the data is made of less use by them.

3. Reliability

People have long carried out face to face business. This mode of business has been quite successful. One of the causes of this success is the Trust that is developed in these transactions, because of face-to-face dealing that is happening in this transaction. This concept of trust has its social & psychological aspects.

As mentioned, trust and reliability are at the center of any business dealing. It becomes even more important when we are dealing over the internet. We will also demonstrate that the definition of trust is not as simple as we may have presumed previously. Moreover, Trust need to be an integral part of any business dealing over the internet for the digital transactions to have the same level of acceptance likes traditional business. For example the vendor who is selling the product should have the trust that the person buying it is capable of paying for what he has bought while at the same time the purchaser need to be assured by the vendor that he would not disclose the private info of the buyer to anyone else.

3.1 Definition of Reliability or Trust

We have had numerous definitions of reliability and trust available. Some have tried to define trust in the general or global perspective, while some define it with reference specific types of applications. Reliability is defined as "The ability of an item to perform a required function under given conditions for a given time interval" [10]

One such attempt of defining trust in electronic commerce was made, where with reference to a system, Trust is defined as "a belief that is influenced by the individual's opinion about certain critical system features"[11]. Whereas reliability "In general, reliability (systemic def.) is the ability of a person or system to perform and maintain its functions in routine circumstances, as well as hostile or unexpected circumstances [12]"

Grandison and Sloman have also argued that because there is a lack of commonality on definition of trust it has directed them to interchangeably use the terms authorization, authentication and trust. Furthermore, they have defined trust as "the firm belief in the competence of an entity to act dependably, securely, and reliably within a specified context (assuming dependability covers trust and timeliness)" [13].

3.2 Authentication and Authorization Interchanged with Trust

Interchanging trust with authorization and authentication cannot be done according our belief, because authorization and authentication, understandability are the security service provided by the application, while we cannot consider the same for trust, it is the outcome (i.e. a belief) resulting from the appropriate application of Authorization and Authentication in a scenario, like reliability. However, we agree with the authors that there is no unanimity about the definition of Trust and Reliability.

We would also agree to the importance imparted to authorization and authentication, as these service are integral in getting the trust of consumers and merchants. In this respect, digital certificates has provided us with a solution that can be termed to have contributed a great deal in increasing the trust on security of e-commerce transaction on the whole and specifically the service of authorization and authentication.

One of the best available solutions to integrate authentication with Internet applications using the digital signatures is the concept of Identity certificates (or public-key certificates) [14]. However Internet application requires that it be clearly specified that which actions are permissible, for the user that has been authenticated by the authorization service. In order to achieve this we should consider putting restriction on the privileges as well. A solution for this issue is provided by Attribute certificates which in conjunction with identity certificates, provide authorization and the access levels as well [15].

Public-Key Infrastructure (PKI) is known to support encryption, integrity and non-repudiation at the same time, along with providing a trustworthy & efficient mechanism to distribute and manage these certificates. It is evident that in the absence of such a mechanism it is not practical to assume that the applications using digital signature would ever be a reality.

This Framework of attribute certificate also provide us with a base on which we can build a PMI (Privilege Management Infrastructure). The information contained in the Attribute certificate and the Identity certificate of each user link the PMI & PKI infrastructure. This link further establishes the fact that authorization must rely on authentication for proving the identity, and it is brings out, that the appropriate use of a combination of these Authorization and authentication certificate greatly increase the reliability and the trust on the system from the user's perspective. Although linked, both systems can be managed independently as well.

The attribute certificates can be used for numerous purposes, like it may contain information related to the membership of a group, clearance level, the role that an individual plays etc. Through the attribute certificate authorization can be provided to decentralized applications. Additionally through attribute certificates, digital business applications can conveniently make authorization information distributable.

3.3 Trust Management

The Management of reliability and Trust for digital commerce is most likely the most critical issue when dealing with this concept. The concept of trust management was introduced by Blaze and others [16]. It was proposed that PolicyMaker can work as a solution for managing trust. PolicyMaker allows a programming language to encode the access level and the identity of the individual.

Additionally to achieve standardization and facilitation of PolicyMaker's integration into applications, the name of KeyNote was suggested. It makes the use of assertion languages, which are accommodate the security policy of different application.
However, we know that trust and reliability of varies with time, thus digital certificate for attribute & identity are considered by some authors for trust management purposes. Procedures & functions provided by PMIs & PKIs can be considered advance methods of trust management. They seem to be better solutions than the ones discussed previously as they are less static, however they have a bias approach for the authorization as well as the authentication service.

3.4 Challenges of Trust & Reliability

We have already discussed, that even the simplest issues related to trust still need solutions and are open. And this too is just a marginal part of the complete set of challenges related to Trust. During a National Science Foundation workshop aimed at discussing the challenges for Management of Data and Information the following challenges were identified by a group of people who are experts in this Field:

a) How can we build & initiate trust?
b) How can we evaluate & maintain trust?
c) How to deal with fraud?
d) How to guarantee scalability, performance, and economic parameters for trust solutions?
e) How to engineer trust-based applications and systems?

They have additionally recommended the need of research being conducted in the following areas related to Trust: (a) Social paradigm (b) Liability. (c) Scalability and adaptability of infrastructure. (d) Benchmarks, developing of applications based on trust. (e) Detection & Prevention of Fraud (f) Interdisciplinary research in areas associated with Trust.

Thus we find it to be evident that still research needs to be conducted in this area in future as well.

4. Security

Five components namely data, procedures, hardware, software and people are interconnected & interacting in any e-commerce scenario. We come to a conclusion that e-commerce system is an information system that comprises of both an organizational framework and a Tech. infrastructure, not just the Tech. infrastructure alone. Thus, information system setting must be used when addressing the issue of security in e-commerce.

Under the present situation, we can define security as an organized framework that consists of policies, procedures, concept, believes, principles, measures and techniques required to protect, against any deliberate or accidental threat to the system or the individual assets of a system [17]. Operationally, when compiling this framework, we must identify the requirement first.

4.1 Requirement of Security

Close inspection of all e-commerce reveals the presence of discrete stages, this allow us to build a universal model that describes them all. A model similar to the one described above was suggested to cater to business dealings [18] and has also been demonstrated to be great for dealing with commercial transactions [19]. This model is based on the observations that the unit of any Business is an exchange transaction. In such a transaction, party X & Y, mutually agree that they will comply with specified condition of satisfaction. The Buyer (X) and the seller (Y) are part of this transaction. X makes a request to Y for providing a commodity to X which Y accepts, in return X pays Y for the commodity. This entire exchange can be explained clearly in a four phased cycle:

1. Request. X makes a request of Y for providing a service. (This often means Y taken up on an offer he has made).
2. Negotiation. X and Y mutually agree what will be provided (X's condition of satisfaction) and what will be the payment made (Y's condition of satisfaction).
3. Performance. Y performs the actions required to fulfill his part of the deal and informs X when completed.
4. Settlement. X accepts the work done by Y, accepting it to be satisfactory, pays.

We can combine Performance and Settlement can be combined in a single phase, and call it Execution phase [20]. The above model depicts all kind of transactions, including electronic transaction.

In the first phase, both the X&Y have unique requirement of security. X needs assurance that the offer in consideration is valid, that is, the information presented to him has not been corrupted by an external party. Similarly, Y needs assurance that the offer he made is available to X. Y may require his offer to be confidential; for fear that an adversary may barge in the Deal, if this is not a retail transaction. Confidentiality is also clearly required, in the second phase, especially when this is related to negotiations of contract. It is Important none of the parties are able to repudiate their offers in this phase. However, non-repudiation becomes more important during the Final leg. In this stage secured delivery of goods along with secure payment should be ensured. It is noteworthy that some items is intangible; and thus the seller can deliver it to the buyer electronically Example, digitally shares. The situation presents us an interesting requirement of security. Lastly, we see the fundamental difference between electronic & traditional Business, i.e. there is no face 2 face interaction in Electronic Commerce. Machine that don't have any way of identifying the person on the opposite side, are presented with pre-agreed information that establishes identification of the users.

Thus, the security requirement of Electronic commerce essentially include the necessity to maintain the integrity, confidentiality and availability of the information and the system, the authenticity of the entities involved that they are not able to repudiate their transaction.

4.2 How to Address the Requirement

For frameworks to preserve the information system's security it must include of technical, legal social & organizational action. Legal actions are to taken at a national & international level by the government. Individual organization need to take Technical and organizational action. Lastly, enhancing public awareness for security requirements and on the right and obligation arising from this requirement must be done on the social level.

The major legal issues associated with e-commerce [21-22], are:

(i) The protection of privacy.

(ii) The protection of intellectual property rights. Related to it is the problem of registering domain names to be sold later at a higher price i.e. cybersquatting. We also have an issue of Patent Protection in electronic commerce. National legislation exists almost everywhere for protecting of intellectual's property right, [23]. The World Intellectual Property Organization – WIPO (www.wipo.org) plays the most prominent role internationally, also administering 23 relevant international treaties [24].

(iii) Rights to freedom of speech contending the necessity to control potentially dangerous, illegal and offensive information. Including the issues spam control.

(iv) To be protected against Fraud, including identity fraud, regulation of taxation. To be protected against computer crime and money laundering etc.

Though we have legislation for maximum of the above subjects in traditional commerce but it cannot be directly applied to e-commerce environment. Therefore, we need to develop new laws or to ensure that the existing ones are applicable to the required situations.

The vast number of entities involved in e-commerce makes is impossible to use of symmetric encryption. As it is not possible to maintain & manage certificates and keys for such large base of users via small scaled, inter-organizational means. No matter if they are completely automated also. Thus, a consolidated approach consisting of five kinds of components, and based on the Public-Key Infrastructure is required [25]:

(i) Certification Authorities (CAs) for issuing & revoking of certificate.
(ii) Organizational Registration Authorities (ORAs) to ensure tying of certificate holder identities & attributes with public keys.
(iii) Certificate holders who are issued certificates and have the ability to sign digital documents and encrypt them.
(iv) Clients who can, based on public key of a trusted Certification Authority, validate digital signatures and their certification paths.
(v) Repositories having the ability of storing & retrieving the available certificate along with CRLs ('Certificate Revocation List').

A TSA (Time Stamping Authority) might also be included as part of Public-Key Infrastructure. Organizations that work as Certification Authorities, Registration Authorities, Repositories and Time Stamping Authorities are commonly called TTP (Trusted Third Parties) / CSP (Certification Service Providers).

Despite the fact that requirements from a PKI are quite dissimilar and have been recorded in numerous applications, a common ground can be found [26]. You may also find a broad list of services satisfying the requirements above in [27]. Subsequently, the list of functions required to perform these services can be defined in [27].

From the discussion above it seems that we know all the issues related to security in e-commerce and we have the technology available that can provide solutions to the issues in security of e-commerce. However, if that was the case then a critical question would be: Why do we still have breaches of Security when e-commerce transactions happen? The incidents should not be happening.

The most common problem is that, while everyone acknowledges the need for securing ecommerce, what they neglect is that security is much more than putting physical and electronic barriers. The most robust firewall and strongest encryptions are practically useless without organizational security, built around a policy that enunciates how these usages, maintenance and management of these tools is to be done. This policy is concerned to and neutral to technology as it is at a High-Level. It is aimed at setting procedures and directions, along with defining penalties & measures for noncompliance [28].

5. Conclusion

Despite Presence of laws dealing with on e-commerce Confidentiality, Reliability and security, agreements between countries are still missing. From a legal point of view it should not make a difference to the buyer or the vendor who are part of the electronic transaction, where the business, user or the intermediary service are situated. We must imitate an effort to encompass protection of all consumer rights, moral or legal, with a common understanding and agreement. Previously, law makers have fought against individual violations when they occurred – which resulted in a kind of patch work of different legal systems. But this helps only safeguard against violation of secluded characteristics of Security, Confidentiality and Reliability in Electronic commerce.

There must be an inherent support in E-commerce businesses for privacy platforms, third-party transaction service, security solutions and trust infrastructure. The Policy must clearly outline what are the applicable laws for which countries where the e-business is located. The policies should also have proper change management, i.e. keeping track, with dates, of all changes made to them. Consumers should choose the product and service more prudently on the basis of privacy & security statements and on the presence of certified features, like site authentication or seals of privacy. Thus increasing the acceptance of the seal and forces the e-business to comply with the statement published. Yet, we know for sure that the privacy enforced via organizational means don't really guarantee privacy of the individual. All these approaches only help in making the decision on which we can trust and rely on. It is the initial step only; we require technology to ensure preservation of confidentiality.

Current technologies have made significant breakthrough in preserving the Security, Confidentiality and reliability of electronic Commerce. However, a lot of research is still required for this to be performed without the user's intervention and with lesser participation of the trusted third parties. To conclude technologies that are better suited for the general security requirements, need to be developed. In the current scenario anonymity is considered to be a threat to the social or national security. We need to carry out more research to comprehend how we can achieve a balance between the two conflicting requirements and have a harmonious existence of both.

References

[1] http://forrester.com/ - Forrester Research. Post-web retail. Sept. 1999.
[2] http://www.census.gov/estats - US Census Bureau
[3] R. Clarke, "Internet Privacy Concerns Confirm the Case for Intervention", Comm. of the ACM. Vol. 42, No. 2, 1999.
[4] W. Chung, and J. Paynter, "Privacy Issues on the Internet", Proc of the 35th Hawaii Int. Conf. on System Sciences. Jan. 2002.
[5] M. Brown, and R. Muchira, "Investigating the relationship between Internet Privacy Concerns and Online Purchasing Behaviour", Journal of Electronic Commerce
[6] I. Araujo, "Privacy Mechanisms supporting the building of trust in e-commerce", Proc. IEEE International Workshop on Privacy Data Management, Tokyo, Japan, April 2005.
[7] M. K. Reiter, and A. D. Rubin, "Anonymous web transaction with Crowds" Comm. of the ACM. Vol. 24, No. 2, 1981.
[8] D. L. Chaum, "Untraceable electronic mail, return address, and digital pseudonyms", Comm. of the ACM. Vol. 24, No. 2, 1981.
[9] J. Lopez, R. Oppliger, and G. Pernul, "Authentication and Authorization Infrastructures (AAIs): A Comparative Survey", Computers & Security Journal. Elsevier (North Holand), Vol. 23, 2004.
[10] http://disi.unitn.it/locigno/didattica/PE/05-06/Support_ Lez01.pdf
[11] Anil Kini, and Joobin Choobineh, "Trust in Electronic Commerce: Definition and Theoretical Considerations", HICSS (4), 1998, 51-61.
[12] Wikipedia - http://en.wikipedia.org/wiki/Reliability
[13] T. Grandison, and M. Sloman, "A Survey of Trust in Internet Applications", IEEE Communications Surveys & Tutorials, 2000.
[14] "Information Technology - Open systems interconnection - The Directory: Authentication Framework", June 1997. - ITU-T Recommendation X.509

[15] "Information Technology - Open systems interconnection - The Directory: Public-key and attribute certificate frameworks", March 2000- ITU-T Recommendation X.509,

[16] M. Blaze, J. Feigenbaum, and J. Lacy, "Decentralized Trust Management", IEEE Symposium on Security and Privacy, pp.164-173, 1996.

[17] E. Kiountouzis, "Approaches to security of information systems", Information Systems Security, New Technologies Publications, Athens, (S. Katsikas, D. Gritzalis and S. Gritzalis Eds.), 2004

[18] T. Winograd and F. Flores, "Understanding Computers and Cognition", AddisonWesley, 1997.

[19] P. J. Denning, "Electronic Commerce - Internet Besieged", Addison-Wesley and ACM Press, (D. E. Denning & P. J. Denning Eds),1998.

[20] G. Pernul, A. Rohm and G. Herrmann, "Trust for Electronic Commerce Transactions", in Proceedings, ADBIS '99, Springer-Verlag, 1999.

[21] R. Burnett, "Legal aspects of e-commerce", Computing & Control Engineering Journal, 2001.

[22] E. Turban, "Electronic Commerce: A Managerial Perspective", Prentice Hall. 2004.-

[23] http://www.wipo.int/clea/en/index.jsp

[24] http://www.wipo.int/treaties/en

[25] A. Arsenault and S. Turner, "IETF PKIX WG, Internet draft, Internet X.509 Public Key Infrastructure PKIX Roadmap", March 10, 2000.

[26] D. Lekkas, S.K. Katsikas, D.D. Spinellis, P. Gladychev and A. Patel, "User Requirements of Trusted Third Parties in Europe", in Proceedings, User identification and Privacy Protection Joint IFIP WG 8.5 and WG 9.6 Working Conference, pp. 229-242, 1999.

[27] S. Gritzalis, S.K.Katsikas, D. Lekkas, K. Moulinos, and E. Polydorou, "Securing the electronic market: The KEYSTONE Public Key Infrastructure".

[28] S.K. Katsikas and S.A. Gritzalis, "A Best Practice Guide for Secure Electronic Commerce", Upgrade, Vol. III, no.6, December 2002. http://www.upgradecepis.org .

Author Biographies

Dr. Abdullah received his B.Sc. degree in Computer Science from University of Southern Mississippi in 1990, Master degree from MIS Department, Information Systems, University of Illinois in 1992 and PhD degree in Computer Science from Computer Science Department, George Washington University in 2003. He has a teaching and research experience of more than 20 Years and his research interests include Information Systems, Software Engineering and Artificial Intelligence. He has more than 20 publications to his credit. Presently he is Chairman/HoD, Information Systems Department, Faculty of Computing and Information Technology, and also in additional charge of Vice-Dean, College of Science and Technology, King Abdulaziz University, Jeddah, KSA.

Design of a Code using System Theoretic Approach

V. C. Kotak[1] and B. K. Lande[2]

[1] **Research Scholar, Veermata Jijabai Technological Institute, Matunga, Mumbai, INDIA**

[2] **School of Engineering & Technology, Himalayan University, Jolly Grant, Dehradun, INDIA**

Abstract

To design the coding and decoding schemes which can be easily implemented has been the focus of researchers. We propose here one such scheme which is based on the definition of controllability of linear discrete-time system. Given a linear controllable discrete-time system, it has been proposed here that a good code with $d_{min} = 2$ can be proposed. Such a code has been shown to satisfy Singleton as well as Plot kin bounds.

Keywords: *Discrete-time system, Controllability, Coset leaders, Decoders.*

1. Introduction

In an effort to find good codes for practical purpose, researchers have moved beyond block codes to other paramounts such as convolution codes, turbo codes, space-time codes, low-density parity check codes and even quantum codes.

All the coding and decoding schemes proposed by now suffer from limitations due to the complexity of the decoder; therefore, all efforts in this direction have been focused towards the design of coding and decoding schemes which could be easily implemented.

Consider a linear controllable discrete-time scalar control system

$$x(k+1) = Ax(k) + bu(k) \qquad (1)$$

Controllability of this system implies that it is always possible to design a control function consisting of a sequence of impulse functions that can steer all the states of a given system in (1) to the origin. Such a control strategy that consists of sequence of impulse functions cannot be designed for continuous-time linear controllable system. One technique of designing such a control strategy for continuous-time systems has been proposed [1]. It was therefore thought that by applying similar idea to discrete-time linear controllable systems, the states reach the origin by a suitable action of control. Such control functions have been proposed in [2]. It was then thought that if the same idea is applied to linear discrete time controllable system then all the states at the n^{th} instant of sampling generate a set of n linearly independent vectors. We propose this set as a kind of code and stud of the paper.

We first introduce the concept of steering the states to the origin for a system described in terms of a transfer function of a linear discrete time controllable system.

Let us consider equation (1) which is an n^{th} order difference equation,

$$y(k+n)+C_1y(k+n-1)+C_2y(k+n-2)+....+C_ny(k)=u(k) \qquad (2)$$

with the initial conditions,

$$\begin{bmatrix} y(0) & y(1) & - & - & - & y(n-1)\end{bmatrix}^T = \begin{bmatrix} d_0 & d_1 & - & - & - & d_{n-1}\end{bmatrix}^T \qquad (3)$$

and it is required that

$$\begin{bmatrix} y(n) & y(n+1) & - & - & - & y(2n-1)\end{bmatrix}^T = \begin{bmatrix} 0 & 0 & - & - & - & 0\end{bmatrix}^T \qquad (4)$$

In order to do so, let us assume that the solution of (2) can be given by

$$y(k) = \sum_{j=0}^{n-1} a_j \partial(k-j) \qquad (5)$$

Using the given initial conditions, we can work out the values of a_j , $j = 0,1,....,n-1$. This leads to,

$$a_0 = d_0, ..., a_{n-1} = d_{n-1} \qquad (6)$$

Looking upon (2) as an expression for u(k), and then substituting for all the difference terms of y(k) in terms of 'δ' functions using (5), we finally get an expression for u(k) , in terms of a sequence of impulse functions that will steer $y(0)$, $y(1)$, ..., $y(n-1)$ to the origin.

This idea can be extended to the discrete-time linear system represented in the state variable form for designing the code.

2. Problem Statement

Consider a system in (1) with the initial conditions

$\left[x(0),\ x(1),\ .....,\ x(n-1)\right]^T \triangleq \left[d_1,\ d_2,\ .....,\ d_n\right]^T$, of which the general solution can be given by,

$$x(n) = A^n x(0) + \sum_{j=0}^{n-1} A^j Bu(n-1-j) \qquad (7)$$

Let us assume that $x_1(k)$ which is the k^{th} component of the n-tuple vector x_i to be given by,

$$x_1(k) = c_1\delta(k) + c_2\delta(k-1) + ... + c_n\delta(k-n+1) \qquad (8)$$

where c_i are the constants to be determined using the initial conditions, note that k is a general variable and n denotes the dimension of the system.

Evaluating (8) at $k = 0$, we get $c_1 = x_1(0)$

Further,

$$x_1(k+1) = c_1\delta(k+1) + c_2\delta(k) + c_3\delta(k-1) + ... + c_n\delta(k-n+2) \triangleq x_2(k) \qquad (9)$$

And evaluating (9) at, $k = 0$ leads to $c_2 = x_2(0)$

continuing in the same way, we shall obtain

$$x_2(k+1) = c_1\delta(k+2) + c_2\delta(k+1) + c_3\delta(k) + ... + c_n\delta(k-n+3) \triangleq x_3(k) \qquad (10)$$

leading to, $c_3 = x_3(0)$ and finally,

$$x_{n-1}(k+1) = c_1\delta(k+n-1) + c_2\delta(k+n-2) + c_3\delta(k+n-3) + ... + c_n\delta(k) \triangleq x_n(k) \qquad (11)$$

and thus, $c_n = x_n(0)$ this results in

$$x_1(k) = d_1\delta(k) + d_2\delta(k-1) + ... + d_n\delta(k-n+1) \qquad (12)$$

Since (1) is controllable, we can express matrices A and b in phase variable canonical form, so that (1) can be represented as,

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ | \\ | \\ | \\ x_{n-1}(k+1) \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & - & 0 & 0 & 0 \\ 0 & 0 & 1 & - & 0 & 0 & 0 \\ | & & & | & & & | \\ | & & & | & & & | \\ 0 & 0 & 0 & - & 0 & 1 & 0 \\ 0 & 0 & 0 & - & 0 & 0 & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & - & -a_3 & -a_2 & -a_1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ | \\ | \\ | \\ x_{n-1}(k) \\ x_n(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ 0 \\ 1 \end{bmatrix} u(k) \qquad (13)$$

Thus,

$$x_1(k+1) = x_2(k)$$
$$x_2(k+1) = x_3(k)$$
$$\vdots$$
$$x_{n-1}(k+1) = x_n(k)$$

and

$$x_n(k+1) = -a_n x_1(k) - a_{n-1}x_2(k) - .... - a_1 x_n(k) + u(k) \qquad (14)$$

or

$$u(k) = x_n(k+1) + a_n x_1(k) + a_{n-1}x_2(k) + ... + a_1 x_n(k) \qquad (15)$$

treating this as the formula for u(k), we replace $x_1(k), x_2(k), ..., x_n(k)\ \&\ x_n(k+1)$

in terms of impulse functions $'\delta'$ by using the assumed (16). This finally leads us to

$$\begin{aligned} u(k) &= \left[d_1\partial(k+n) + d_2\partial(k+n-1) + ... + d_n\partial(k+1)\right] \\ &+ a_n\left[d_1\partial(k) + d_2\partial(k-1) + ... + d_n\partial(k-n+1)\right] \\ &+ a_{n-1}\left[d_1\partial(k+1) + d_2\partial(k) + ... + d_n\partial(k-n+2)\right] \\ &\qquad | \qquad\qquad | \qquad\qquad | \\ &\qquad | \qquad\qquad | \qquad\qquad | \\ &+ a_1\left[d_1\partial(k+n-1) + d_2\partial(k+n-2) + ... + d_n\partial(k)\right] \end{aligned}$$

thus we get,

$$\begin{aligned} &\therefore\ u(0) = a_n d_1 + a_{n-1}d_2 + a_{n-2}d_3 + ... + a_2 d_{n-1} + a_1 d_n \\ &\therefore\ u(1) = a_n d_2 + a_{n-1}d_3 + ... + a_3 d_{n-1} + a_2 d_n \\ &| \qquad | \qquad | \\ &| \qquad | \qquad | \\ &\therefore\ u(n-1) = a_n d_n \end{aligned} \qquad (16)$$

Now, for $x(1)$ we have

$$x(1) = Ax(0) + bu(0)$$

$$x(1) = \begin{bmatrix} 0 & 1 & 0 & - & 0 & 0 & 0 \\ 0 & 0 & 1 & - & 0 & 0 & 0 \\ | & & & | & & & | \\ | & & & | & & & | \\ 0 & 0 & 0 & - & 0 & 1 & 0 \\ 0 & 0 & 0 & - & 0 & 0 & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & - & -a_3 & -a_2 & -a_1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ | \\ | \\ | \\ d_{n-1} \\ d_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} a_n d_1 & a_{n-1}d_2 & \cdots & a_2 d_{n-1} & a_1 d_n \end{bmatrix}$$

$$x(1)=\begin{bmatrix} d_2 \\ d_3 \\ | \\ | \\ | \\ d_{n-1} \\ -a_n d_1 - a_{n-1}d_2 - .....- a_2 d_{n-1} - a_1 d_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ 0 \\ a_n d_1 + a_{n-1}d_2 + ...+ a_2 d_{n-1} + a_1 d_n \end{bmatrix}$$

$$x(1)=\begin{bmatrix} d_2 \\ d_3 \\ | \\ | \\ | \\ d_{n-1} \\ -a_n d_1 - a_{n-1}d_2 - .....- a_2 d_{n-1} - a_1 d_n + a_n d_1 + a_{n-1}d_2 + ...+ a_2 d_{n-1} + a_1 d_n \end{bmatrix}$$

$$x(1)=\begin{bmatrix} d_2 \\ d_3 \\ | \\ | \\ | \\ d_{n-1} \\ 0 \end{bmatrix} \qquad (17)$$

further for $x(2)$ we have,

$$x(2) \;=\; A^2 x(0) + bu(1) + Abu(0)$$

$$x(2)=\begin{bmatrix} 0 & 1 & 0 & - & 0 & 0 & 0 \\ 0 & 0 & 1 & - & 0 & 0 & 0 \\ | & & & | & & & | \\ | & & & | & & & | \\ 0 & 0 & 0 & - & 0 & 1 & 0 \\ 0 & 0 & 0 & - & 0 & 0 & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & - & -a_3 & -a_2 & -a_1 \end{bmatrix}^2 \begin{bmatrix} d_1 \\ d_2 \\ | \\ | \\ d_{n-2} \\ d_{n-1} \\ d_n \end{bmatrix}$$

$$+\begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} a_n d_2 & a_{n-1}d_3 & - & - & - & a_3 d_{n-1} & a_2 d_n \end{bmatrix}$$

$$+\begin{bmatrix} 0 & 1 & 0 & - & 0 & 0 & 0 \\ 0 & 0 & 1 & - & 0 & 0 & 0 \\ | & & & | & & & | \\ | & & & | & & & | \\ 0 & 0 & 0 & - & 0 & 1 & 0 \\ 0 & 0 & 0 & - & 0 & 0 & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & - & -a_3 & -a_2 & -a_1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} a_n d_1 & a_{n-1}d_2 & \cdots & a_2 d_{n-1} & a_1 d_n \end{bmatrix}$$

$$x(2)=\begin{bmatrix} 0 & 1 & 0 & - & - & - & 0 \\ 0 & 0 & 1 & - & - & - & 0 \\ | & & & | & & & | \\ | & & & | & & & | \\ | & & & | & & & | \\ -a_n & -a_{n-1} & -a_{n-2} & - & - & -a_2 & -a_1 \\ a_n a_1 & -a_n + a_1 a_{n-1} & -a_{n-1} + a_1 a_{n-2} & - & - & -a_3 + a_1 a_2 & -a_2 + a_1^2 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ | \\ | \\ | \\ d_{n-1} \\ d_n \end{bmatrix}$$

$$+\begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ 0 \\ a_n d_2 + a_{n-1}d_3 + ...+ a_3 d_{n-1} + a_2 d_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ 1 \\ -a_1 \end{bmatrix} \begin{bmatrix} a_n d_1 & a_{n-1}d_2 & \cdots & a_2 d_{n-1} & a_1 d_n \end{bmatrix}$$

$$+\begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ 1 \\ -a_1 \end{bmatrix} \begin{bmatrix} a_n d_1 & a_{n-1}d_2 & a_{n-2}d_3 & - & - & a_2 d_{n-1} & a_1 d_n \end{bmatrix}$$

$$x(2)=\begin{bmatrix} d_2 \\ d_3 \\ | \\ | \\ d_{n-2} \\ -a_n d_1 - a_{n-1}d_2 - a_{n-2}d_3 - ....- a_2 d_{n-1} - a_1 d_n \\ a_1 a_n d_1 - a_n d_2 + a_1 a_{n-1} d_2 - .....- a_3 d_{n-1} + a_1 a_2 d_{n-1} - a_2 d_n + a_1^2 d_n \end{bmatrix}$$

$$+\begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ 0 \\ a_n d_2 + a_{n-1}d_3 + ...+ a_3 d_{n-1} + a_2 d_n \end{bmatrix}$$

$$+\begin{bmatrix} 0 \\ 0 \\ | \\ | \\ | \\ a_n d_1 + a_{n-1}d_2 + a_{n-2}d_3 + ...+ a_2 d_{n-1} + a_1 d_n \\ -a_1 a_n d_1 - a_1 a_{n-1} d_2 - a_1 a_{n-2} d_3 - .....- a_1 a_2 d_{n-1} - a_1^2 d_n \end{bmatrix}$$

$$x(2)=\begin{bmatrix} d_2 \\ d_3 \\ | \\ | \\ d_{n-2} \\ -a_n d_1 - a_{n-1}d_2 - a_{n-2}d_3 - ... - a_2 d_{n-1} - a_1 d_n + a_n d_1 + a_{n-1}d_2 + a_{n-2}d_3 + ... + a_2 d_{n-1} + a_1 d_n \\ a_1 a_n d_1 - a_n d_2 + a_1 a_{n-1} d_2 - a_{n-1} d_3 + a_1 a_{n-2} d_3 - .. \\ \qquad + a_3 d_{n-1} + a_2 d_n - a_1 a_n d_1 - a_1 a_{n-1} d_2 - a_1 a_{n-2} d_3 - .... - a_1 a_2 d_{n-1} - a_1^2 d_n \end{bmatrix}$$

$$x(2)=\begin{bmatrix} d_2 \\ d_3 \\ | \\ | \\ d_{n-2} \\ 0 \\ 0 \end{bmatrix} \qquad (18)$$

Continuing the same process to obtain the values of $x(3),\ x(4), \ldots\ldots, x(n)$, we observe that the elements in the vector $x(k)\ \ where\ \ k = 0,1,\ldots\ldots,n-1$ go on shifting upwards and finally we get the vector,

$$x(n-1) = \begin{bmatrix} d_{n-1} \\ 0 \\ | \\ | \\ 0 \\ 0 \end{bmatrix} \tag{19}$$

and

$$x(n) = \begin{bmatrix} 0 \\ 0 \\ | \\ | \\ 0 \\ 0 \end{bmatrix} \tag{20}$$

which assures that the terminal state i.e., the state at the n^{th} instant of sampling has been steered to the origin.

In fact, taking any values for $x(0)$ and solving for $x(1),\ x(2), \ldots\ldots,\ x(n)$ will continue to display the same nature of $x(k),\ k = 0,1,\ldots.,n$ as described above.

However, with arbitrary values of $d_i\ \ i = 1,2,\ldots\ldots,n$ we require to use the modulo (2) operation on every element of the vectors $x(0),\ x(1),\ldots\ldots,\ x(n)$ in order to bring them in the binary form.

In order to avoid using modulo (2) operation, we can assume all the values of x (0) to be given by unity.

For example,

$$\begin{aligned}
&x(0) &&= [1\ \ 1\ \ -\ \ -\ \ -\ \ -\ \ -\ \ 1]^T \\
&then \\
&x(1) &&= [1\ \ 1\ \ -\ \ -\ \ -\ \ -\ \ 1\ \ 0]^T \\
&| \\
&| \\
&x(n-1) &&= [1\ \ 0\ \ -\ \ -\ \ -\ \ -\ \ -\ \ 0]^T \\
&x(n) &&= [0\ \ 0\ \ -\ \ -\ \ -\ \ -\ \ -\ \ 0]^T
\end{aligned} \tag{21}$$

3. Design of the Code

From (21) it can be observed that all the vectors, $x(0),\ x(1),\ldots\ldots,\ x(n-1)$ are linearly independent and hence let us consider them to form a generating set of vectors to form the code C, such that

$$G\ [I \quad P]$$

where P is the set of all the above linearly independent vector $x(0),\ x(1),\ldots\ldots,\ x(n-1)$, and I is correspondingly sized unit matrix.

Thus

$$G = \begin{bmatrix} 1 & 0 & 0 & - & - & 0 | d_1 & d_2 & - & - & - & d_k \\ 0 & 1 & 0 & - & - & 0 | d_2 & d_3 & - & - & d_k & 0 \\ & & & | & & | & & & | & & \\ & & & | & & | & & & | & & \\ 0 & 0 & - & - & 0 & 1 | 0 & 0 & - & - & - & 0 \end{bmatrix}$$

Since this is linear code, we obtain all the code words by taking the sum of all the rows with each other and finally appending a zero vector to this set of code words so that this set of code words forms a sub-space.

This, in general, gives rise to the linear (n,M,d) code given below :

$$\begin{aligned}
&\begin{cases} 0\ \ 0\ \ 0\ \ -\ \ -\ \ -\ \ -\ \ 0, \\ 1\ \ 0\ \ 0\ \ -\ \ -\ \ d_1\ \ d_2\ \ -\ \ -\ \ d_k, \end{cases} \\
&0\ \ 1\ \ 0\ \ -\ \ -\ \ d_2\ \ d_3\ \ -\ \ d_{k-1}\ \ 0, \\
&-\ \ -\ \ -\ \ -\ \ -\ \ -\ \ -\ \ -\ \ -\ \ -, \\
&0\ \ 0\ \ -\ \ -\ \ 0\ \ 1\ \ 1\ \ 0\ \ -\ \ 0, \\
&1\ \ 1\ \ 0\ \ -\ \ -\ \ 0\ \ d_1 + d_2\ \ d_2 + d_3\ \ -\ \ -\ \ d_{k-1} + d_k\ \ d_k, \\
&1\ \ 1\ \ 1\ \ 0\ \ -\ \ -\ \ 0\ \ d_1 + d_3\ \ d_2 + d_4\ \ -\ \ -\ \ d_{k-2} + d_k\ \ d_{k-1}\ \ d_k, \\
&-\ \ -\ \ -\ \ -\ \ -\ \ -, \\
&0\ \ 0\ \ -\ \ -\ \ 1\ \ 1\ \ d_k + 1\ \ 0\ \ -\ \ -\ \ -0\}
\end{aligned}$$

In the next section, we state and prove the main result of the paper, we establish that the total number of code words will be q^k, with $d_{\min} = 2$.

4. Main Result

Before we state the main result of the paper, we reproduce the following Lemma from [3] which is crucial in the proof of the main result.

Lemma 1:[3] *For a linear code, the minimum distance d^* satisfies*

$$d^* = \min_{c \neq 0} w(c) = w^*$$

where the minimum is over all code words except the all-zero codeword.

***Proof* :**

$$d^* = \min_{\substack{c_i, c_j \in \mathbb{C} \\ i \neq j}} d(c_i, c_j)$$

$$d^* = \min_{\substack{c_i, c_j \in \mathbb{C} \\ i \neq j}} d(0, \; c_i - c_j)$$

$$d^* = \min_{\substack{c \in \mathbb{C} \\ c \neq 0}} w(c)$$

$$d^* = w^* \qquad (7)$$

Theorem 1: Given a p^{th} order discrete time completely controllable linear system

$$x(k+1) = Ax(k) + bu(k)$$

there exists an (n, M, d) code over $\mathbb{F}_q$ with q = 2 such that

1. $n = 2p$
2. $M = 1 + p + \left[{}^pC_2 + {}^pC_3 + ... + {}^pC_p \right]$
3. $d_{\min} = 2$

***Proof* :**

Note that the set of code words is generated by the basis vectors $x(0), x(1), ..., x(n-1), x(n)$ which are obtained by solving the given linear discrete-time controllable system, hence it is a vector space.

1. This is obvious from the size of G.
2. Since this is a linear Code, total number of code words shall be given by
   M = (all zero code) + (all the p code vectors) + { [Sum of all the code vectors taken two at a time] + [Sum of all the code vectors taken three at a time] + ……. + [Sum of all the code vectors taken (p-1) at a time] +[Sum of all the p code vectors] }
   $$M = 1 + p + \left[{}^pC_2 + {}^pC_3 + ... + {}^pC_p \right]$$
   $$M = q^p$$
3. This is obvious from the proof of Lemma 1.

The set of codewords so designed exhibit, all the properties which are stated in the main result and summarized below :

Order of the system (p)	Total no. of Codewords (q^p)	Size of the Generator Matrix	Length of the Codewords ($2p$)
2	4	4 x 2	4
3	8	6 x 3	6
4	16	8 x 4	8
5	32	10 x 5	10

Table 1 : Summary of the Main Result

5. Decoding

The existing conventional methods of coset leaders are applicable to decode the code which we have been able to design. However, since we have proved that $d_{\min} = 2$ for the codes designed in this paper, we propose a decoding algorithm which successfully performs error correcting upto 1-bit.

An algorithm devised by us for Decoding the proposed code

1. Consider our code consists of m code words C_i, each of length n
2. For each code word C_i we generate a table consisting of n words which are at a distance of 1 from C_i by changing each of the n symbols at a time.
3. Then via bit by bit comparison we eliminate those C_i *'s* whose set of words at a distance 1 are the same or some of the words within their sets are same
4. Thus finally the total number of code words C_i within our code will be reduced, but then we are guaranteed of perfect decoding by our technique
5. Whenever a code word is received (say R_i) , we look up the word in each of these m tables via bit by bit comparison
6. If the received code word is found in a table of particular code word C_i, we conclude that the receive word R_i maps onto that particular code word C_i .

5. Conclusion

The design of a binary code proposed here is based on the definition of controllability of discrete-time system. It has been proved that it is an (n, M, d) code with, $M = 2^p$ and $d_{\min} = 2$. It has also been observed that, this code satisfies the Plotkin as well as the Singleton upper bounds. The methodology for the generation of the code is very simple. For the generation of the code it is sufficient to simply know the vector formed by a set of initial conditions and the parity check matrix will automatically be formed.

The set of codewords so designed exhibit, all the properties which are stated in the main result. The existing conventional decoding scheme of coset leaders is applicable to decode the code proposed here. Further to this, for the codes designed in this paper, we have also devised an algorithm to propose a methodology to decode this code, with an error-correcting capability of upto 1-bit, its minimum distance being $d_{\min} = 2$.

References

[1] Kalman R. E., 'On the General Theory of Optimal Control', Proc. First Intl. Congr. of the Fed. of Automat Contr., (IFAC) Moscow, Vol. 1, pp. 481-493, (1960)

[2] Agashe S. D., Lande B. K., 'A New Approach to State-transfer Problem, J. Franklin Inst., Vol. 333(B), No. 1, pp. 15-21, (1996)

[3] Blahut R. E., 'Theory and Practice of Error Control Codes', Addison-Wesley Publishing Company, pp. 46-47, (1984)

Reducing the Computational Cost in Multi-objective Evolutionary Algorithms by Filtering Worthless Individuals

Zahra Pourbahman[1], Ali Hamzeh[2]

[1] **Department of Electronic and Computer Engineering, Shiraz University, Shiraz, Iran**

[2] **Department of Electronic and Computer Engineering, Shiraz University, Shiraz, Iran**

Abstract

The large number of exact fitness function evaluations makes evolutionary algorithms to have computational cost (especially in Multi Objective Problems (MOPs)). In some real- world problems, reducing number of these evaluations is much more valuable even by increasing computational complexity and spending more time. To fulfil this target, we introduce an effective factor, in spite of applied factor in Adaptive Fuzzy Fitness Granulation NSGAII (AFFG_NSGAII) algorithm, to filter out worthless individuals more precisely. Our proposed approach is compared with respect to AFFG_NSGAII, using the Hypervolume (HV) and the Inverted Generational Distance (IGD) performance measures. The proposed method is applied to 1 traditional and 1 state-of-the-art benchmarks with considering 3 different dimensions. From an average performance view, the results indicate that although decreasing the number of fitness evaluations leads to have performance reduction but it is not tangible compared to what we gain.

Keywords: Multi objective evolutionary algorithm optimization, Fitness approximation, Information granulation, Pareto optimal solutions.

1. Introduction

Evolutionary algorithms (EAs) have a population-based procedure in which the population evolves repeatedly by applying some stochastic operators in order to generate better population members until a termination control criterion is met [1]. They seem to be one of promising optimizers among recent proposed optimization methods [2] since they have a number of unique features as follows. (i) EAs can be implemented simply [1], (ii) EAs can find multiple optimal solutions ideally while classical optimization methodologies can't find such solutions efficiently [3], and (iii) EAs perform the parallel search as a computationally quick procedure in contrast with classical optimization methodologies [1].

In a wide variety of real-world optimization problems EAs can be applied [1] because most of the time, practical problems have two or even more normally conflicting objectives which should be optimized simultaneously as MOPs in which a set of optimal solutions (effective solutions) needs to be obtained and EAs can find these effective solutions efficiently in a single run [2] whereas they use a population-based approach. This trend is known as Multi Objective Evolutionary Algorithms (MOEAs). Also, set of these effective solutions is known as Pareto optimal set and their vectors are called non-dominated. Non-dominated vectors are plotted in objective space and constituted the Pareto front. In MOPs, to assign a fitness value to an individual, all objectives should be evaluated. Therefore, when MOEAs are applied to a complex problem, computational cost can grow increasingly [4]. Also, it can be time-consuming. To handle these difficulties, fitness approximation can be integrated into MOEAs [5].

In our work, we aim to reduce the number of exact fitness function evaluations in one of the state-of-the-art proposed approaches for fitness approximation, which is called AFFG_NSGAII [6, 7] by introducing an effective factor to diagnose valuable individuals more logically.

The remainder of this paper proceeds as follows. Literature review is provided in Section 2. Section 3 expresses the contributions of the proposed approach. Section 4 and Section 5 present the experimental setup and the experimental results respectively. Section 6 is devoted to discussion. Finally, Conclusion of the paper and future directions are described in Section 7.

2. Related Work

In some real-world problems, metaheuristics like evolutionary algorithms are used to find a set of solutions over a unique run. The large number of exact fitness function evaluations makes such problems computationally prohibitive. The computational cost becomes more critical in MOPs since more objectives are involved. To deal with this difficulty, it is common to use approximation techniques, which are divided into three

levels, namely, problem approximation, functional approximation, and evolutionary approximation [8].
Problem approximation tries to substitute an easier computationally solvable problem for the original one. As an example, performance evaluation of turbine blade wind tunnel experiments, which is led to Euler equations, can be replaced with Computational Fluid Dynamics (CFD) simulation, which is led to Navier-Stokes equations [8, 9].
Functional approximation tries to estimate a model based on objective functions known as the fitness function in the evolutionary computation [8]. To approximate fitness function, surrogate-assisted evolutionary computation can be used [9]. In recent researches, an Aggregate Surrogate Model (ASM) for multi objective optimization is introduced [10, 11, 12]. This surrogate model is built based on the combination of One-Class Support Vector Machine (SVMs) to change (randomly) unsupervised population into supervised one in each generation and Support Vector Regression (SVR) to estimate fitness of each individual and provide a Pareto front at last. Since ASM is used for extrapolation, to have enough diversity in the search space, informed operators are applied. In the other hand, because there are some errors in the surrogate model, pre-children generated by informed operators are not filtered in the basis of ASM lonely. In [10, 11] it is done based on ASM gain with regard to the lowest distance to non-dominated solutions in each generation. In [12] offsprings are filtered in a greedy manner (The highest amount of {ASM (k) – ASM (z_k) | k is an offspring and z_k is the nearest neighbour of k}), but premature convergence forces to make a probabilistic selection besides by using a normal distribution. Another research on surrogate-assisted MOEAs was proposed in [13]. In this work a Pareto Rank Learning procedure is used to predict rank of each new offspring. To learn this surrogate model, a population evolves iteratively and the value of each new offspring is evaluated using original objective functions and then is archived into an embedded database. Archiving new offsprings continue until having enough training data. After that, the non-dominance sorting is applied over the archived solutions. Then, an ordinal SVM model is learned based on the sorted database. At last, the rank of each new offspring is predicted in terms of the learned model; if the model produces an output of rank 1 its fitness is evaluated and is archived in the database. This updated database is used for updating the model after each generation. Even though fitness function approximation models presented can reduce cost of solving problems with expensive objective functions, they have some general defects as follows. Since model is updated in each generation, computational burden rises. Moreover, the precision of the model is dependant to the primary training data. Additionally, complexity of model grows exponentially as the number of problem parameters increases [14].
Evolutionary approximation specifically is used in evolutionary algorithms. Fitness inheritance is a major class of this type of approximation, which was basically introduced in [15]. In this method, the number of fitness approximation in contrast with fitness evaluation is controlled based on *inheritance proportion* parameter. For an individual, its similarity to its parents and fitness of its parents are applied to form a weighted average formula for fitness approximation. Despite the simplicity of this method, since similarity of each individual to its parents is evaluated just in the decision space, its performance is not acceptable [14].
To address the above-mentioned difficulties (in functional and evolutionary approximation), a new method for fitness approximation based on information granulation was introduced by Davarynejad et al. [7] Called Adaptive Fuzzy Fitness Granulation NSGAII (AFFG_NSGAII). In this method, a pool of solutions is constituted in the objective space. Each member of the pool is called a fuzzy granule. Each fuzzy granule is a Gaussian Membership Function (GMF) where its center is an individual and its width is computed based on its fitness and some problem dependant parameters. But approximated fitness sometimes leads to have not sufficient precision in such calculations. So, the weighted rank of each member of the pool besides a problem dependant parameter which is the minimum width of GMFs is used instead [6]. Additionally, each fuzzy granule has a life index used in fitness approximation while it can control the computational complexity of the algorithm. In this method, fitness of each new individual, generated by an evolutionary algorithm (like the standard NSGAII) in the decision space, being approximated or evaluated explicitly is determined based on its maximum similarity to the granules of the pool. The maximum similarity is evaluated in terms of a predefined similarity metric, which is Gaussian similarity function. In this criterion, the computed width of each fuzzy granule is used as a parameter for controlling the degree of the similarity among a new individual and that fuzzy granule; if the maximum similarity of the new individual to the granules of the pool be lower than an adaptive threshold, its fitness is approximated by increasing the granule's life index. As a point, in [6] the threshold is considered fixed (0.9) to simplify their evaluations. In the other hand, in AFFG_NSGAII, the pool size is controlled in terms of the life index parameter in which the granule with the lowest life index is removed from the pool when its size becomes more than a predefined threshold.
Since AFFG_NSGAII is one of the state-of-the-art proposed approaches in fitness approximation, vast

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

researches over it was done recently. For example, as mentioned earlier, by the reason that the width of each granule is used in the similarity metric as an important factor, in [16] a fuzzy logic controller is embedded to propose a width for each fuzzy granule. Input of this controller is Number of Decision Variable (NDV), Maximum Range of Decision Variables (MRDV), and number of completed generations. As an application-based work, effectiveness of the granule-based fitness approximation on Spread Spectrum Watermarking was investigated in [17].

In some applicatory problems like simulation, reducing number of expensive objective function evaluations is taken into consideration. AFFG_NSGAII can be used to deal with this limitation as a recent and promising method for fitness approximation. However, in this method, exploring valuable solutions for fitness evaluations more accurately can have a considerable effect on reducing the number of exact fitness function evaluations without sacrificing its performance. We contribute to this area and acquire some achievements.

3. The Proposed Approach

As mentioned in the previous section, AFFG_NSGAII as an evolutionary algorithm integrated into fitness approximation was improved in [6], in which filtering worthless individuals for fitness approximation was performed in terms of a fixed threshold and an applied factor that is a similarity metric called Gaussian similarity function. It means that the value of each new offspring generated by the standard NSGAII is determined based on its maximum similarity to the granules of the pool which is computed by the similarity metric; if the maximum similarity of a new offspring to the pool be lower than a fixed threshold, it is considered as a valuable individual. So, it is added to the pool and its fitness is evaluated explicitly; otherwise, fitness of the most similar granule of the pool is assigned to the new offspring (fitness is approximated) and then granule's life index is increased.

However, we found that in such evolutionary process, there are some potential factors despite applied factor to filter out worthless individuals more precisely. Even if applying these factors leads to have an increase in the computational complexity but decreasing the computational cost in many applicatory problems, in which there are multiple expensive-to-evaluate objective functions, is much more significant under a limited resource budget. In our proposed approach, we introduce the most effective factor among the potential factors in spite of Gaussian similarity function, to identify more logically whether a new offspring is worthy enough for exactly evaluating its fitness. Before the introduction of the new effective factor, a preprocessing should be explained in the following.

In each generation, granules of the pool are ranked based on non-dominance sorting [18]. Then, Non-inferior solutions are considered as the Current Pareto Set.

Inspired by the fact that in most MOEAs the population is driven toward the best Pareto points [10], we introduce an influential factor in order to guide the search in the vicinity of the Current Pareto Set in each generation.

Suppose the phenotype of *jth* individual and the center of *lth* fuzzy granule in *ith* generation to be like

$$X_j^{(i)} = \left\{x_{j,1}^{(i)}, x_{j,2}^{(i)}, \ldots\ldots, x_{j,n}^{(i)}\right\},$$

$$C_l^{(i)} = \left\{c_{l,1}^{(i)}, c_{l,2}^{(i)}, \ldots\ldots, c_{l,n}^{(i)}\right\}$$

, respectively and *d* be considered as the dimension of each individual, the minimum distance between *jth* individual and *k'* elements of the Current Pareto Set in *ith* generation is computed based on Euclidean distance, as Eqn. (1).

$$D_{j,l}\left(X_j^{(i)}, C_l^{(i)}\right) = \sqrt{\sum_{n=1}^{d}\left(X_{j,n}^{(i)} - C_{l,n}^{(i)}\right)^2} \qquad (1)$$

$, where\ l = 1,2, \ldots, k'$

Now assume that the maximum similarity of the new offspring to the granules of the pool be lower than the determined threshold like before. To decide fitness of that offspring is either evaluated or approximated, the minimum distance of both the new offspring and its parents to the Current Pareto Set are computed by Eqn. (1); if the new offspring be closer to the Current Pareto Set compared with at least one of its parents, the new offspring is considered as a valuable individual, is added to the pool as a new fuzzy granule and its fitness is evaluated explicitly. We call our proposed approach as Modified_AFFG_NSGAII.

Even if applying our powerful factor leads to make algorithm more complicated but in some real-world problems like expensive simulation-based and mechanical design problems decreasing the computational cost is much more considerable even by increasing computational complexity and spending more time.

In Modified_AFFG_NSGAII, we use our influentially promising factor, which is the minimum distance to the Current Pareto Set despite applied factor, thereby deciding that fitness of a new offspring is either evaluated or

approximated. Consequently, the estimation of the proximity of solutions to the real Pareto set locally leads to have more precise selections of valuable individuals for fitness evaluations. In this way, we deal with the computational cost burden of such expensive problems by remarkably reducing the number of exact fitness function evaluations without having any tangible effect in the viewpoints of efficiency and efficacy.
In order to prove that our proposed approach is promising, 14 test problems are applied. Additionally, 2 well-known performance metrics are used for validation of our proposed approach.

4. The Experimental Setup

This section describes comprehensive assessments by means of two well-known performance metrics and adopting wide varieties of test problems to compare our results with respect to those obtained with a state-of-the-art algorithm for fitness approximation (AFFG_NSGAII) [6].

4.1 Performance Measures

In this section, we present 2 indicators, which are commonly used specially in MOEAs for evaluation of our proposed approach.

4.1.1 Hypervolume

For a minimization problem, the volume in the objective space covered by non-inferior solutions (N) is evaluated by this metric. The set of the worst values of objectives forms a vector as the reference set. As explained mathematically by Deb in [2], for each non-inferior solution, $s \epsilon N$, a hypercube, V_s, is constructed with a reference point, r. After all, Hypervolume is calculated based on the union of all hypercubes, as follow:

$$HV = Volume\left(U_{s=1}^{|N|} V_s\right) \quad (2)$$

To make it sensible, it is showed in Fig 1 [2].

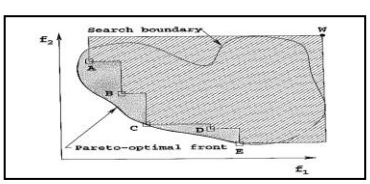


Fig.1 The Hypervolume enclosed by non-dominated solutions.

4.1.2 Inverted Generational Distance

A real Pareto front and a set of candidate solutions

$$PF = \{y_1, y_2, \ldots, y_N\},$$
$$F = \{X_1, X_2, \ldots, X_k\}$$

are given; the Inverted Generational Distance (IGD) is defined as follow:

$$IGD(F, PF) = \frac{1}{N}\left(\sum_{j=1}^{N} \acute{D}_j^t\right)^{\frac{1}{t}}, 1 \leq t \leq \infty \quad (3)$$

Where $\acute{D}_j$ is minimal Euclidean Distance from y_j to F [19].

4.2 Benchmarks

In this section, we present 1 traditional and 1 state-of-the-art benchmarks in order to perform comprehensive assessments of our proposed approach.

4.2.1 Congress on Evolutionary Computation 2009 (CEC09)

In the CEC 2009 algorithm competition, a set of bound constrained MOP test problems as UF family and a set of constrained test problems as CF family are suggested [20]. In our experiments we adopt 5 test problems from CF family, CF1 to CF5, and 4 test problems from UF family, UF1 to UF5 except UF4.

4.2.2 Zitzler-Deb-Thiele (ZDT)

As it was emerged in [21], ZDT family test problems have sufficient complexity to compare different types of multi objective optimizers. In our experiments, we adopt 5 test problems, ZDT1 to ZDT6 except ZDT5 as a binary problem.

5. The Experimental Results

Some parameter settings need to be performed in our experiments. The population size is set to 50. A set of new offsprings are generated by Simulated Binary Crossover (SBX) with probability of 0.9 and Polynomial Mutation (PM) with the probability of 1/L, where L is the number of decision variables. Distribution indices for crossover and mutation are taken from the literature (η_c = 20 and η_m = 20). Furthermore, binary tournament selection is applied. Tables 1, 2, 3 show amounts of mentioned design parameters per test problem.

Table 1: Utilized parameter values and reference points used for calculating I_H in ZDT family and their number of decision variables.

Problem	σ_{min}	N_G	Reference point	Number of decision variables
ZDT1	2^{-4}	100	[1.1,3.5]	6
ZDT2	2^{-5}	100	[1.1,5.0]	6
ZDT3	2^{-5}	100	[1.1,6.0]	6
ZDT4	2^{-6}	100	[1.1,140]	10
ZDT5	2^{-5}	100	[1.1,9.0]	10

Table 2: Utilized parameter values and reference points used for calculating I_H in CF family and their number of decision variables.

Problem	σ_{min}	N_G	Reference point	Number of decision variables
CF1	2^{-4}	100	[3,3]	10
CF2	2^{-4}	100	[8,7]	10
CF3	2^{-4}	100	[68,59]	10
CF4	2^{-4}	100	[18,19]	10
CF5	2^{-4}	100	[31,32]	10

Table 3: Utilized parameter values and reference points used for calculating I_H in UF family and their number of decision variables.

Problem	σ_{min}	N_G	Reference point	Number of decision variables
UF1	2^{-4}	100	[8,7]	30
UF2	2^{-4}	100	[6,5]	30
UF3	2^{-4}	100	[12,10]	30
UF5	2^{-4}	100	[18,15]	30

Additionally, all numerical results are the average of 30 independent runs, which are presented in Tables 4, 5, 6, 7, 8, and 9. These attainments are related to HV and IGD of Pareto front while ZDT, CF, and UF test problem families are applied, respectively and both AFFG_NSGAII and Modified_AFFG_NSGAII converge. Indeed, these results indicate that both methods approximately have the same performance.

Table 4: IGD of both AFFG_NSGAII and Modified_AFFG_NSGAII after convergence in ZDT family.

Problem	AFFG_NSGAII IGD	Modified_AFFG_NSGAII IGD
ZDT1	0.0364	0.0370
ZDT2	0.0235	0.0254
ZDT3	0.0284	0.0381
ZDT4	2.6393	3.2972
ZDT6	1.5948	1.7387

Table 5: HV of both AFFG_NSGAII and Modified_AFFG_NSGAII after convergence in ZDT family.

Problem	AFFG_NSGAII HV	Modified_AFFG_NSGAII HV
ZDT1	3.4501	3.4504
ZDT2	4.7906	4.7874
ZDT3	6.6715	6.6600
ZDT4	150.4215	149.7011
ZDT6	5.1082	5.1251

Table 6: IGD of both AFFG_NSGAII and Modified_AFFG_NSGAII after convergence in CF family.

Problem	AFFG_NSGAII IGD	Modified_AFFG_NSGAII IGD
CF1	0.6615	0.5578
CF2	0.2872	0.3018
CF3	1.0518	1.2223
CF4	0.2580	0.2981
CF5	1.4652	1.6357

Table 7: HV of both AFFG_NSGAII and Modified_AFFG_NSGAII after convergence in CF family.

Problem	AFFG_NSGAII HV	Modified_AFFG_NSGAII HV
CF1	8.4249	8.3584
CF2	53.8690	53.3946
CF3	3.9161e+003	3.8987e+003
CF4	334.5758	332.9370
CF5	928.7142	929.7592

Table 8: IGD of both AFFG_NSGAII and Modified_AFFG_NSGAII after convergence in UF family.

Problem	AFFG_NSGAII IGD	Modified_AFFG_NSGAII IGD
UF1	0.2389	0.3185
UF2	0.1134	0.1456
UF3	0.4946	0.6081
UF4	2.1404	2.2922

Table 9: HV of both AFFG_NSGAII and Modified_AFFG_NSGAII after convergence in UF family.

Problem	AFFG_NSGAII HV	Modified_AFFG_NSGAII HV
UF1	54.4400	53.5675
UF2	28.8462	28.7701
UF3	112.5421	110.7408
UF4	218.7533	217.6718

In the other hand, the average number of exact fitness function evaluations (of 30 independent runs) was plotted against the number of generations, which is determined in terms of the convergence time per test problem per algorithm. Derivative Figures (from Fig. 2 to Fig. 7) signify that Modified_AFFG_NSGAII reduces the computational cost considerably compared with AFFG_NSGAII.

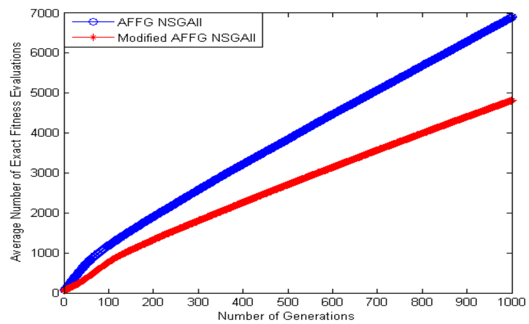


Fig. 2 Computational Cost Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over ZDT1 Problem.

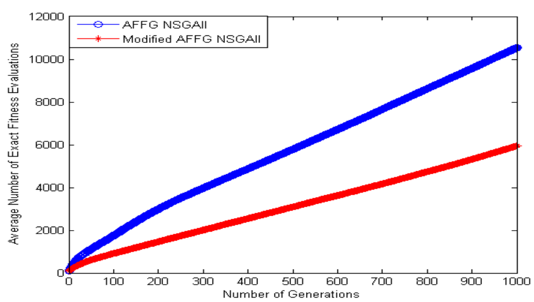


Fig. 3 Computational Cost Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over ZDT4 Problm

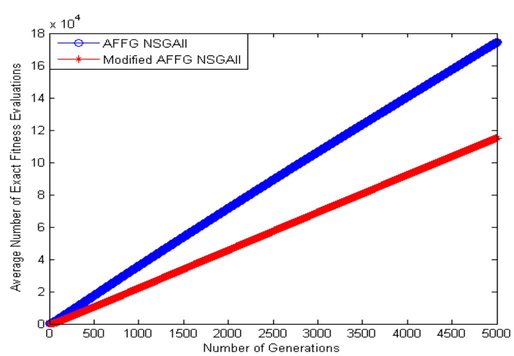


Fig. 4 Computational Cost Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF1 Problem.

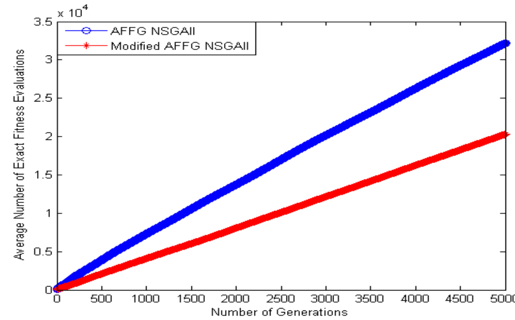


Fig 5. Computational Cost Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF3 Problem.

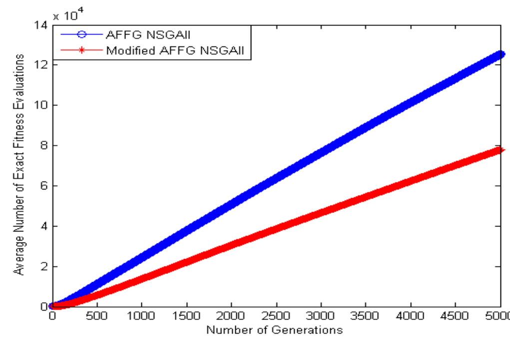


Fig.6 Computational Cost Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF1 Problem.

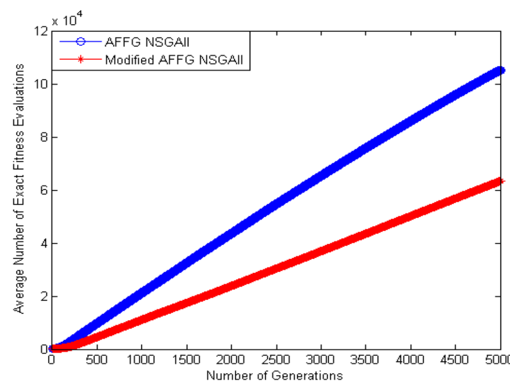


Fig.7 Computational Cost Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF3 Problem.

6. Discussion

As some state-of-the-art MOEAs integrated with fitness approximation in the literature, it is common to perform fitness approximation for some individuals besides fitness evaluation as usual for others. In some applicatory problems like simulation-based and mechanical design problems, there are expensive objective functions to evaluate. Therefore, contribution to this area has attracted more attention, recently. In this paper, we have contributed to this area in order to decrease the computational cost.

We believe that if we have had even lower individuals for fitness evaluations as usual in each generation but higher confidence about their qualities, termination control criterion would be met sooner. Therefore, there is a trade-off between the computational cost and the computational complexity to achieve this fidelity. To achieve the above target, inspired by the fact that in most MOEAs the population is driven toward the best Pareto points, we proposed an effective and powerful factor in order to guide the search in the vicinity of the Current Pareto Set in each generation. Derivative Figures (Fig. 8 to Fig. 19) prove that Modified_AFFG_NSGAII mostly outperforms AFFG_NSGAII in terms of HV and IGD metrics per adopted test problem.

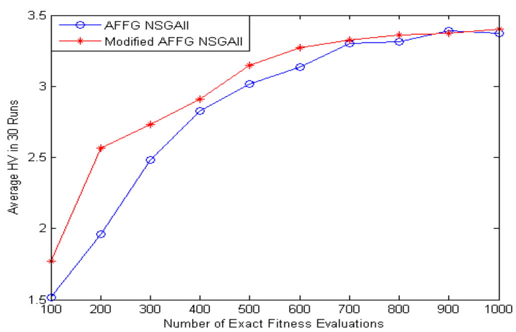


Fig.8 Performance (HV) Comparison of Modified_AFFG_NSGA2

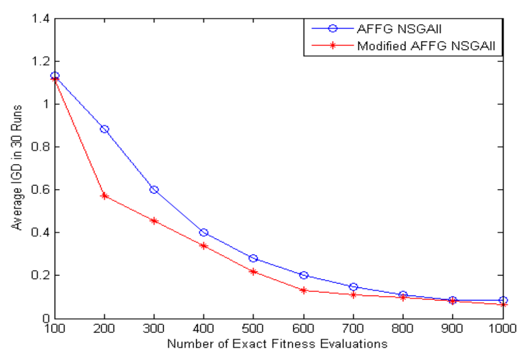


Fig.9 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over ZDT1 Problem.

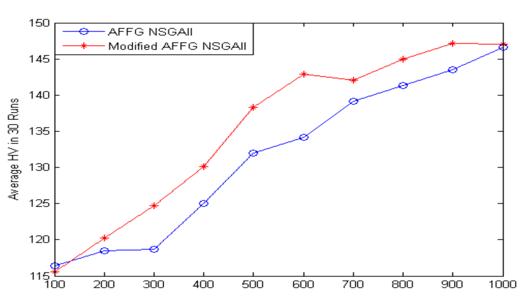


Fig.10 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over ZDT4 Problem.

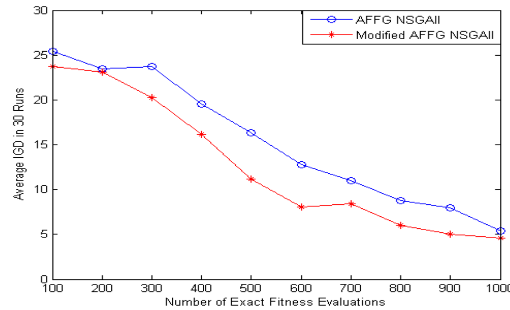


Fig.11 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over ZDT4 Problem.

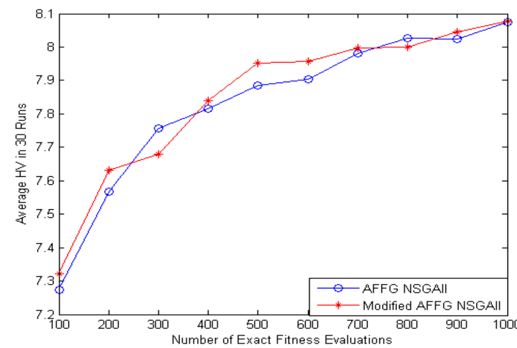


Fig.12 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF1 Problem.

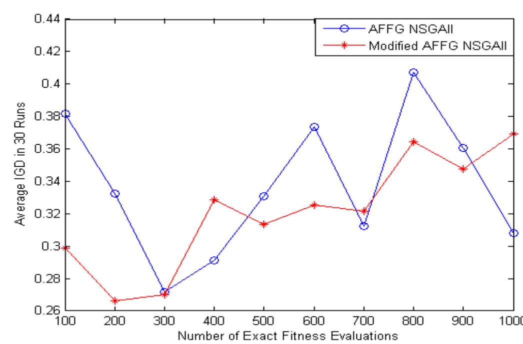


Fig.13 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF1 Problem.

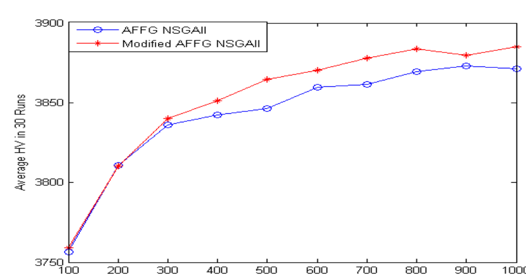


Fig.14 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF3 Problem.

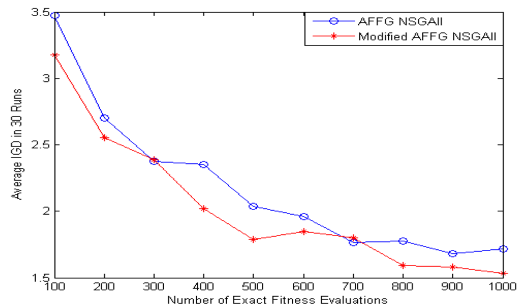


Fig.15 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF3 Problem.

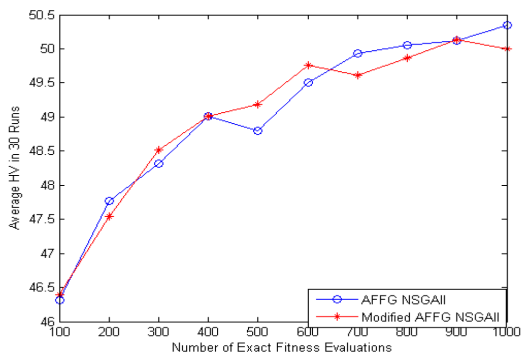


Fig.16 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF1 Problem.

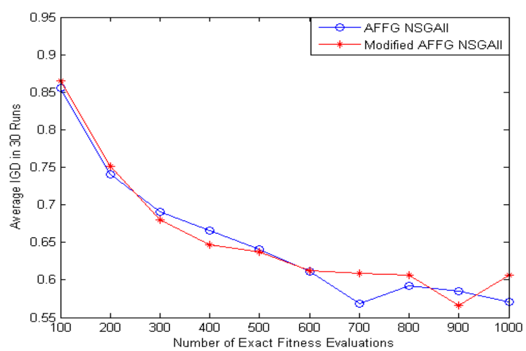


Fig.17 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF1 Problem.

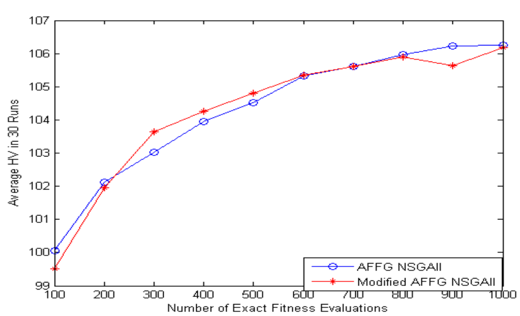


Fig.18 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF3 Problem.

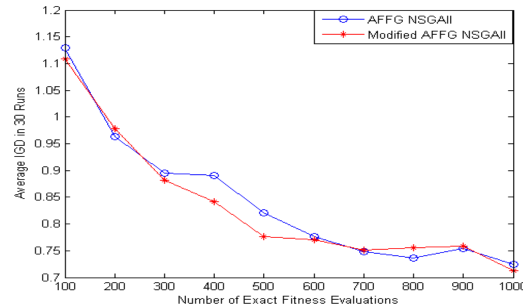


Fig.19 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF3 Problem.

As mentioned in Section 4, both methods were run until a fixed number of generations characterized in terms of the convergence time per test problem. To have a deep analysis of the proposed approach, the average HVs, and the average IGDs (of 30 independent runs), like the average number of exact fitness function evaluations in the previous section, were plotted separately against the determined number of generations. Derivative Figures in Section 4 (Fig. 1 to Fig. 7) and those are demonstrated in this section (Fig. 20 to Fig. 31) indicate that in our proposed approach the computational cost remarkably decreases while the convergence speed reduces. Fortunately, reduction in the convergence speed is negligible in comparison with the amount of decreasing the computational cost per test problem.

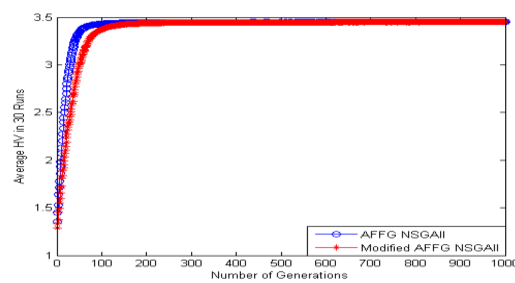


Fig.20 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over ZDT1 Problem.

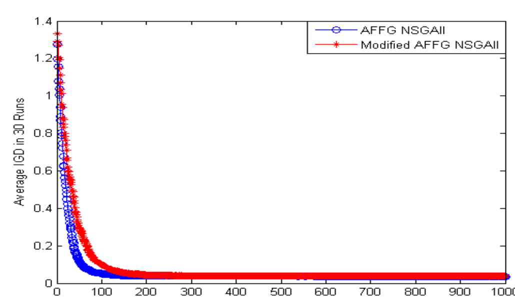


Fig.21 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over ZDT1 Problem.

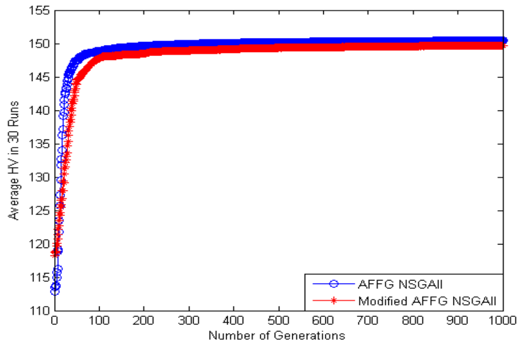


Fig.22 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over ZDT4 Problem.

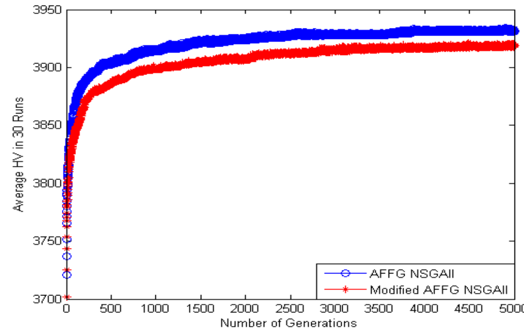


Fig.26 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF3 Problem.

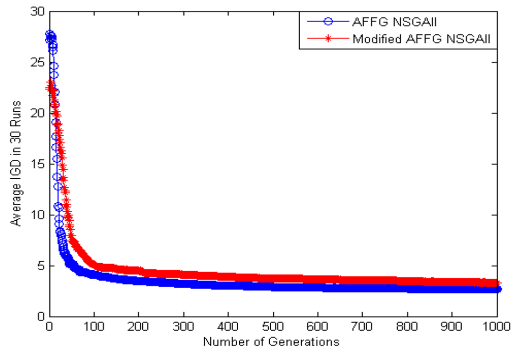


Fig.23 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over ZDT4 Problem.

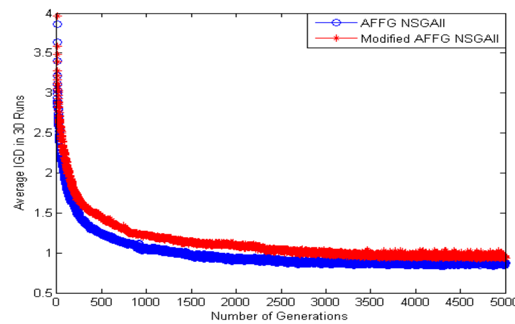


Fig.27 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF3 Problem.

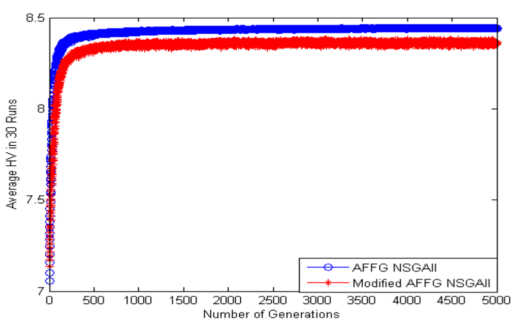


Fig.24 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF1 Problem.

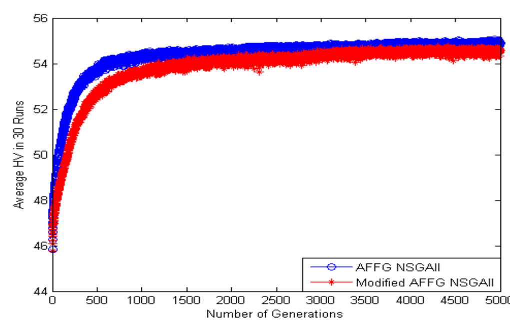


Fig.28 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF1 Problem.

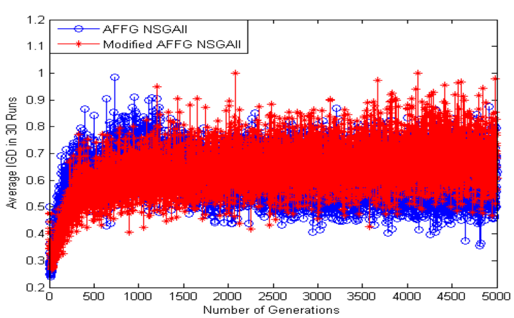


Fig.25 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over CF1 Problem.

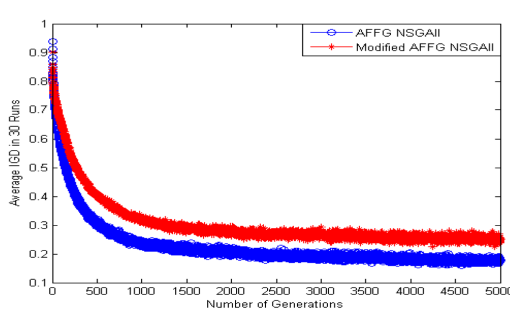


Fig.29 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF1 Problem.

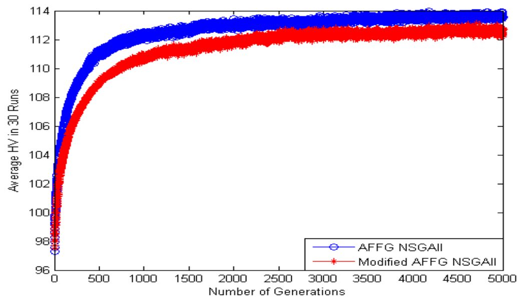


Fig.30 Performance (HV) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF3 Problem.

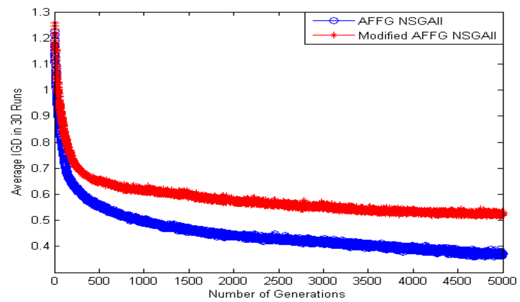


Fig.31 Performance (IGD) Comparison of Modified_AFFG_NSGA2 and AFFG_NSGA2 over UF3 Problem.

To find a better understanding of the usefulness of the proposed approach, numerical results are provided. In Tables 10, 11, and 12 the comparative amounts of the area under both AFFG_NSGAII and the Modified_AFFG_NSGAII curves (correspond to the Figures 44 to 71) are presented.

Table 10: The area under both AFFG_NSGAII and Modified_AFFG_NSGAII curves in ZDT family correspond to Figs 20 to 23.

Problem	Average IGD – Number of Fitness Evaluations AFFG - Modified AFFG	Average HV – Number of Fitness Evaluations AFFG - Modified AFFG
ZDT1	331.1715 - 260.0680	2587.1 – 2726.9
ZDT2	701.3847 - 557.1235	3386.3 – 3582.6
ZDT3	563.4949 - 481.2454	4834.4 – 4998
ZDT4	13893 - 11228	118390 – 122180
ZDT6	5802 – 5535	1375.8 - 1600.4

Table 11: The area under both AFFG_NSGAII and Modified_AFFG_NSGAII curves in CF family correspond to Figs 24 to 27.

Problem	Average IGD – Number of Fitness Evaluations AFFG - Modified AFFG	Average HV – Number of Fitness Evaluations AFFG - Modified AFFG
CF1	302.5250 - 287.1000	7063.3 - 7080.3
CF2	523.86 - 514.0750	45962 - 45822
CF3	1924.6 – 1792.2	3461400 - 3470000
CF4	736.4810 - 714.8476	291500 - 291630
CF5	2139 – 2120.6	817870 - 815850

Table 12: The area under both AFFG_NSGAII and Modified_AFFG_NSGAII curves in UF family correspond to Figs 28 to 31.

Problem	Average IGD – Number of Fitness Evaluations AFFG –Modified AFFG	Average HV – Number of Fitness Evaluations AFFG – New Modified AFFG
UF1	1.1369e+003 - 1.5205e+003	2.6903e+0052.7172e+005 -
UF2	569.7237 - 695.9495	1.4502e+005 - 1.4418e+005
UF3	2.2725e+003 - 2.9205e+003	5.6256e+005 - 5.5642e+005
UF5	1.2000e+004 - 1.2880e+004	1.3309e+006 - 1.3243e+006

Moreover, to be more understandable, Tables 13, 14, and 15 signify the percentage of differences of those comparative amounts. In particular, for some test problems such as CF4 the computational cost improved to more than 43% while the convergence speed reduced to less than 1%.

Table 13: The percentage of differences of the area under the AFFG_NSGAII and the Modified_AFFG_NSGAII curves in ZDT family correspond to Figs 20 to 23 and Figs 2 to 3.

Problem	HV – Number of Generations	Evaluations – Number of Generations
ZDT1	0.9376	30.1030
ZDT2	0.3101	36.8103
ZDT3	0.7941	33.0024
ZDT4	0.7585	46.0977
ZDT6	4.6630	38.4838

Table 14: The percentage of differences of the area under the AFFG_NSGAII and the Modified_AFFG_NSGAII curves in CF family correspond to Figs 24 to 27 and Figs 4 to 5.

Problem	HV – Number of Generations	Evaluations – Number of Generations
CF1	0.9642	35.1581
CF2	0.8242	30.6539
CF3	0.3839	39.4853
CF4	0.3092	43.6179
CF6	0.4056	26.4578

Table 15: The percentage of differences of the area under the AFFG_NSGAII and the Modified_AFFG_NSGAII curves in UF family correspond to Figs 28 to 31 and Figs 6 to 7.

Problem	HV – Number of Generations	Evaluations – Number of Generations
UF1	0.9906	39.2636
UF2	0.5766	41.2803
UF3	1.0910	43.0236
UF4	0.4973	24.3047
UF6	0.9906	39.2636

As it was showed in Tables 1, 2, and 3 in Section 4, the number of decision variables is considered 6, 10, and 30 for "ZDT1 to ZDT3", "ZDT4, ZDT5, and CF1 to CF5", and "UF1 to UF3 and UF5", respectively to have further investigation. The results, from first to end, illustrated that increasing individuals dimension has greater negative impact on Modified_AFFG_NSGAII form the viewpoint of both efficiency and efficacy. To explore the reason, we found the following observations. First of all, we suppose that the maximum similarity of a new offspring to the pool be more than 0.9 (predefined threshold); if the number of decision variables be set 100 in one time and 10 in the other time, the ratio of dissimilarity (between the new offspring and the granule whose similarity to it is more than predefined threshold) of the first case to the second one is 10 to 1. So, fitness is approximated with a lower accuracy in the first case. Furthermore, according to what was explained before, a greater number of fitness approximations are performed by Modefied_AFFG_NSGAII rather than AFFG_NSGAII until evolution control criterion is met. Therefore, decreasing both speed and accuracy is more tangible in our proposed approach rather than AFFG_NSGAII while the number of decision variables is increased.

7. Conclusion and Future Directions

In this study, we have introduced an effective factor for fitness approximation inspired from information granulation that affirmatively impress on reducing cost of MOEAs optimization. Our comprehensive experiments illustrate that the proposed approach is promising.

As a future work, we can explore some extra factors to find valuable individuals more and more precisely. Also, our proposed approach can be employed in many objective problems.

References

[1] K. Deb, "Multi-objective Optimization Using Evolutionary Algorithms: An Introduction," Technical Report 2011003, Indian Institute of Technology Kanpur, 2011.

[2] K. Deb, "Multi-Objective Optimization Using Evolutionary Algorithms," Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons, Chichester, 2001.

[3] K. Deb, "Multi-objective Evolutionary Optimization: Past, Present and Future," *Proceedings of the Fourth International Conference on Adaptive Computing in Design and Manufacture (ACDM'2000), UK, London, pp. 225–236, 2000.*

[4] G. Rohling, "Multiple Objective Evolutionary Algorithms for Independent for Independent, Computationally Expensive Objective Evaluations," PhD. Thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, Georgia(Atlanta), USA, 2004.

[5] J. E. Rodríguez, A. L. Medaglia, and C. A. Coello Coello, "Design of a Motorcycle Frame Using Neuroacceleration Strategies in MOEAs," Journal of Heuristics, Vol. 15, No. 2, 2009, pp. 177-196.

[6] M. Davarynejad, J. Rezaei, J. Vrancken, J. Berg, and C. A. Coello Coello, "Accelerating Convergence Towards the Optimal ParetoFront," Proceeding of IEEE Congress on Evolutionary Computation, , 2011, pp. 2107 – 2114.

[7] M. Davarynejad, M. R. Akbarzadeh-T, and N. Pariz, "A Novel General Framework for Evolutionary Optimization: Adaptive Fuzzy Fitness Granulation," Proceeding of IEEE Congress on Evolutionary Computing, 2007, pp. 951 – 956.

[8] Y. Jin, "A Comprehensive Survey of Fitness Approximation in Evolutionary Computation," Soft Computation, Vol. 9, No. 1, 2005, pp. 3–12.

[9] Y. Jin, "Surrogate-Assisted Evolutionary Computation: Recent Advances and Future Challenges," Swarm and Evolutionary Computation, Vol. 1, No. 2, 2011, pp. 61–70.

[10] I. Loshchilov, M. Schoenauer, and M. Sebag, "A Mono Surrogate for Multiobjective Optimization," Proceedings of the Twelfth Annual Conference on Genetic and Evolutionary Computation, 2010, pp. 471-478.

[11] I. Loshchilov, M. Schoenauer, and M. Sebag, "A Pareto-Compliant Surrogate Approach for Multiobjective Optimization," Proceedings of the Twelfth Annual Conference on Genetic and Evolutionary Computation, 2010 Vol. pp. 1979-1982.

[12] I. Loshchilov, M. Schoenauer, and M. Sebag, "Dominance-Based Pareto-Surrogate for Multi-Objective Optimization," Proceedings of the 8th international conference on Simulated evolution and learning (SEAL), USA, pp. 230-239, 2010.

[13] C.W. Seah, Y. S. Ong, I. W. Tsang, S. Jiang, "Pareto Rank Learning in Multi-objective Evolutionary Algorithms," In Proceedings of the IEEE Congress on Evolutionary Computation, 2012, pp. 1-8.

[14] M. Davarynejad, "Adaptive Fuzzy Fitness Granulation in Evolutionary Algorithms for Complex Optimization," M.S. Thesis, Department Electrical Engineering-Control Program, Ferdowsi University of Mashhad, Iran, 2007.

[15] R.E. Smith, B. A. Dike, and S. A. Stegmann, "Fitness Inheritance in Genetic Algorithms," Proceedings of the ACM symposium on Applied computing (SAC), 1995, PP. 345-350.

[16] M. Davarynejad, M. R. Akbarzadeh-T, and C. A. Coello Coello, "Auto-Tuning Fuzzy Granulation for Evolutionary Optimization," Proceedings of IEEE International Conference on Evolutionary Computation (CEC), IEEE Service Center, 2007, pp. 951–956.

[17] M. Davarynejad, C. W. Ahn, J. Vrancken, J. van den Berg, and C.A. Coello Coello, "Evolutionary hidden information detection by granulation-based fitness approximation," Soft Computing, Vol. 10, No. 3, 2010, pp. 719-729.

[18] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," IEEE Transactions on Evolutionary Computation, Vol. 6, No. 2, 2002, pp. 182–197.

[19] O. Schuetze, X. Esquivel, A. Lara, and C. A. Coello Coello, "Some Comments on GD and IGD and Relations to the Hausdorff Distance n," In proceeding of Genetic and Evolutionary Computation Conference, 2010, pp.1971-1974.

[20] Q. Zhang, A. Zhou, S. Zhao, P. N. Suganthany, W. Liu, and S. Tiwari, "Multiobjective optimization Test Instances for the CEC 2009 Special Session and Competition," Proceeding

of IEEE Congress on Evolutionary Computing, 2009, pp.18-21.

[21] E. Zitzler, K. Deb, and L. Thiele, "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results," Evolutionary Computation, Vol. 8, No. 2, 2000, pp. 173–195.

Z. Pourbahman is the MSc student in artificial intelligence of CSE and IT Department of Shiraz University. As one of her research interests, she focuses on evolutionary algorithms specially fitness approximation domain. Also, she is co-author of several articles in Machine Translation area especially in Natural Language Processing domain. She was programmer team leader in Telecommunication Company of Iran in 2008.

A. Hamzeh received his Ph.D. in artificial intelligence from Iran University of Science and Technology (IUST) in 2007. Since then, he has been working as assistant professor in CSE and IT Department of Shiraz University. There, he is one of the founders of local CERT center which serves as the security and protection service provider in its local area. As one of his research interests, he recently focuses on cryptography and steganography area and works as a team leader in CERT center to develop and break steganography method, especially in image spatial domain. Also, he works as one of team leaders of Soft Computing group of shiraz university working on bio-inspired optimization algorithms. He is co-author of several articles in security and optimization.

Compression of an AVI Video File Using Fractal System

Nevart A. Minas , Faten H. MohammedSediq

Computer Systems, Technical Institute,
Kirkuk, Iraq

Abstract

Compression of video files has become a necessity and very important because of transmission and storage of uncompressed video would be extremely costly and impractical.In this paper, the theory of fractal image compression is used to compress an Audio/ Video Interleaved file (AVI) Video file. The problem that the typical fractal compression approaches always discussed, is the long encoding time. A fast fractal image compression (FFIC) approach is used based on centralized moment descriptors to encode the video file images with low encoding time and high quality.

The first stage is opening the AVIvideo file and split its signal into separated images (frames) of type bitmap (BMP). In the second stage the digital image or frame of digital video is transformed from Red-Green-Blue(RGB) color space to the YC_bC_r color space image which is based on luminance (L) and chrominance (C) values.In third stage, each of the YC_bC_r components (y, C_b, and C_r) is compressed alone using FFIC and the fractal codes are savedgenerating a highly scalable layered bit stream that can be decoded at different qualities in terms of spatial resolution. The decoded images can be displayed at arbitrary resolution, with a high compression ratio. Furthermore, the algorithm is designed to require only a minimal coding delay. In the decoding stage all used compression techniques are executed in inverse sequence to obtain the decompressed image of video file and then reconstruct the AVI video file.

Keywords: *Video Compression, Video Coding, Fractal Image Compression, IFS, Centralized Moment.*

1. Introduction

Digital video has become very important form of information technology and is now used in many different areas, such as board casting, teleconferencing, mobile telephone, surveillance, and entertainment. People now expect to be able to access video through a wide range of different devices and over various networks.

Compression refers to the process of reducing the number of bits required to represent the image and video comes in two forms lossless and lossy. The lossless compression is a process to reduce image or video data for storage and transmission while retaining the quality of original image (i.e. the decoded image quality is required to be identical to image quality prior encoding. In lossy compression, on the other hand, some information present in the original image or video is discarded so that the original raw representation of image or video can only be approximately reconstructed from the compressed representation with high compression ratio. For more compression can be achieved with approximate quality to source image lossy compression is almost usually used than lossless compression to compress digital video. In general, in video compression the video sequences contain significant amount of statistical and subjective redundancy within and between frames. The ultimate goal of video source coding is the bit rate reduction for storage and transmission by exploring both spatial and temporal redundancy and to encode "minimum set" of information using entropy coding techniques .This is usually results in compression of coded video data compared to original source data. These statistical and subjective redundancy that can be exploited for compression are called spatial and temporal redundancy.[1]

The main idea of video compression is to exploit redundancies that are present in the video. There are two compression standards present that have been developed for low bit-rate applications. These are the H.263 [2] and MPEG-4 [3] video compression standards. Both standards utilizes wavelet transform based techniques. Wavelet transform techniques combine both transform and sub band coding. The other compression technique that has the potential for high compression is the fractal coding technique. Fractal compression differs from the standard transform coder methods. They were created as a result of the study of iterated function systems (IFS). Fractal methods store images as contraction maps of which the image is the fixed point. The decoding procedure in which the image is recovered by iterating the maps to its fixed point is simple. However, the recovered image suffers from the tiling effect, especially at low bit-rates[4].

Fractal Image Compression (FIC) has been an area of intensive research. Despite the diverse advantages offered by fractal compression, such as high decompression speed, high bit rate and resolution independence, the greatest disadvantage is the high computational cost of the coding phase, which means fractal coding cannot compete with other techniques (wavelets, etc.) in [5, 6]. During more than two decades of the development in FIC, a lot of research has been done to improve the performance. Generally, the attempts to speed up the fractal encoding consist of modifying the following aspects: the composition of the domain pool, the type of search used in block matching, or the representation/quantization of the transform parameters. Apart from these attempts focusing on the coding speed, some hybrid coders, such as wavelet-based and DCT-based fractal encoders, have been developed in [7, 8]. Various FIC side applications now are also explored in many other fields such as image database indexing and even face recognition [9].

2. Digital Video Coding

A digital video sequence is a collection of pictures, also called frames, spaced at fixed time intervals. In a color video sequence, each frame consists of three components, which can be either red-green-blue (RGB primaries) or luminance and two chrominance (YC_bC_r format) components. The luminance (Y) component is a monochrome image containing the structural information of the frame. The two chrominance (C_b and C_r) components contain color hue and saturation information of the frame. RGB format is used in displaying. RGB color space can be converted into YC_bC_r format as in equation (1).

$$\begin{pmatrix} Y \\ C_b \\ C_r \end{pmatrix} = (A) \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 16 \\ 128 \\ 128 \end{pmatrix} \qquad (1)$$

$$\text{where , } A = \begin{pmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.347 \\ 0.439 & -0.368 & -0.071 \end{pmatrix}$$

Three types of chrominance sampling formats relative to the luminance are used in video coding. These are labeled as 4:4:4, 4:2:2 and 4:2:0. In 4:4:4 formats; the same sampling grid for all three components is used. In 4:2:2 formats; the chrominance is sampled 2:1 horizontally but not vertically. The 4:2:0 formats have the chrominance sampled 2:1 both horizontally and vertically. Each component of a frame is a two-dimensional (2-D) signal, which can be represented by a matrix. The elements of the frame matrix are called as pixels. Therefore, a video sequence can be considered as a three-dimensional (3-D) signal for each spectral (color) component.

Digital video coding or compression is concerned with reducing the number of data storing units (bps) used to represent given information content in a video sequence. In addition to inter pixel spatial redundancy within a frame, video sequences contain high temporal (inter frame) redundancy, which is usually exploited in video coding algorithms by coding some frames using motion compensated prediction with reference to previously coded frames [10].

3. AVI File Format

The Audio Video Interleave (AVI) file is a file that is essentially a multimedia container format that was first pioneered by Microsoft to show video content for the Windows Operating System. This file allows both video and audio to play together, as well as making possible the concept of multiple streaming of video with sound. In general, AVI files tend to use file format extensions that were created in February of 1996 by Matrox Open DML group. These extensions, also commonly known by the name AVI 2.0, are completely supported by Microsoft.

The AVI file format is based on the RIFF (Resource Interchange File Format) which works by dividing all the raw data in one file into what are known as 'chunks' of data, where each chunk has its own 4 character tag for identification purposes. In any RIFF formatted file, an AVI file is just one chunk of the total data. This AVI file (one chunk) is further sub-divided into two obligatory chunks and one elective chunk of data.

Here's how the AVI file (chunk) works: The 1st sub-chunk holds integral information such as the metadata regarding the video and its characteristics: frame rate, height and width. This chunk is known as the file header, who identification is the four-character string hdrl. The 2nd sub-chunk, identification stream movi, is the one that consists of the actual data that comprises of the movie. The third sub-chunk, which is optional, is used to index all the chunks within the AVI file[11].

4. Fractal Block Coder

Fractal compression techniques are based on IFS, in which the image is described by sets of equations that provide contractive mappings. The first effective fractal coder was introduced by Jacquin [4], who noticed that a part of an image is similar to another part of the image. This coder makes use of this local self-similarity for compression, by forming images from properly transformed copies of parts of itself. This system requires the image to be divided into smaller, non-overlapping blocks called range blocks. Thereafter, larger domain blocks are constructed from the same image. The idea of this coder is to find an affine transformation of a domain block that closely matches each range block.

Two issues need to be considered in the design of fractal based compression systems. The first is the size and shape of range and domain blocks. Since range blocks are the attractors, the types of affine transformations become limited, since it depends on the size and shape of the range block. The simplest partitioning scheme is to divide the image into non-overlapping square range blocks of a fixed size. The domain blocks are normally twice the size of the range blocks and can overlap. Results reveal that small block sizes yields small compression ratios and a large block size, a large compression ratio. However, the clarity of the recovered image suffers with large block sizes. The downfall of the fixed size partition is that the image is partitioned without considering the contents of the image. There are regions of the image that will be covered well using small range block sizes and similarly, there are regions that could be covered well with larger range blocks. This will increase the compression ratio and maintain the clarity. This observation leads to the use of the variable block size partitioning technique.

The second matter to consider is the type of affine transformations to perform on domain blocks. It is required that these transforms be contractive in order for a fixed point to be reached in the decoding stage. The transforms are normally flip operations such as: horizontal, vertical, and diagonal flip; or rotation operations such as: 90^o, 180^o, and 270^o, and rotation. Also the identity block forms one of the transformations [12].

PIFS image encoder consists of a set of transforms applied on the regions of the image (i.e., range blocks). The transforms are, firstly, used to generate the overlapped domain regions. Secondly, a set of spatial contractive affine transforms are used to approximate the image range blocks by linearly mapping the most similar domain block. For a range block with pixel values (r0, r1,....,rm-1), and the domain block (d0,d1,....,dm-1), the contractive affine approximation is:

$$r_i' = sd_i + o \tag{2}$$

Where

s (scale) and o (offset) are the affine transform coefficients,

r's are the approximate (constructed) range values.

The affine transform is changed to become:

$$r_i' = s(d_i - \bar{d}) + \bar{r} \tag{3}$$

Where $\bar{r} = \frac{1}{m}\sum_{i=0}^{m-1} r_i$,

$$\bar{d} = \frac{1}{m}\sum_{i=0}^{m-1} d_i \tag{4}$$

Objects are represented as a collection of pixels in an image. Thus, for the purpose of recognition the properties of groups of pixels are needed to be determined. The description is often just a set of numbers (i.e., the object's descriptors). Moments describe the shape's layout (i.e., the arrangement of its pixels), a bit like area. Compactness and irregularity order descriptions.

Moments describe the shape's layout (i.e., the arrangement of its pixels), a bit like area. The calculation of moment invariants for any shape requires knowledge about both the shape boundary and its interior region. The moments used to construct the moment invariants are defined to be continuous but for practical implementation they are computed in the discrete form. Given a function f(x,y), the regular moments are defined as follows[13]:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \tag{5}$$

m_{pq} is a two-dimensional moment of the function f(x,y), the order of the moment is (p+q), where p and q are both integer numbers. The coordinates of the centre of gravity of the image are calculated using the following equations[13]:

$$x = \frac{M_{10}}{M_{00}}, \quad y = \frac{M01}{M_{00}} \tag{6}$$

The central moments can be defined in their discrete representation as:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x,y) \tag{7}$$

For an image block f(x, y) the central moment of order (p+q), around the block's central point (xc,yc), is defined as:

$$M(p,q) = \sum_x \sum_y (x - x_c)^p (y - y_c)^q f(x,y) \tag{8}$$

When this definition is applied to determine the nth order central moments of the zero mean range and domain image blocks, we get:

$$M_d(n,0) = \sum_{i=0}^{m=1} (x_i - L_c)^n (d_i - \bar{d}) \tag{9}$$

$$M_d(0,n) = \sum_{i=0}^{m=1} (x_i - L_c)^n (d_i - \bar{d}) \tag{10}$$

$$M_r(n,0) = \sum_{i=0}^{m=1} (x_i - L_c)^n (r_i - \bar{r}) \tag{11}$$

$$M_r(0,n) = \sum_{i=0}^{m=1} (y_i - L_c)^n (r_i - \bar{r}) \tag{12}$$

Where, $L_c = \frac{L-1}{2}$, (13)

L is the block width of the image

m is the number of block elements

(x_i, y_i) are the x and y coordinates of ith elements (pixel).

From the pair of nth moments {i.e., $M(n,0)$ & $M(0,n)$ }the following moments blocks descriptors could defined[13]:

$$R_n = \frac{M^2(0,n) - M^2(n,0)}{M^2(0,n) + M^2(n,0)} \tag{14}$$

Combining equation (3) with equations (9-13), and substitute the result in equation (14) we can easily prove that:

$$R_n^d = R_n^r \tag{15}$$

Where, R_n^d is the descriptor value of the domain block, R_n^r is the range block descriptor value

The above implies that "if any two blocks (from range and domain) satisfy the contractive affine transform, then their moments-based descriptor values should have similar values ($R_n^d = R_n^r$) whatever their isometric state. This does not mean that any two blocks have similar R factors are necessarily similar to each other" .

The main disadvantage of classical affine transform scheme is the greedy search in domain pool which is time consuming. In this paper some improvements have been made on the searching scheme to be selective instead of exhaustive. The improvements aimed to speed up the affine transform coding drastically without causing degradation in secret image quality. To make the searching process selective two moments based descriptor have been used to index (i.e., classify) the cover and secret blocks listed in domain and range pools, respectively. As mentioned before, the first introduced block index parameter is used to describe the isometric state of the block, while the second parameter is used to classify the blocks into categories.

Decompression is simpler and much faster than the coding process. Using the stored transform data, an iterative process commences, until the final image is reached [13].

The number of blocks partitioned relates directly to the compression ratio and PSNR. The number of blocks is related on the compression ratio and PSNR. As the number of blocks increases, the higher the PSNR and the lower the compression ratio, and vice versa. The peak signal-to-noise ratio (PSNR) per color channel is defined as:

$$\text{per} - \text{channel} - \text{PSNR} = 10 \log \frac{255^2}{\text{MSE}} \text{dB}, \quad (16)$$

$$\text{Where} \quad \text{MSE} = \left(\frac{1}{\text{nxn}}\right) \sum_{i=1}^{n} \sum_{j=1}^{n} \left(x_{ij} - x^{\hat{}ij}\right)^2 x_{ij} \ ,$$

xij is the value of a specific color channel in the original image,
x̂ij the decompressed one [14].

The first widely used signal coding technique was Pulse Code Modulation (PCM),consisting of the independent digitization and coding of signal samples. Limited compressionis achievable, however, since inter-sample dependence is completely ignored. The dependence between pixels may be taken into account by coding the prediction errorfor each pixel, using a prediction based on the values of previously encountered pixels in each scan line.Linear prediction provides a practical alternative only requiring knowledge of the autocorrelation function of the scan lines. The coding of the resulting linear prediction errors is known as Differential PCM (DPCM) [15]. In this paper, DPCM encoding is applied to the fractal coefficients (scale and average) values for each frame.

The compression performance is computed by the following equation [16]:

Compression-Rate=(Size after compression)/
(Size before compression) (17)

Compression-Ratio=100(1 − Compression Rate) (18)

5. The Proposed System

The system consists of two stages: encoding and decoding. In the decoding stagethefirst step is to open AVI structure and get the sequence of frames / images, each image then will be compressed using the FFIC separately.In FFIC the image is converted from RGB to YC_bC_r color space. The fractal domain blocks for each component (Y, C_b, and C_r), are obtained by down sampling the image by 2. Differentparameters like jump step and block size are used in the matching process between the range and domain blocks. The range and domain blocks can have any length, not like in the classical FIC systems where the block width and height must be divisible by the block length. To solve this problem, a band resizing operation is applied, it is done by inserting some additional lines and/or columns to the images in order to make the width and height of band as multiple of block length. The difference in length is managed at the encoding and decoding stages using image re-sampling method.

By using the centralized moment descriptor the domain blocks are sorted to speed up the search operation to find the best matching block among the domain blocks. The moment descriptor is also used to determine the suitable rotation state for each matching, so the fractal image compression becomes very fast and effective. After encoding each component, the fractal codes are quantized and saved into a file.

The following algorithm shows the encoding steps of the AVI video file (see Fig. 1):

Input: AVI video file
Output: Compressed file.

Step 1: read the AVI file.
Step2: split the AVI file into a series of frames (bitmap images (BMP)).
Step 3: for each frame do the following:
Step 4: read the color image (frame) and convert it to YC_bC_r color space.
Step 5: apply FFIC to compress each component (Y, C_b, and C_r)
Step 6: fractal code is quantized, DPCM is applied for the components (Y, C_b, and C_r)
Step 7: save fractal codes for the components (Y, C_b, and C_r) into the compressed file.
Step 8: goto step 3

End compression.

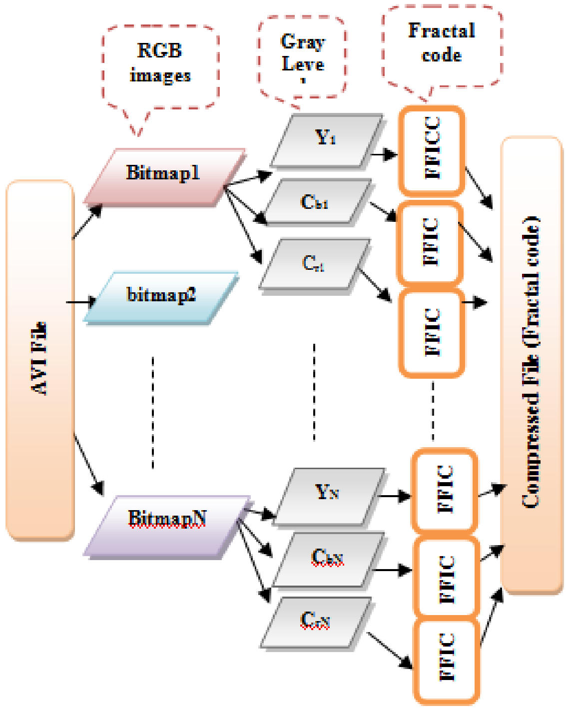


Fig. 1: Encoding flowchart of AVI file (Compression)

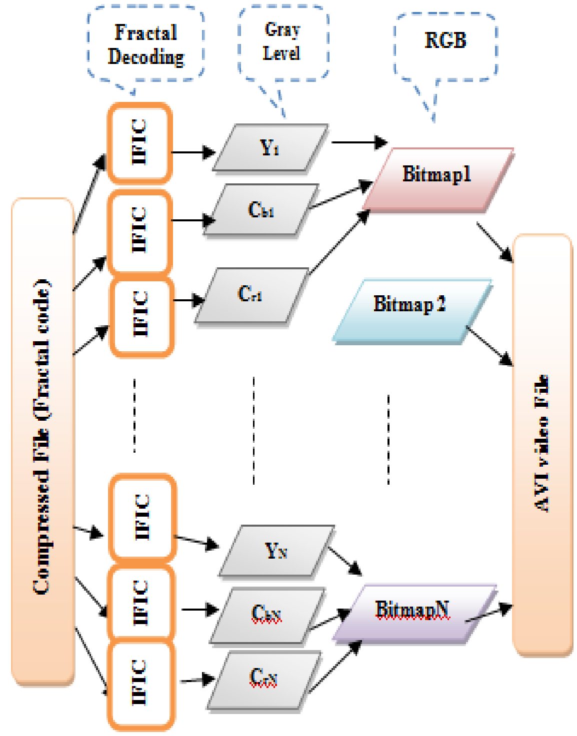


Fig. 2: Decoding steps of the compressed AVI file (Decompression)

The decoding stage in the reverse steps are performed. The fractal codes for each image (with its three components) are loaded from the compressed file and decoded using fractal inverse coding and then converted to the RGB color space. Then the resulted images (frames) are combined together in an AVI video file. The decoding steps are performed as listed in the following algorithm (Fig. 2): The following algorithm shows the decoding steps of the AVI video file:

Input: Compressed video file.
Output: AVI video file

Step 1: read the fractal code for each image (frame) components (i.e. Y, C_b, and C_r) from the compressed file in sequence.
Step 2: for each image component (fractal code) do the following:
Step 3: perform the inverse fractal image compression (IFIC) to each component and retransform them from YC_bC_r to RGB color space images (bitmap images).
Step 4: combine the series of bitmap images into AVI video file decompression program.

End decoding

6. Experimental Results:

The compression algorithm is conducted on the three AVI videos (Bfly, Dance and Cat) which are selected to be different in details and number of frames. The videos runs at (30) frames/second.

The algorithm was tested and performance parameters (encoding time, average PSNR, size before and after compression, compression ratio and rate) were registered. The algorithm used to compress the AVI video many times using different fractal parameters like block size as shown in table (3) and different jump step as shown in table (4). The highlighted rows in these tables represents the best results after decompressing the videos which proved that the three videos have a good quality, good compression ratio and low encoding time when the jump step is 2 and the block length is 5.

Fig. (4) shows that the encoding time was decreased with the increasing of the block size for all videos, the same happens with the jump step in Fig. (5), but making these parameters higher than the suggested values will affect the quality of the video as shown in Fig. s (6-7). The compression ratio (CR) was increased when the block size and the jump step were decreased (Fig. s 8-9), but making these parameters lower than the suggested values will make the compression operation slower.

Table 3: The effect of jump step on the average testing results for the video images

AVI	Jump step	Total E-time (sec)	Total D-Time (sec)	Total size after compr. (MB)	Average PSNR	Average CR	Average Rate
Bfly	1	168.7	30.2	3.4	25.8	14.2	1.68
	2	160.2	33.22	3.35	26	14.5	1.5
	3	154	32.99	3.25	24.9	14.7	1.4
Dance	1	44.85	32.07	2.57	38.5	18.3	1.31
	2	36.78	32.32	2.31	35.1	20.3	1.7
	3	34.36	31.68	2.2	35.2	21.3	1.12
Cat	1	160.0	24.76	2.25	35.3	14.7	1.62
	2	70.3	24.21	3.1	34.4	15.9	1.5
	3	43.6	24.77	1.97	34.2	16.7	1.43

Table 4: The effect of block size on the average results for the three video images

AVI	Block size	Total-E-time (sec)	Total-D-time (sec)	Average size(MB)	Average PSNR	Average CR	Average Rate
Bfly	4	176.86	30.91	5.18	28.1	9.22	2.62
	5	160.2	33.22	3.35	26	14.99	.5
	6	179.4	31.96	2.29	24.22	20.77	1.15
Dance	4	35.46	30.69	3.63	34.75	13	1.8
	5	37.61	32.23	2.31	33.12	20.38	1.7
	6	32.04	33.61	2.2	31.77	28.99	0.83
Cat	4	57.63	24.91	3.25	36.22	10.17	2.35
	5	64.38	24.21	3.1	34.4	15.93	1.5
	6	73.53	25.42	1.47	33.13	22.52	1.06

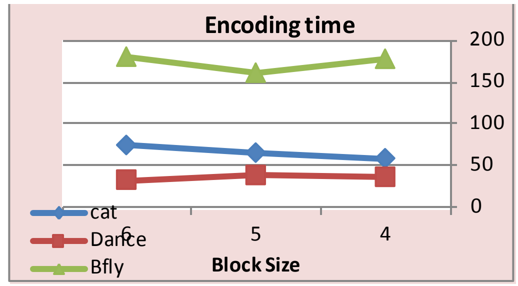


Fig. 4: The effect of block size on the Encoding time

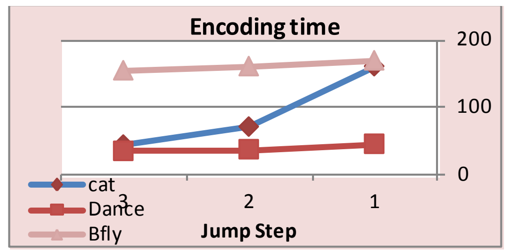


Fig. 5: The effect of Jump step on the Encoding time

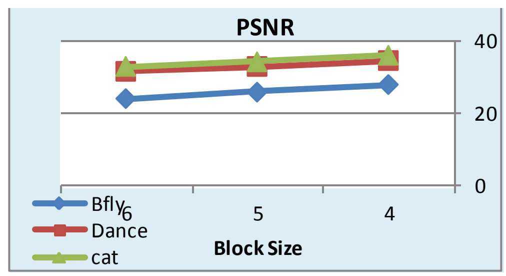


Fig 6: The effect of block size on the PSNR

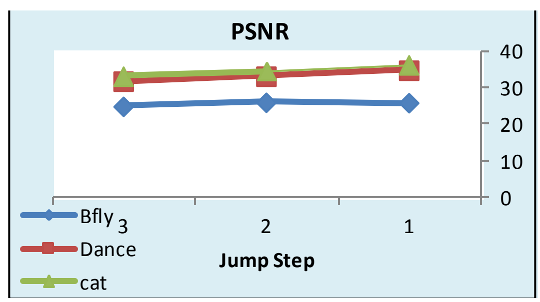


Fig. 7 : The effect of Jump step on the PSNR

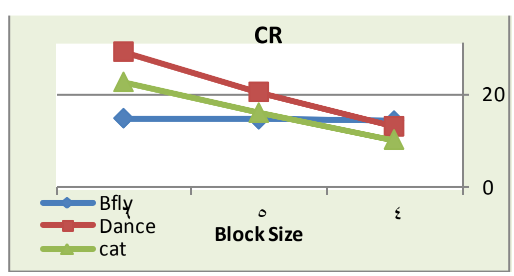


Fig. 8: The effect of block size on the CR

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

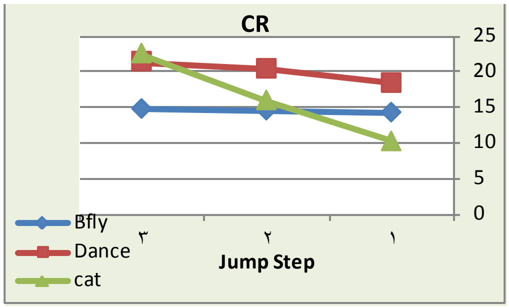


Fig. 9: The effect of Jump step on the CR

The Total results for the whole AVI video frames for all videos are computed and registered in (table 5), which are the encoding time, the average PSNR computed by the equation (16) , the best block size selected from table (3), jump step selected from table (4), the file size to determine the compression ratio that computed from the equations (17 and 18), and sizes of these videos before and after compression.

Table 5: The total tested results for all of the three AVI videos using best block size and jump step

AVI	No-of-Frames	Total E-time (sec)	Av. PSNR	Size before Comp. (MB)	Size after Comp. (MB)	Total CR	Rate
Butt	150	160.2	26	47.4	3.35	92.932	0.071
Man	162	36.78	35.1	476	2.31	95.147	0.049
Cat	150	70.38	34.4	33.1	3.1	93.716	0.063

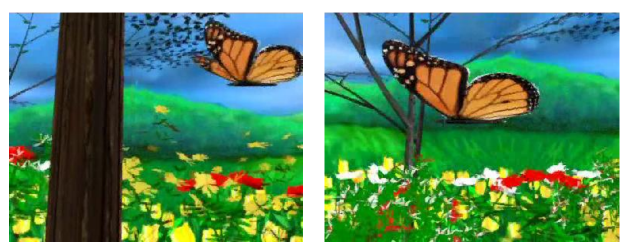
A comparison was done between the proposed research and other researches that used the video compression using fractal image compression in table (60). The comparison shows that the proposed research have the best compression ratio with good quality and rate. Although the encoding time did not mentioned in some of these researches, it is an important parameter used to evaluate any fractal compression method because of the low compression speed it suffers from. The proposed research proved low encoding time with low rate in all tested videos comparing them with other researches.

Table 6: Comparison table with others related researches

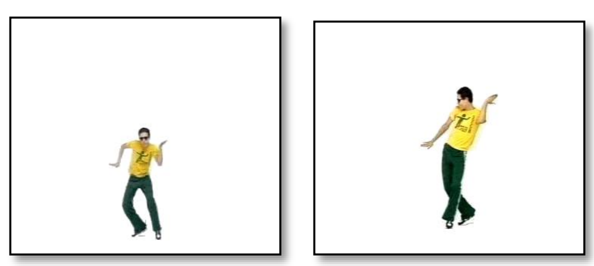
Research	No-of-Frames	Cr	Av. PSNR	Time (sec)	Rate (bit/pixel)
[2]	--	74.39	33	--	0.056
[8]	150	--	33.51	--	--
[9]	--	50	40	--	--
[17]	25	--	37	>309.7	8.46
Proposed (Bfly)	150	92.93	26	160.2	0.071
Proposed (Dance)	162	95.147	35.1	36.78	0.049
Proposed (Cat)	150	93.72	34	70.38	0.063

Considering the samples of the reconstructed videos frames in Fig. (3), it is obvious that the reconstructed AVI videos have a very good quality and FFIC is a perfect compression method to be use.

(a) Frames 40 and 140 of the reconstructed "Cat"



(b) Frames 20 and 75 of the reconstructed "Bfly"



(c) Frames 20 and 130 of the reconstructed "Dance".
Fig. 3: Selected frames for each of the three reconstructed videos

7. CONCLUSIONS

Fractal image compression always suffer from long encoding time, but in our research, this problem is managed using FFIC which is a fast method that speeded up the compression by using the moment descriptors. It is an effective way to compress AVI images when comparing it with the traditional and other fractal video compression methods.

The investigation and performance analysis shows the advantages of the proposed compression method using the FFIC on the three AVI videos, and the resulted values in all above tables and graphs proves that reconstructed videos have high PSNR, also the CR in all tested videos was about (92-94)% of the actual size of the original video for tested videos. The selected block size was 5 with jump step 2, and these parameters are very important to get good quality, CR and low encoding time . To have more compression and faster encoding time, these parameters can be increased, but this will affect the quality of the video.

8. REFERENCES

[1] Nabeel H., Tawfiq A. and Wafaa H., "Video Clip Image Compression Using DCT Technique" , Information technology, University of Babylon, Iraq, 2011.

[2] M. S. Lazar, L. T. Bruton, “Fractal Block Coding of Digital Video”, IEEE Trans. Circuits and Systems for Video Technology, vol.4, no.3, pp. 297-308, June 1994.

[3] A. Jacquin, “Image coding based on a fractal theory of iterated contractive image transformations”, IEEE Trans. Image Processing, vol. 1, no. 1, pp.18-30, Jan. 1992.

[4] Y. Brijmohan and S. H. Mnene, "Video Compression for Very Low Bit-Rate Communications Using Fractal and Wavelet Techniques ", e-book browse, No 19, 2010.

[5] Y. Fisher, "Fractal Image Compression: Theoryand Application", Springer-Verlag, New York, Inc., 1995.

[6] B. Wohlberg and G. d. Jager,"A Review of the Fractal Image Coding Literature", IEEE Transaction on Image Processing, Vol. No. 12, pp. 1999.

[7] G. Melnikov and A. K. Katsaggelos, "A Jointly Optimal Fractal/DCT Compression Scheme", IEEE Trans. on Multimedia, vol. 4, No.4, pp.413-422, 2002.

[8] G. M. Davis, "A Wavelet-Based Analysis of Fractal Image Compression", IEEE Transaction Image Processing, USA, Vol. No. 2, pp. 1998.

[9] H. Wang, Q. Wu, X. He and T. Hintz, "Preliminary Research On Fractal Video Compression On Spiral Architecture", IEEE, Proc. IPCV , pp.557-562, 2006.

[10] Abhayaratne G. , "Lossless and Nearly Lossless Digital Video Coding", for the degree of Doctor of Philosophy of the University of Bath, 2002.

[11] "File Extension Database", VLC media player for Windows, http://file.ms/extension/avi/.

[12] James D. Murray and William V. Ryper, "Encyclopedia of Graphics File Formats", Second Edition, O'Reilly & Associates, Inc, ISBN: 1-56592-161-5, 1996.

[13] Loay E. George, Suad K. Ahmad, “Hiding Image in Image Using Iterated Function System (IFS)”, European Conference of Computer Science (ECCS '10), Image, Sound and Signal Processing, ISBN: 978-960-474-250-9, 2010.

[14]"Encyclopedia of File Format", http://netghost.narod.ru/gff/graphics/book/ch10_03.htm.

[15] Wohlberg B., "Fractal Image Compression and the Self-Affinity Assumption: A Stochastic Signal Modeling Perspective", Ph.D thesis, University of Cape Town, 2003.

[16] Mustafa M. , "Using Shift Number Coding with Wavelet Transform for Image Compression", UK Journal of Information and Computing Science, England, Vol. 4, No. 3, pp. 311-320, ISSN , 746-7659, 2009.

[17] Mukhopadhyay, J. and Ghosh, S.K., "Low bit rate video compression using relative fractal coding" , Proc. of EUSIPCO, France,Vol.II, 125-128, 2002.

Faten H. MohammedSediq Assistant professor in the Faculty of Technical Institutes/ Technical Institute / Kirkuk-Iraq, Graduated in Mathematical Department /College of Science-University of Mosul with the degree of B.Sc., In 1987 graduated from the Computer Science in the Computer National Center-Iraq with the degree of Msc., Ph.D. in computer sciences, Technology University-Iraq in 2006; Research Interests: Detection and Addition in 3D Object Animation, Image Processing.

Nevart A. Minas Assistant teacher in the Faculty of Technical Institutes/ Technical Institute of Kirkuk - Iraq, B.Sc. in Computer Science in Mosul University-Iraq in 1989, M.Sc in Image Processing , collage of Science, University of Sulaimania -Iraq in 2009; Research Interests: image and video compression techniques (transform coding, fractals, wavelet…), and Simulation.

Multi-data embedding in to RGB Image with using SVD method

Zineb EL GORFTE[1], Noureddine CHERKAOUI EDDEQAQI[1] ,Abdenbi BOUZID[1] and Ahmad.ROUKH[1]

[1] Department of physics, Faculty of Science, Moulay ismail University
B.P 11201 zitoune Meknes, Morocco

Abstract

Owing to the rapid of the communication and multimedia technology, documents are distributed through the internet simpler than before, but at the same time, these Advantages can also be used by pirates. The concept of digital watermarking is associated with data- hiding technique known Steganography. This mechanism of leading insertion should respect two conditions at least: mark must be indiscernible; the human eye should not differentiate between a not marked and marked image; the watermark should resist any voluntary or involuntary changes. Extraction should be also blind: to extract mark they don't need original picture. Many watermarking algorithms in the field DCT, wavelet and CDMA have been proposed seeking to optimize a compromise imperceptibility / robustness. Many recent techniques of watermarking are inspired by usual methods of coding and compression. The singular value decomposition (SVD) is one example. In this paper, we present the most recent method of SVD algorithm applied in color image; this blind method is based on embedding different data into the matrix of singular values. The results show that this method is robust against JPEG compression, color reduction (GIF) and the spreading of histogram.

Keywords: *Watermarking, robustness, singular value decomposition.*

1. Introduction

Nowadays, the usage of computers and the Internet has become increasingly common and widespread. The security of information exchange is very important on the network. Authentication and information hiding have also become important issues. To achieve these issues Digital watermarking plays a role to prove the rightful ownership or protect a customer's right of using the medium. The work in this field has led to several watermarking techniques such as correlation-based techniques, frequency domain techniques, DFT based techniques and DWT based techniques. The novel watermarking algorithm based on singular value decomposition (SVD) performs well in manipulations such as re-sampling, filtering, compression, cropping, scaling, rotation, etc.

We concentrate on a singular value decomposition (SVD) watermarking scheme. SVD is a linear algebra scheme that can be used for many applications, particularly in image compression, embedding watermark into the singular value decomposition (SVD) domain. First the SVD is performed on the original M x N image A

$$\boldsymbol{A = U\,S\,V^{H}}$$

Where U and V are an MxM orthogonal matrix and N x N orthogonal matrix, respectively, and S is an MxN diagonal singular value matrix. A watermark matrix W is then added into the matrix S, and SVD is performed on the new matrix $S+\ \alpha W$ to get Uw, Sw and $V\ w$, where is the scaling parameter that determines the strength of the watermark W.

$$\boldsymbol{S + \alpha W = Uw\ Sw\ V\ H}$$

The watermarked image Aw is obtained by:

$$\boldsymbol{U\ \ Sw\ V^{H} = Aw}$$

In this paper, we improve the SVD watermarking method, and extend it to the color images. Because the process of SVD does not involve the sensitivity of the human visual system, three components (R, G and B) on the color space are able to be utilized. Our proposed method can hide three grayscale images simultaneously to the color image. From a series of experiments, we will show it could resolve the rightful ownership of the watermarked image with good robustness, good imperceptibility and higher security

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

2. Numerical Results

The host color image A with size m× m is known, it contains three channels called red, green and blue channels. We can easily divide a color image into its red, green and blue channels and similarly red, green and blue channels can simply be combined to get the corresponding color image. We consider three images M1, M2 and M3 with size l× l as watermarks.
First, we apply the SVD to each component of the original image (red, green and blue component)

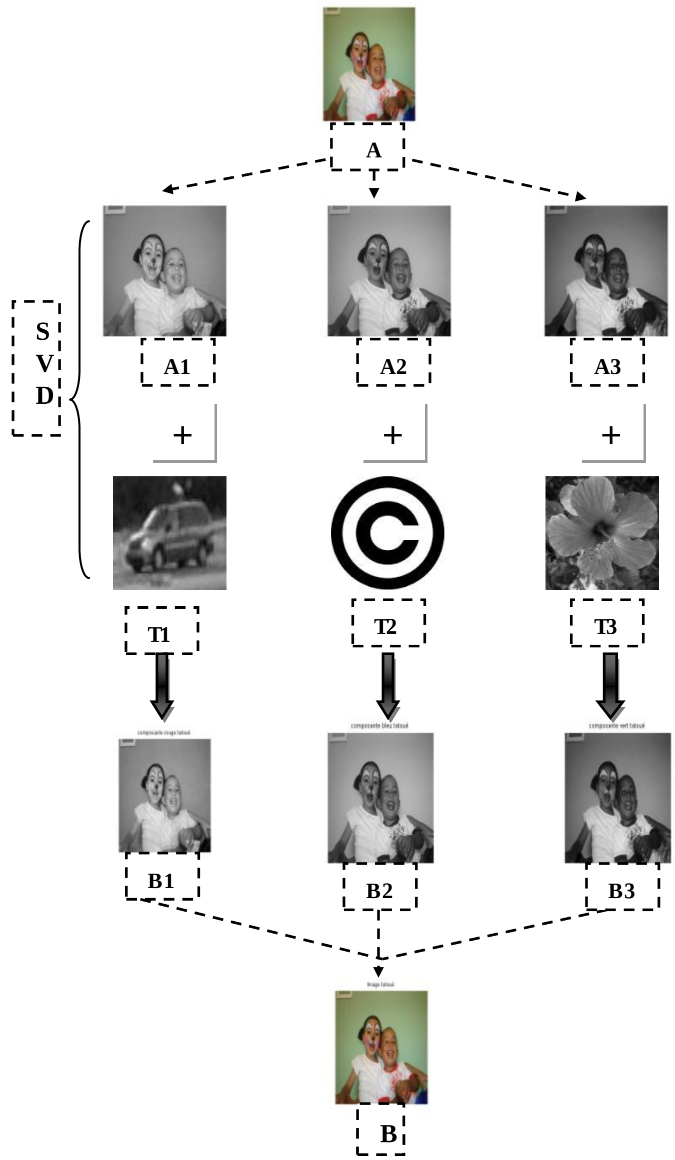
$$\mathbf{A} = \begin{pmatrix} A_R \\ A_G \\ A_B \end{pmatrix} \xRightarrow{\text{SVD}} \begin{pmatrix} U_1 S_R V_1 \\ U_2 S_G V_2 \\ U_3 S_B V_3 \end{pmatrix}$$

We can perform the embedding procedure by the following steps.

$$\begin{pmatrix} S_{wR} \\ S_{wG} \\ S_{wB} \end{pmatrix} \xLeftarrow{\text{Insertion}} \begin{pmatrix} S_R+\alpha T_1 \\ S_G+\alpha T_2 \\ S_B+\alpha T_3 \end{pmatrix}$$

α is a scaling factor, which controls the strength of the watermark to be inserted. Because the embedded watermark can't degrade the host image quality, the value of **α** should be small. During the process of SVD, three matrices U, V and S are produced.
(U_1, V1); (U_2, V2); (U_3, V3) these matrices are the user's secret keys, which don't contain any information about three watermarks.
A method of desirable watermarking should satisfy the ownership of imperceptibility, where the integrated watermark is not visible for the observer.
(A) Represent the Host image.
(T1,T2,T3) Represent the three watermarks for the protection of copyright, (A1,A2,A3) represent the three components of RGB image, (B1, B2, B3) represent the three components watermarked of RGB image ,(B) The watermarked image.
Here, we define the value of scaling factor α=0.1 and inject these watermarks into the RGB components of the host image

Fig1: Procedure for Digital Watermarking image RGB using SVD method



IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

3. Results:

We implemented the three watermarks in the host image. Several calculations were made to evaluate the watermarking algorithms. The watermarked images were examined and suitable tests were performed as explained below.

3.1 Imperceptibility

The imperceptibility of the watermark is tested through comparing the watermarked image with the original one.
As a measure of the quality of a watermarked image, Peak Signal to Noise Ratio (PSNR) is used.

To compute the Peak signal to noise ratio, first we need to compute the Mean Square Error (MSE) as in (1) [1]:

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

Peak Signal-to-Noise Ratio (PSNR) takes the signal strength into consideration (not only the error), is a better test here. It is written according to the following

$$PSNR = 20\log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right)$$

K: represents the matrix data of our original image
I: represents the matrix data of our watermarked image in question
m: represents the numbers of rows of pixels of the images
i: represents the index of that row
n represents the number of columns of pixels of the image
j: represents the index of that column
f represents the matrix data of our original image
MAX_f: is the maximum signal value that exists in our original "known to be good" image.

The result table shows the influence of insertion rate on the degradation value of psnr but not on the imperceptibility of the mark.

Number of watermarks	***1***	***2***	***3***
Image watermarked			
Psnr	***36.0151***	***32.5601***	***30.6426***

Number of watermarks	***1***	***2***	***3***
Image watermarked			
Psnr	***43.6500***	***40.6288***	***38.9514***

Table1: PSNR values for the multi data embedding using SVD method in to color image

4. Conclusions

In this paper, a new method of RGB image watermarking was proposed, the basic idea is to insert multiple watermarks in an RGB image
Based on the results obtained we can deduct that there is no visual difference between the original image and the watermarked image. Experimental results show that our method maintains a high quality of watermarked images and very robust against more conventional attacks.

5. References

[1] Hashida_K, Shiozaki_A, "A method of embedding robust watermarks into digital color images", IEICE Trans. on Fundamentals of Electronics Communications and Computer Sciences, vol. E81A, no. 10, pp. 2133-2137, 1998.

[2] R.Z. Liu and T.N. Tan, "An SVD-Based Watermarking Scheme for Protecting Rightful Ownership", IEEE Trans. on Multimedia, vol. 4, no. 1, pp. 121-128, 2002.

[3] Yongdong Wu, "On the Security of an SVD-based Ownership Watermarking", IEEE Trans. on Multimedia, vol. 7, no. 4, pp. 624-627, Aug. 2005.

[4] Xiao-Ping Zhang and Kan Li, Comments on "An SVD-Based Watermarking Scheme for protecting Rightful Ownership", IEEE Trans. on Multimedia, vol. 7, no. 2, Apr. 2005.

IJCSI
www.IJCSI.org

[5] C. I. Podilchuk and E. J. Delp, "Digital Watermarking: Algorithms and Applications," IEEE Signal Processing Magazine, July 2001, pp. 33-46.

[6]B. Zhou and J. Chen.A Geometric Distortion Resilient Image Watermarking algorithm Based on SVD.Chinese Journal of Image and Graphics, Vol. 9, April 2004, pp. 506-512.

[7] A. Benhocine, L. Laouamer, L. T. Nana, and A. C. Pascu,Measuring watermarking robustness using fractal dimensions, Journal of Communication and Computer, vol. 7, no. 1.

[8] H.C. Andrews and C.L. Patterson, "Singular value decomposition (SVD) image coding ", IEEE Trans. on Communications, vol. COM-24, no. 4, pp. 425-432, Apr. 1976.

[9] Kutter, F. Jordan, and F. Bossen, "Digital signature of color images using amplitude modulation", in Storage and Retrieval for Image and Video Databases V, SPIE vol. 3022, pp. 518~526, Jan. 1997.

A Novel Steganography with Preserving Statistical Properties

Dr. Thamir Rashed Saeed
Department of Electrical Eng.- Unv. Of Technology/ Iraq
Electronic Design for Signal Processing Research Group
Ciphering Research Group

Abstract

In this paper a novel algorithm is present to improve the capacity and security of stegaography technique. Where, the proposed work are based on three stages, first one is checking the statistical performance of the cover image. Second stage is splitting the pixels of cover image depending on the pixels values (edges). Then , at the third stage replacement one or two bits of cover image with massage or not depend on the pixels values in second stage. The results were, appear the detection probability by statistical calculation is 0 - 7.65 % with capacity 30-60 % and false-negative ratio is 76.8% and with distortion is 1.1-1.5%. The improvement that satisfied is 7.8% , 2.56% and 20% respectively.

***Keywords**: Steganography, LSB, Probability of Detection.*

1. Introduction

In the electronic era of revolutionary changes to the nature of information, and With the development of Internet technologies, digital media can be transmitted conveniently over the networks. Therefore, it is become difficult to confirm the safety or truth of data, due to the spread of tools used for distortion and manipulation and the capacity of some people to harness technology to serve malevolent ends[1,2]. Therefore, How to protect secret messages during transmission becomes an important issue[3], and this demand can be satisfied by hiding of the data that is, represent a key to safe communication[4]. That hiding can be in the content or form or, both, where, cryptography hides the contents of a secret message from malicious people, whereas Steganography even conceals the existence of the message[2,5]. The difference between steganography and cryptography is that steganography is a stealthy method of communication that only the communicating parties are aware of; while, cryptography is an overt method of communication that anyone understands, despite its payload is scribbled, and can use both of them in steganography. Where cryptography could use as a key of steganography[3,6].

2. Steganography

Steganography is the information hiding technique in which covering secured data into a computer carrier file without damaging the file or changing its size[4,7]. The purpose of steganography is covert communication-to hide the existence of a message from a third party[8]. Many different carriers file formats can be used, but digital images are the most popular because of their frequency on the Internet. For hiding secret information in images, there exist a large variety of steganographic techniques some are more complex than others and all of them have respective strong and weak points[9,10]. Stegaography contains a few branches as cryptography, Watermarking, Fingerprinting. The stego medium that a hidden message is inserted in it is called cover medium. It may be a picture, a sound, and a film. The image steganography algorithms can be categorized into two categories, spatial domain and frequency domain[11]. After inserting a message by Algorithm containing, it is called placed medium span (stego medium). The data that we insert in cover medium is secret message. The key that is used to insert message and take out secret message is called placing key. The techniques that help us to recognize cover medium and span are called disclosing[12,13].

The stego-process could be represented the following Formula[3,14]:

cover medium + embedded message + stegokey = stego-medium.

There are many embedded techniques, the Least Significant Bit (LSB) algorithm has a larger amount of capacity than other embedding techniques and it is recognized now, due to many advantages such as the algorithm is simple, the embedded velocity is fast and so on[2,11].

Image steganography systems have three conflicting conditions contend with one another: capacity, imperceptibility and robustness. Capacity refers to the amount of information that can be hidden in the cover medium, imperceptibility to an eavesdropper's inability to detect hidden information, and robustness to the number of modification the stego medium can withstand before an adversary can destroy that hidden informations[10]. There are three types of steganographic protocols used. They are Pure Steganography, Secret Key Steganography, and Public Key Steganography . The work of this paper is considered inside pure staganography.

In the existing technical literature, many related, studies on staganography have been reported. The researchers are focused on four axis, that are; the security[1,2,4,16], security and capacity[11,13,15], capacity and reduces the distortion[5,8,9,12,16,17,18], as;

In [1], two stages are made the work, first one is hidden a massage in a LSBs of the image. Then, the second is by hashing the cipher image using MD5 technique. While [7] use public key for hidden the massage in LSB, then, apply private key (RSA) for reconstruct the massage. By use particle swarm optimization for finding the best pixel locations to transfer the image to the new one was present in [4]. [2] was introduced a robust technique of hiding secret data in image based on LSB insertion and RSA encryption technique. For improving the security and capacity [13] was proposed a half tone picture method for secure the data communication. In [15] was combining the pixel-value differencing module and LSB for same demand. [11] is present a hiding massage in cover image by 3 LSB bits depend on the sequences of the massage bits. For reduce the embedding error and increasing the embedding capacity adaptive steganographic algorithm that use in [16]. Also [5] use fore-pixel difference and modified LSB to satisfy the same requirement. In [9] converting gray image to the RGB one to avoid gray color sequence and to declare the black region for hidden the massage in it. By shifting the word and synonym the test [8] was present a Steganography method. A steganographic technique based on genetic algorithm to find a near-beast structure for the pair-wise Least-Significant-Bit (LSB) matching scheme was presented in [12]. Improving LSB by decreasing the several changes in cover image by regarding to statistical properties of it was presented in [17]. In [18] a statistical properties of the cover image are taken before hidden the massage then adding a noisy bits in other location to rearrange the statistical properties after hiding a massage.

3. Proposed Algorithm

The most common and well-known steganographic method is called least significant bit (LSB) substitution, which embeds secret data by replacing k LSBs of a pixel with k secret bits directly. The human perceptibility has a property that it is sensitive to some changes in the pixels of the smooth areas, while; it is not sensitive to changes in the edge areas. Not all pixels in a cover image can tolerate, equal number of changes without causing noticeable distortion. Adaptive steganography which take care, about the important characteristics & statistics of the cover image.

The proposed algorithm based on the assumption that the intruder has full knowledge of the design and implementation details of the steganographic system. Also, its adopted according to statistics of image before, and after hiding the data.

3.1 Algorithm architect ;

For regarding the security with distortion, we must take care of the architect of the cover-image by taking the architect of the pixels as follows;

- Divide the pixels into three groups:-
  1- Whose value less than 6 (with respect to the most nibble)--LPC
  2- Whose value less than 12 and greater than 6 (with respect to the most nibble)--MPC
  3- Whose value greater than 12 (with respect to the most nibble)—HPC

- The hidden data will be depending on the above groups as in flow chart in figure (1) where;
  1. If the current pixel CP ∈ LPC no data will be hidden in that pixel.
  2. If the current pixel CP ∈ MPC one bit of data will be hidden in LSB of the pixel.
  3. If the current pixel CP ∈ HPC two bits of data will be hidden in LSB and beside it bits of the pixel.

Figure (2) represent the flowchart of data hidden in pixels.

■ Check number of pixels in MPC and HPC groups;

$$TNoP = Nop(MPC) + NoP(HPC) \quad (1)$$

■ Number of data (bits) can hidden in the cover image is;

$$TNoB = Nop(MPC) + 2 * NoP(HPC) \quad (2)$$

■ Select a prime number which has primitive numbers equal to TNoB

3.2 The algorithm steps are;

1. The image size Mz= 209 X 270 =56430 pixels, of the image in figure(3-a)[18].
2. Splitting the image matrix into two matrices Mz1 and Mz0 according to the LSB of each pixels before hide the massage;
   Where:
   Mz1= 28221pixels ,(LSB =1)and
   Mz0=28209pixels , (LSB =0).
   Then
   Splitting the image matrix into two matrices M2z1 and M2z0 according to the beside of LSB (bit number 2) of each pixels before hide the massage;
   Where:
   M2z1= 28141pixels ,(bit no. 2=1)and
   M2z0=28289pixels , (bit no. 2=0).
3. Splitting the image matrix into two matrices MM87 and MM76 according to the bits 8,7 and bits 7,6 of each pixels before hide the massage;
   Where:
   MM87= 4216 pixels , (bits 8, 7 =1)and
   MM76= 8850 pixels , (bits 7, 6 =1).
4. Hid two bits in LSB(bit 1,2) for the pixels in MM87 group, and one bit in LSB for the pixels in MM76 group.
5. Maximum text (massage or data) which can be hiding according to the proposed algorithm is 17282 bits (2.160) Kbyt.
6. Text (massage) size Tz =463 bytes (3707 bits).
7. The massage bits hidden distribution by nonrepated randomly to the LSB and LSB+beside according to bit 8,7 and bit 7,6 as ;
   Pixel No. which hides the bit of massage in LSB or LSB+its beside is
   $$PHB = a^i \bmod p \quad (3)$$
   Where;
   a- Primitive root of prime p
   $0 \leq i \leq p\text{-}1$
   P- prime number which has primitive root numbers equal to the message size.
8. Then splitting the image matrix as in step 2 after hiding a text. At this step its clear the number of pixels which have LSB equal one and zero are different from that before the hidden a text. This means the distortion can be detected.
9. To overcome the above weakness, determine the difference of the number of pixels which have a one and zero at LSB and beside bit from the pixels which not have a text.
10. Add one's and zero's to that pixel to satisfy the numbers in step 2, the cipher image as in figure (3-b)

3.3 Probability calculations:

The probability of any bit in a binary number is;

$$P(b)_n = ½ \quad (4)$$

Where;

b – bit in binary number.

n- No. of bits for that binary number.

The probability of any bits in a binary number is;

$$P(b)_n = P(b1) * P(b2) * \ldots P(b_N) = ½ * ½ * \ldots * ½ = \frac{1}{2*N} \quad (5)$$

$\therefore$ *for two bits*

$$P(b1 and\ b2) = \frac{1}{2*2} = \frac{1}{4} \quad (6)$$

Probability of certain number in the matrix is;

$$P(N)_m = \frac{the\ frequency\ of\ that\ number}{total\ matrix\ number} \quad (7)$$

$\therefore$ *for two bits in certain number in matrix is*

$$P(bM) = P(b1 and\ b2) * P(N)_m \quad (8)$$

The true-positive rate the probability that an image detected by Stegdetect really has steganographic content—as follows [19];

$$p(S|D) = \frac{p(S).P(D|S)}{P(D)} = \frac{P(S)\ .\ P(D|S)}{P(S).P(D|S) + P(\neg S).P(D|\neg S)} \quad (9)$$

Where

P(S) - probability of steganographic content in images,

P(¬S) - complement of P(S)

P(D|S) - probability that we'll detect an image that has steganographic content,

P(D|¬ S) is the false-positive rate. Conversely,

$$p(\neg D|S) = 1 - P(D|S) \quad (10)$$

Where;

$p(\neg D|S) -$ false-negative rate.

The probabilities with and without the proposed algorithm and for ref. [11] are shown in table (1).

Table (1) measurements f the proposed algorithm

Measurements / algorithm	Destortion %	PD by Human Vision	PD by Statistical	False-nagaitive ratio	Capacity
Without Proposed Algorithm	0.39% - 100%	≈ 0	100 %	0	100 %
With Proposed Algorithm	1.1% -1.5%	≈ 0	≈ 0 % - 7.65 %	76.8%	30 -60%
Ref. [11]	30%	≈ 0	~ 60%	30%	30-60%

4. Results:

From the results in Table (1), many strength points can be see, the distortion is reduces because the statistical properties are taken into account and the probability of detection as in that table and Figure (4) are improved about other algorithms. While, the capacity is not high with this algorithm because this algorithm was, focuses on the reduces the distortion and increase the security, but at the same time, it's not bad. The preference of this work with respect to the previous [18] in the false-negative ratio by 20%.

5. Conclusion

Often the statistical analyzed of stego- image which reflects or show the presence of certain hidden data in the image .In our proposed algorithm, who overcome this weakness of the steganographic system cause reduce the statistical detection, to reduce the distortion and increase the capacity by use hidden in LSB or LSB+beside bits of the image pixels depends on the value of pixel. Also, the detection rate weakness was overcome, by non-repeated, which the probability of detection nearly equal to zero.

However the two strength points in proposed algorithm are inversely related and depend on the text (massage) size. Where the text size decrease the non–repeated distribution factor effect on detection will reduce, and the adding noisy data to re-statistical of an original image factor effect will increase and the capacity will decrease, and vice versa. We examine our work by wifi internet network for many cases and get same results with the improvement which satisfied is 7.8% , 2.56% and 20% for probability of detection, false negative ratio and distortion respectively with respect of ref. 19. The weakness of this algorithm is still the capacity and we will overcome it in the next paper.

References

[1] Mohammad A. Ahmad, Dr. Imad Alshaikhli, Sondos O. Alhussainan, "Achieving Security for Images by LSB and MD5", Journal of Advanced Computer Science and Technology Research Vol.2 No.3, Pag., 127-139, September 2012.

[2] Emad T. Khalaf and Norrozila Sulaiman, "A Robust Data Hiding Technique based on LSB Matching", World Academy of Science, Engineering and Technology 58 2011.

[3] Jayeeta Majumder and Sweta Mangal, " An Overview of Image Steganography using LSB Technique", National Conference on Advances in Computer Science and Applications with International Journal of Computer Applications (NCACSA 2012).

[4] Parisa Gerami, Subariah Ibrahim and Morteza Bashardoost, " Least Significant Bit Image Steganography using Particle Swarm Optimization and Optical Pixel Adjustment", International Journal of Computer Applications (0975 – 8887) Volume 55– No.2, October 2012.

[5] Xin Liao, Qiao-yan Wen, Jie Zhang, "A steganographic method for digital images with four-pixel differencing and modified LSB substitution", ELSIVER, J. Vis. Commun. Image R., 2011.

[6] Youssef Bassil, "Image Steganography Method Based on Brightness Adjustment", Advances in Computer Science and its Applications (ACSA), ISSN: 2166-2924, Vol. 2, No. 2, 2012.

[7] Arpan Ghorai, Dibyendu Chowdhury and Satyajit Das, "Design and Implementation of Public Key Steganography", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-2, May 2012.

[8] Sharon Rose Govada, Bonu Satish Kumar, Manjula Devarakonda, and Meka James Stephen, " Text Steganography with Multi level Shielding", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012.

[9] Nitin Jain, Sachin Meshram, and Shikha Dubey, "Image Steganography Using LSB and Edge – Detection Technique",International Journal of Soft

Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.

[10] Ashish kumari , Shyama Sharma , Navdeep Bohra," Implementation of IMAGE STEGANOGRAPHY Based on Random LSB", IJCSMS International Journal of Computer Science and Management Studies, Vol. 12, Issue 01, January 2012.

[11] Gandharba Swain, and Saroj Kumar Lenka, " A Technique for Secret Communication Using a New Block Cipher with Dynamic Steganography", International Journal of Security and Its Applications Vol. 6, No. 2, April, 2012.

[12] M. Soleimanpour, S. Talebi and H. Azadi-Motlagh, " A Novel Technique for Steganography Method Based on Improved Genetic Algorithm Optimization in Spatial Domain", Iranian Journal of Electrical & Electronic Engineering, Vol. 9, No. 2, June 2013.

[13] Amin Hashemi Pour, and Ali Payandeh, "A New Steganography Method Based on the Complex Pixels", Journal of Information Security, 3, 202-208, 2012.

[14] Manish Mahajan and Navdeep Kaur, " Adaptive Steganography: A survey of Recent Statistical Aware Steganography Techniques", I. J. Computer Network and Information Security, 10, 76-92, Sept. 2012.

[15] Manjunath Gadiparthi, Keshav Sagar, Divya Sahukari and Rakesh Chowdary, "A High Capacity Steganographic Technique based on LSB and PVD Modulus Methods", International Journal of Computer Applications (0975 – 8887), Volume 22– No.5, May 2011.

[16] Yeuan-Kuen Lee, and Ling-Hwei Chen, " An Adaptive Image Steganographic Model Based on Minimum-Error LSB Replacement", 1999.

[17] Kazem G., Shahrokh G. and Saeed R. "LSB^{++}: An Improvement to LSB^{+} Steganography", IEEE TENCON Conf. Pag. 364-368, Bali, 2011.

[18] Thamir R. Saeed and Shaymaa Abd-Elghany, " EFFICIENT ADAPTIVE STEGANOGRAPHY ALGORITHM", Accepted for publication in International Journal of Pure and Applied Research in Engineering and Technology, 2013

[19] N. Provos and P. Honeyman, "Hid and Seek: An Introduction to Steganography," *IEEE Security and Privacy* , Vol. 1 Issue 3, Pag. 32-44, 2003.

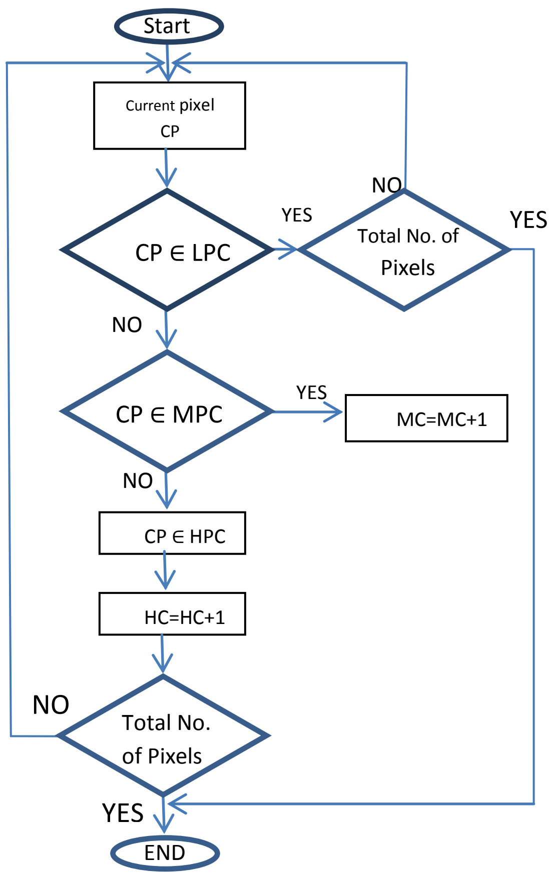


Figure (1) The architect of the cover-image pixels

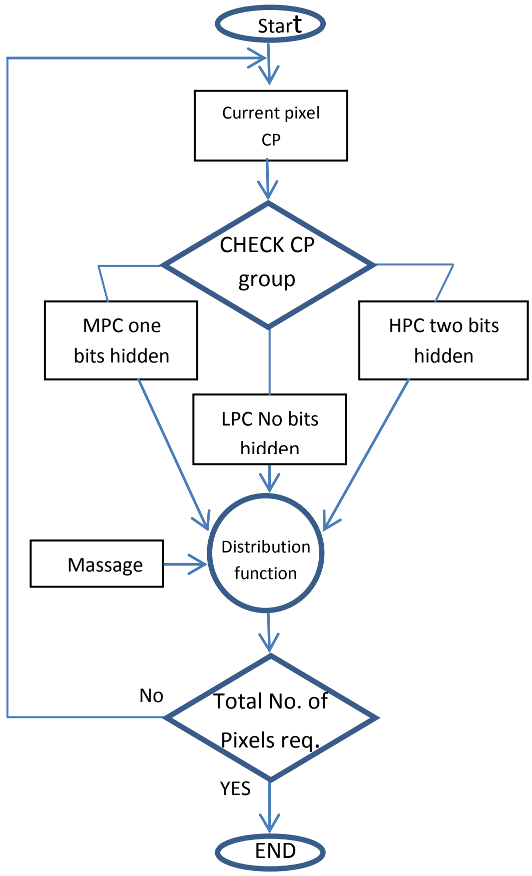


Figure (2) The flowchart of proposed algorithm

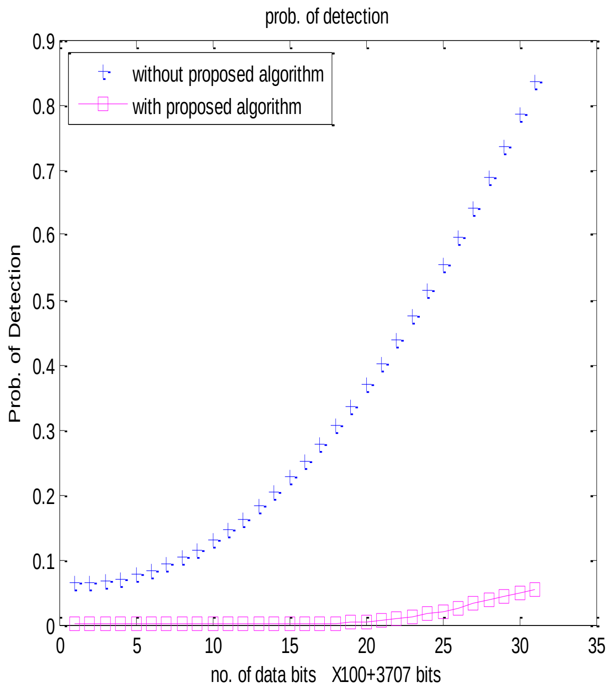
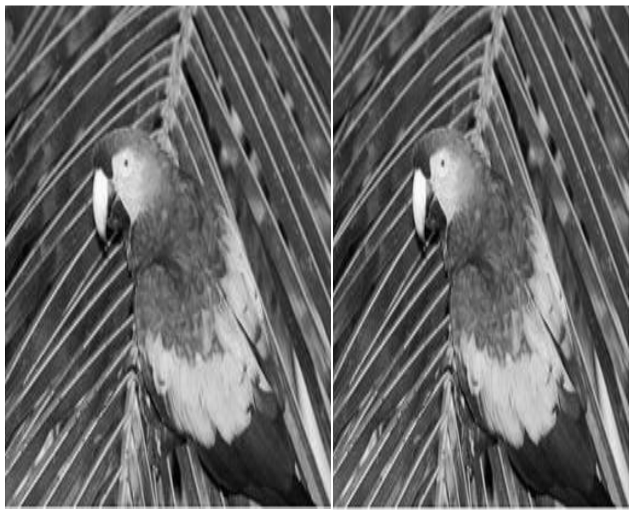


Figure (4) Prob. Of Detection with and without a proposed Algorithm



a-before hiding text

b-after hiding text

Figure (3) Cover-Image before and after hiding text

Propose HMNIDS Hybrid Multilevel Network Intrusion Detection System

Dr. Saad K. Majeed[1] Dr. Soukaena H. Hashem[1] Ikhlas K. Gbashi[1]

[1]Computer Sciences/ University of Technology/ Baghdad-Iraq

Abstract

This research present a proposal Hybrid Multilevel Network Intrusion Detection System (HMNIDS) which is a "hybrid multilevel IDS", is hybrid because use misuse and anomaly techniques in intrusion detection, and is multilevel since it apply the two detection techniques hierarchal in two levels. First level applies anomaly ID technique using Support Vector Machine (SVM) for detecting the traffics either normal or intrusions, if normal then passes it else the system input the intrusion traffic to the second level to detect the class of intrusion where this level apply Misuse ID technique using Artificial Neural Networks (ANN).

The proposal depend on Data mining is a DM-based HMNIDS since mining provide iterative process so if results are not satisfied with optimal solution, the mining steps will continue to be carried out until mining results are corresponding intention results. For training and testing of MHNIDS in our experiment, we used NSL-KDD data set. It has solved some of the inherent problems of the KDD'99. NSL-KDD similar to KDD99 their connections contains 41 features and is labeled as either normal or attack type, many of these features are irrelative in classification process. In our proposal we used Principle Component Analysis (PCA) as feature extraction to reduce no. of features to avoid time consuming in training and real-time detecting. PCA introduce 8 features as subset of correlated intrinsic features present the basic point in classification. The sets of features that have been resulted from PCA and the all features set will be the feeding of HMNIDS.

The results obtained from HMNIDS showing that accuracy rate of SVM and ANN classifiers separately are both high but they are higher with PCA (8) features than all (41) features. Confusion matrix of HMNIDS gives high detection rates and less false alarm rate, also they are higher with (8) PCA than all (41).

Keywords
NIDS, PCA, SVM, ANN and Confusion Matrix

1. Introduction

An intrusion can be defined as "any set of actions that attempt to compromise the integrity, confidentiality or availability of a resource". Also intrusion is any set of deliberate, unauthorized inappropriate, and/or illegal activity by perpetrators either inside or outside a system, which can be deemed a system penetration, that attempt to compromise the integrity, confidentiality or availability of a resource [1]. An intrusion detection system is a system that can analyze in real time or delayed events from a computer system. It detects overflows, among other rights and prevents visible signs of attacks against information systems. It's sort of a device to monitor the activity of a machine or network to detect intrusion attempts and generate alerts for possible against reactions and procedures. The detection of intrusions is an important component of infrastructure protection mechanisms and it analyzes the occurring events in the aim to identify intrusive behavior and establish a response plan [2, 3].

Intrusion detection systems have been widely applied to prevent and reduce damage to information systems. The criteria of an intrusion system are described as follows. Robustness, because IDSs are susceptible to attack, survivability is the essential ability such as redundancy, health checking, mobility, and dynamic recovery. Efficiency, the IDS modules can execute the sniffing and analysis software in time to detect the anomaly behaviors. Adaptability, the intrusion detection model does not depend on any special system, application environments, system vulnerabilities, or sorts of intrusions. Scalability, IDS can scale sizes for different traffics growing to avoid IDS services bottleneck. Feedback, the feedback measures the system whether to detect an intrusion and take certain actions automatically [4]. IDS can be classified according to IDS's environment as: a network-based IDS (NIDS) that is a dedicated computer, or special hardware platform, with detection software installed that captures packets in a promiscuous mode, or as a host-based IDS (HIDS) that monitors the resource usage of the operating system (OS) and the network. HIDS can only monitor the resource usage of the applications and not the applications themselves. Intrusion detection techniques are classified into two broad categories: misuse detection and anomaly detection. Misuse detection works by searching for the traces or patterns of well- known attacks. Clearly, only known attacks that leave characteristic traces can be detected that way. Anomaly detection, on the other hand, uses a model of normal user or system behavior and ages significant deviations from this model as potentially malicious. This model of normal user or system behavior is commonly known as the user or system profile. Strength of anomaly detection is its ability to detect previously unknown attacks [2, 3]. In order to overcome the limitations of traditional intrusion detection system, a systematic and higher automation method should be employed in the design of intrusion detection system. Data mining is the kind of effective method. Data mining concentrates on analyzing from a lot of noise, fuzzy and random data and extracts meaningful, potential useful information and knowledge. Mining is a repeated execution of the following three steps: Collect the data need for mining, Do the mining operation for the data and Get the corresponding result and then express the result with a certain way [5].

2. Related works

In [6] Ahmad I. et al, propose instead of using traditional approach of selecting features with the highest eigenvalues such as PCA, they applied a Genetic Algorithm (GA) to search the principal feature space for genetic eigenvectors that offers a subset of features with optimal sensitivity and the highest discriminatory power. Therefore, in this research, a mechanism for optimal features subset selection is proposed to overcome performance issues using PCA, GA and Multilayer Perceptron (MLP) is used for classification purpose. Consequently, this method provides optimal intrusion detection mechanism which is capable to minimize amount of features and maximize the detection rates. In [7] Reddy E. K. et al, they see network security technology has become crucial in protecting government and industry computing infrastructure. Modern intrusion detection applications facing complex problems. These applications has to be require reliable, extensible, easy to manage, and have low maintenance cost. In recent years, data mining-based intrusion detection systems (IDSs) have demonstrated high accuracy, good generalization to novel types of intrusion, and robust behavior in a changing environment. Still, significant challenges exist in the design and implementation of production quality IDSs. Instrumenting components such as of data transformations, model deployment, cooperative distributed detection and complex engineering endeavor. In [8] Suebsing A. et al, see in the previous researches on feature selection, the criteria and way about how to select the features in the raw data are mostly difficult to implement. Therefore, this work presents the easy and novel method, for feature selection, which can be used to separate correctly between normal and attack patterns of computer network connections. The goal in their work is to effectively apply Euclidean Distance for selecting a subset of robust features using smaller storage space and getting higher Intrusion detection performance. Experimental results show that the proposed approach based on the Euclidean Distance can improve the performance of a true positive intrusion detection rate especially for detecting known attack patterns. In [9] Bensefia H. et al., propose a new approach for IDS adaptability by integrating a Simple Connectionist Evolving System (SECOS) and a Winner-Takes-All (WTA) hierarchy of XCS (eXtended Classifier System). This integration puts in relief an adaptive hybrid intrusion detection core that plants the adaptability as an intrinsic and native functionality in the IDS. In [10] Vaarandi R., proposed a novel unsupervised DM based approach for IDS alert classification. With this strategy, knowledge is mined from IDS logs and processed in an automated way, in order to build an caution classifier. The classifier is then used in real-time for discerning important IDS warns from frequently occurring false positives and events of low significance. In [11] Vaarandi R. et al., extended their previous work (Risto, 2009) on IDS alert classification, and present a novel unsupervised real time alert classification method which is based on frequent itemset mining and data clustering techniques. Their method first applies a frequent itemset mining algorithm to past IDS alert logs, in order to discover patterns that describe redundant alerts. After that, data clustering methods are used for finding detailed subpatterns for each detected pattern. Finally, the detected knowledge is explained and used for real time classification of IDS alerts, in order to characterize critical alerts from irrelevant ones. In [12] Mohammad M. N. et al., introduced an improved approach for IDS based on combining DM and expert system that is presented and implemented in WEKA (Waikato Environment for Knowledge Analysis). They aimed to design and develop intelligent DM IDS and its core part a composite detection engine with anomaly detection and misuse detection features and the two detection engines work serially to detect the user's activity in turn. The system collects the data of DB audit system in real time, analyzes the audit data, judges that it is a normal behavior, abnormal behavior or aggressive behavior and responds to the result obtained by the operation behavior and finally reports the result to the manager in a comprehensible form. In [13] Guojun Z. et al., presented a cooperative IDS based on IPv6 to address this challenge. Such a system consists of four parts: data flow tracking and analysis, capturing packets and rules matching, disaster recovery, and blocking. The technique of cooperative ID is introduced into the system for realizing the coordination control among parts. The system has a perfect detection rating. In [14] Al-Janabi S. T. and Saeed H. A. proposed an anomaly based IDS that can promptly detect and classify various attacks. Anomaly-based IDSs need to be able to learn the unstable behavior of users or systems. The proposed IDS experimenting with packet behavior as parameters in anomaly ID. There are several methods to assist IDSs to learn system's behavior, the proposed IDS uses a back propagation artificial neural network (ANN) to learn system's behavior and uses the KDD CUP'99 dataset in its experiments and the obtained results satisfy the work objective. In [15] Haldar N. et al presented IDS which employs usage of classification methods to model the usage patterns of authenticated users and uses it to detect intrusions in wireless networks. The key idea behind the proposed IDS is the identification of discriminative features from user's activity data and using them to identify intrusions in wireless networks. The detection module uses statistical methods to accumulate interested statistical variables and compares them with the thresholds derived from users activities data. When the variables exceed the predestined thresholds, an alarm is put forward to alert about a sensible intrusion in the network.

3. Analysis for Critical Points in Current NIDS

From our survey of traditional and current NIDS there are number of significant drawbacks, we see these critical issues must be taken in consideration to construct our proposed model, these drawbacks are:

Traditional dataset KDDCUP'99 is the mostly widely used data set for the intrusion detection. But this dataset has huge no. of records and high no. of redundant records. That make selecting subsets of records for training and testing fatiguing for researchers and affect on performance of IDSs.

NIDS are usually tuned to detect known service level network attacks. This leaves them vulnerable to original and novel malicious attacks.

That amount of data needs to look at seems to be growing rapidly. Depending on the intrusion detection tools employed by a company and its size there is the possibility for logs to reach millions of records per day.

False positives, a common complaint is the amount of false positives an IDS will generate. A false positive occurs when normal attack is mistakenly classified as malicious and treated accordingly.

False negatives, this is the case where IDS does not generate an alert when an intrusion is actually taking place.

In anomaly detection the subject's normal behavior is modeled on the basis of the (audit) data collected over a period of normal operation. If undiscovered intrusive activities occur during this period, they will be considered normal activities. Because of some technical reasons, the current anomaly detection approaches usually suffer from a high false-alarm rate.

Another difficult problem in building anomaly detection models is how to decide the features to be used as the input of the models. It is not guaranteed that all and only the features related to intrusion detection will be selected as input parameters. Although missing important intrusion-related features makes it difficult to distinguish attacks from normal activities, having non-intrusion-related features could introduce "noise" into the models and thus affect the detection performance.

Current misuse detection systems usually work better than anomaly detection systems for known attacks. That is, misuse detection systems detect patterns of known attacks more accurately and generate much fewer false alarms. This better performance occurs because misuse detection systems take advantage of explicit knowledge of the attacks.

Limitation of misuse detection is that it cannot detect novel or unknown attacks. As a result, the computer systems protected solely by misuse detection systems face the risk of being comprised without detecting the attacks.

4. The Proposed Model of HMNIDS

The proposed HMNIDS, see figure (1), is a "hybrid multilevel IDS", is hybrid because it trend to detect intrusions with both techniques misuse and anomaly, and is multilevel because it trend to detect intrusions with two levels for more accurate intrusions types detections. First level of proposed HMNIDS will apply anomaly ID technique should first learn the characteristics of normal activities and abnormal activities of the system, and then the HMNIDS detect traffic that deviate from normal activities. So, the result of first level is detecting the traffics either normal or intrusions, if normal then passes it else enter the intrusion traffic to the second level to detect the class of intrusion. Second level of proposed HMNIDS will apply Misuse ID technique is able to automatically retrain ID models on different input data that include new types of attacks, as long as they have been labeled appropriately. The results of this level are detecting the type of intrusions.

The proposal is a DM-based HMNIDS in which both the misuse and anomaly detection techniques depended in the detection of intrusion, where each instance in a dataset is labeled as "normal" or "intrusion (specify intrusion type)" and a learning algorithms is trained over the labeled data. If the mining results are not satisfied with what we want to, the mining steps will continue to be carried out until mining results are corresponding with our intention results. For training and testing of the proposed MHNIDS in our experiment, we used NSL-KDD data set. It has solved some of the inherent problems of the KDD'99. It is considered as standard benchmark for intrusion detection evaluation. The training dataset of NSL-KDD similar to KDD99 consist of approximately 4,900,000 single connection vectors each of which contains 41 features and is labeled as either normal or attack type ,with exactly one specific attack type. We see the demand for the number of samples for the training the classifier grows exponentially with the dimension of the feature space. This limitation is called the ‘curse of dimensionality’. In our proposed HMNIDS model we indicate that feature reduction technique is capable of reducing the size of dataset. The time and space complexities of most classifiers used are exponential function of their input vector size.

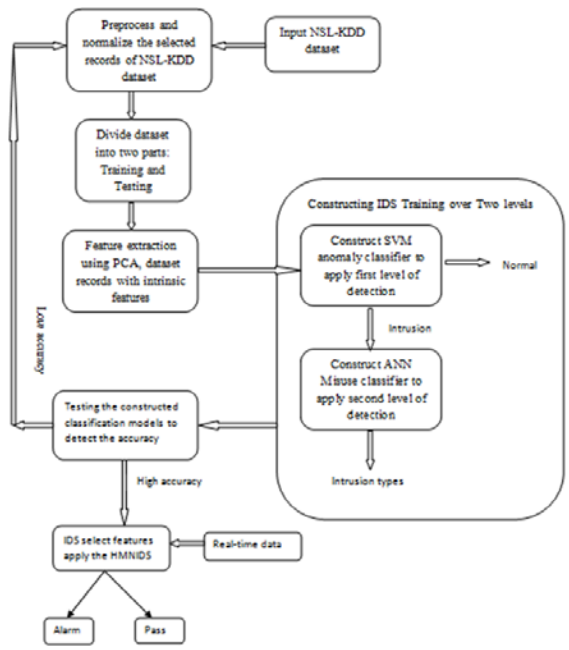


Fig.1. Detailed Proposed HMNIDS

4.1 Description of Dataset

KDD'99 training dataset consists of approximately 5 million connection records (a connection is a sequence of TCP packets starting and ending at some well-defined times, between which data flows to and from a source IP address to a destination IP address under some well-defined protocol) each of which contains 41 features and is labeled as either normal or an attack, with exactly one specific attack type. These features are grouped into four categories: basic features, content features, timebased traffic features and host-based traffic features. The simulated attacks fall in one of the following four categories: Denial of Service Attack (DoS), User to Root Attack (U2R), Remote to Local Attack (R2L), and Probing Attack. A total of 22 training known attack types and additional 17 unknown types are summarized.

The KDD dataset consists of three components: "Whole KDD", "Corrected KDD" and "10% KDD", KDDCUP’99 is the mostly widely used data set for the intrusion detection. But researchers conducted a statistical Analysis on this data set and found important issues which highly affect the performance of evaluated systems, and results in a very poor evaluation of anomaly detection approaches. To solve these issues, they have proposed a new data set, NSL-KDD, which consists of selected records of the complete KDD data set. NSL-KDD has many advantages over the original KDD dataset; It does not include redundant records in the train set, so the classifiers will not be biased towards more frequent records. The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set. As a result, the classification rates of distinct machine learning methods vary in a wider range, which makes it more efficient to have an accurate evaluation of different learning techniques. The numbers of records in the train and test sets are reasonable, which makes it affordable to run the experiments on the complete set without the need to randomly select a small portion. Consequently, evaluation results of different research works will be consistent and comparable.

4.2. Preprocessing on NSL-KDD Dataset

The following processes have been applied to the "NSL-KDD dataset" before it being used in design of the proposed system:

Converting the original NSL-KDD dataset from a text file to SQL server.

Since type of some of NSL-KDD dataset's features is continuous, thus a process for normalization these features have been done in order to become of categorical type so it becomes more convenient with the used DM classification algorithms. Normalization is used for data preprocessing, where the attribute data are scaled so as to fall within a small specified range such as -1.0 to 1.0 or 0.0 to 1.0. Since using neural network back propagation algorithm for classification, normalizing the input values for each attribute measured in the training samples will help speed up the learning phase.

The resulted dataset from process two will be split into two distinct datasets by using, one for classifiers' training which nearly equal 2⁄3 of resulted dataset from process two and the other for classifiers testing which nearly equal 1⁄3 of the same resulted dataset.

4.2 Process of Feature Reduction

Feature selection is intended to suggest which features are more important for the prediction, to find out and get rid of irrelevant features that reduce classification accuracy, discover relations between features and throw out highly correlated features which are redundant for prediction. An earlier general task in data mining is to extract outstanding features for the prediction. This function can be broken into two groups; feature extraction or feature transformation, and feature selection. Feature extraction refers to the process of creating a new set of combined features (which are combinations of the original features). On the other hand, feature selection is different from feature extraction because it does not produce new variables. Feature selection also known as variable selection, feature reduction, attribute selection, feature ranking, or variable subset selection, is a widely used dimensionality reduction technique, which has been the focus of much research in machine learning and data mining.

In our proposal we used Principle Component Analysis (PCA) as feature extraction rather than using any techniques of feature selection. That because we have 41 features, like this no. of features will cause time consuming in training and real-time detecting, so we need to transform these set of 41 features into small subset of correlated intrinsic features present the basic point in classification. Principle Component Analysis (PCA) is a useful statistical technique that has found application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension. The complete subject of PCA statistics is based on the idea that you have this huge set of data, and you want to analyze that set expressions of the relationships between the single points in that set. PCA is applied to the training dataset to find the intrinsic features, see algorithm (1), the resulted set of features in addition to the original set present all features in the training dataset will be used in design (learning) of the classifiers.

Algorithm (1): Suggested-PCA

Input: NSL-KDD training dataset.

Output: PCA set of most frequent and related features.

Steps:

Obtain training NSL-KDD'99 transactions.

Represent every transaction I_i as a vector x_i.

Compute the average transaction $\Psi = 1/M \sum_{i=1}^{M} x_i$ …………………………...(1)

Subtract the mean transaction $\varphi_i = x_i - \Psi_i$……………………………...…. (2)

Compute the covariance matrix $C = 1/M \, [\![\phi_n \varphi]\!]_n^T = AAT$……………………..……(3)

From C Compute eigenvectors u_i of AAT:

Consider matrix AAT as a M×M matrix.

Compute the eigenvectors v_i of AAT such that:

$ATAv_i \rightarrow \mu_i V_i \rightarrow AATAV_i = \mu_i Av_i \rightarrow Cu_i = \mu_i u_i$ where $\mu_i = Av_i$………..(4)

Compute the μ best eigenvectors of AAT: $\mu_i = Av_i$ …………………….(5)

Keep only K eigenvectors, (K features with their values).

End.

4.3. Suggested Hybrid Multilevel NIDS

After the intrinsic features had been selected, the two popular DM classification algorithms: Support Vector Machine SVM from statistical field and Artificial Neural Networks from soft computing field, these two

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

classifiers will used in the design of the suggested HMNIDS in shape of two levels, So both of SVM and ANN classifiers will be trained 2 times, one training with dataset of all 41 features and other with dataset of PCA features to design the suggested HMNIDS classifier, these levels are:

First Level (SVM-Anomaly Detection), Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. SVMs attempt to separate data into multiple classes (two in the basic case) though the use of a hyper-plane. Here will use in more conventional SVM approach, we used one SVM, as anomaly detection techniques to identify normal traffic from intrusion traffics. Algorithm (2) explains SVM algorithm with Anomaly Intrusion detection learned on NSL-KDD dataset.

Algorithm (2): Suggested-SVM

Input: NSL-KDD for training and testing.

Output: Results of Anomaly detection on NSL-KDD testing dataset using SVM.

Steps:

1.Initialize all points in training dataset as (Xi, Yj) where X is a vector of data x1, ……, xn and Y is vector of classes.

2.Initialize vector of weight W.

3.Distribute all points (x, y) and extract the hyper plane separator.

4.If the hyper plane give optimal separation then depend hyper plane as classifier model to classify testing dataset and go End

5.Else must do the following steps

6.Maximize the hyper plan using equation of Getting Maximum Margin = MM = 2 / ||w||. …………. (6)

7.For minimum using equation same as maximizing $\Phi(w) = \frac{1}{2} w^T w$ ………. (7)

8.Initialize Lagrange multiplier αi vector α1…αn using equation

Q(α) =Σαi - ½ΣΣαiαjyiyjxiTxj …………. (8)

9. Apply classification function using equation f(x) = ΣαiyixiTx + b …………. (9)

10. Determine the support vectors xi with non-zero αi (support vectors are the points determine the area of hyper plan)

11. Depend the hyper plan resulted after determining support vectors as the classifier model to classify testing dataset

12. End

Second Level (ANN-Misuse Detection), an artificial neural network is a system simulating work of the neurons in the human brain. The neuron consists of some inputs emulating dendrites of the biological neuron, a summation module, an activation function and one output emulating an axon of the biological neuron. The importance of a particular input can be intensified by the weights that simulate biological neuron's synapses. Then, the input signals are multiplied by the values of weights and next the results are added in the summation block. The sum is sent to the activation block where it is processed by the activation function. Thus, we obtain neuron's answer to the input signals "x". Here will use in more conventional MLP (Multi Layer Perceptron) approach as misuse detection techniques to identify class of intrusion traffics (these are detected intrusion in the first level classifier and sent to the second level to determine it is class). Algorithm (3) explains ANN algorithm with Misuse Intrusion detection learned on NSL-KDD dataset.

Algorithm (3): Suggested-ANN

Input: NSL-KDD for training and testing.

Output: Results of Misuse detection on NSL-KDD testing dataset using ANN.

Steps:

Main Assumption for the Training Process of MLP:

Learning method: Quasi Newton BFGS and Levenberg-Marquardt

Number of Epochs: 1000.

MSE (Mean Square error): 0.01.

Learning rate: 0,9.

Activation function: log-signoid.

Number of neurons in the Input layer: (41 or according no. of PCA set).

Number of neurons in the hidden layer: (21 for 41 input neurons and with no. of PCA set equal to half of this no.).

Number of neurons in the output layer: (4 cause no. of intrusions classes are 4).

Update of weights – batch mode (after presentation of the entire training data set).

Train and Test on NSL-KDD to construct final ANN misuse NIDS.

End.

5. Discussion and Experimental works

The proposal had been implemented on the following platform: Windows 7 Ultimate Service Pack1 and 32-bit OS, 16GB RAM, and Intel® Core (TM) 3 Duo CPU with 2.00 GHz; and by using Visual Basic.Net and SQL server.

Training results of the proposed MHNIDS will presented with two stages; first one will introduce training results of anomaly classifier (SVM) and misuse classifier (ANN) separately. Second level will introduce training results of the complete proposed HMNIDS. That partitioning in presenting the results of training to discover the strength and weak points in the proposed system in all it is parts. Before explaining the results will introduce the no. of records taken from NSL-KDD as samples for training and testing, see table (1), then present the set of intrinsic features obtained by applying PCA on training dataset, this subset is {Protocol_type, Service, Flag, count, srv_count, same_srv_rate, dst_host_srv_count, dst_host_same_srv_rate, dst_host_same_src_port_rate}.

No. of Records/Type of Records	Training	Testing
DOS	Normal	23,000
Probe	500	1,000
R2L	150	2,000
U2R	100	500
Normal	10,000	7,000

Table (1): No. of Records selected from NSL-KDD for

First Training (SVM and ANN separately)

In this study, as in table (1) Training data set in the paper contained 50,750 records, which were randomly generated from the NSL-KDD for training data set that consists of 10,000 normal patterns, 40, 000 known DoS patterns, 500 known Probe patterns, 150 known R2L patterns and 100 known U2R patterns. Testing data set in the paper contained 33,500 records, which were randomly generated (with omitting the records of training) from the NSL-KDD for testing data set that consists of 7,000 normal patterns, 23, 000 known and unknown DoS patterns, 1,000 known and unknown Probe patterns, 2000 known and unknown R2L patterns and 500 known and unknown U2R patterns.

Training which consist of two hierarchy classifiers (SVM and ANN) on Training dataset has been done with two sets of features (All_Features, PCA_Features), so the proposed system has been experimented (i.e., trained and tested) for two times to assess the accuracy of the classifiers. We performed three different experiments and selected a subset of eight features that indicates better performance as compared to others. Our aim is to select minimum features that produce optimal results in accuracy. This definitely impact on overall performance of the system. The features are reduced to 8 from the 41 raw features set. The above experiments show that optimal features increased accuracy, reduced training and computational overheads and simplified the architecture of intrusion analysis engine.

Results of three conducted experiments (Exp1, Exp2, Exp3), which producing the most accurate results, have been presented in this section. Four classification models have been constructed in each of these three experiments. Next these models have been applied on the same Testing dataset, which has been constructed during Exp1, to assess the validation and accuracy of these constructed models on the same testing dataset. The classification results of testing are either TP (intrusion), TN (normal), false positive (FP) (misclassified as intrusion), false negative (FN) (misclassified as normal), Unknown (new user behavior or new attack). From classification results we calculate the detection rate (DR) of IDS is the ratio between the number of TP and the total number of intrusion patterns presented in the testing dataset. It has been calculated using

DR = TP/(TP+FN+Unknown2) *10………………(10)

,and the false alarm rate (FAR) of an IDS is the ratio between number of "normal" patterns classified as attacks (FP) and the total number of "normal" patterns presented in the testing dataset. It has been computed using

FAR=FP/(TN+FP+Unknown1)*100%……………(11)

Values for both of DR and FAR for each classifier in the three experiments have been illustrated in table (2).

Table (2): DRs and FARs of both of them SVM and ANN

SVM-Anomaly Classifier				ANN-Misuse Classifier			
Feature Selection Measure	Experiment No.	DR	FAR	Feature Selection Measure	Experiment No.	DR	FAR
PCA	1	1	0	PCA	1	0.999	0
	2	0.998	0		2	0.999	0
	3	0.999	0		3	0.999	0
ALL	1	0.995	0.02	ALL	1	0.994	0.03
	2	0.996	0.03		2	0.988	0.03
	3	0.995	0.07		3	0.995	0.03

DR are higher with SVM anomaly classifiers and also with ANN misuse classifiers and FAR often ranging between (0 - 0.07) with SVM classifiers and ANN classifiers. It is very clear from these that SVM classifiers are better with anomaly detection and ANN classifiers are better with misuse detection. Selection of the best classification model would be done significantly according to its classification accuracy, which is introduced as the ratio between the number of the correctly classified patterns (TP, TN) and the total number of patterns of the testing dataset. The accuracy (Accu) of each classifier has been calculated using

$$Accu = (TP+TN)/(TP+FP+TN+FN+unknown) *100\% \quad \ldots\ldots\ldots (12)$$

Table (3) summarizes Accu of both SVM and ANN classifiers with PCA_F and ALL_F in the three experiments. According to these results, the classifiers SVM and ANN were more accurate with PCA_F Accu.

Table (3): Accuracy of SVM and ANN Classifiers

Classifier	Experiment No.	PCA_F	ALL_F
SVM-Anomaly	1	1	0.995
	2	1	0.996
	3	0.999	0.997
ANN-Misuse	1	1	0.994
	2	0.999	0.997
	3	0.999	0.995

Second training (Full HMNIDS)

Here will explain the overall accuracy of the proposed HMNIDS according confusion matrix calculations, we will display the confusion matrix for two cases; all 41 features and 8 features, see table (4) and table (5). We see the confusion of classification with 8 features less than confusion of classification with 41 features.

Table (4): confusion matrix of HMNIDS with 41 features

Confusion Matrix	DOS	Probe	R2L	U2R	Normal
DOS	22,981	3	6	0	10
Probe	10	985	1	0	5
R2L	6	1	1984	0	9
U2R	1	0	0	448	1
Normal	6	9	3	0	6992

Table (5): confusion matrix of HMNIDS with 8 features

Confusion Matrix	DOS	Probe	R2L	U2R	Normal
DOS	22,990	2	5	0	3
Probe	8	988	1	0	3
R2L	5	1	1989	0	5
U2R	1	0	50	448	1
Normal	1	3	1	0	6995

6. Conclusion

Data mining are introduced for helping IDS to detect intrusions correctly, and accordingly IDSs have shown to be successful in detecting known attacks. Feature selection is an important task of Network Intrusion application. Using PCA feature selection approach, intrusions are detected with less error rate and high accuracy. Usage of ANN for misuse intrusion detection and SVM for anomaly detection with the input data from the NSL-KDD gives good performance of HMNIDS as comparison with depended related works [6-15]. Tables (2 and 3) present results of detection for SVM and ANN separately where notice the higher rates of detection and very less rates of false alarms especially with PCA set of features. The same in tables (4 and 5) present results of detection for HMNIDS where notice from confusion matrices the higher rates of detection and very less rates of false alarms especially with PCA set of features.

References

1.Lalli and Palanisamy, “Modernized Intrusion Detection Using Enhanced Apriori Algorithm”, International Journal of Wireless & Mobile Networks (IJWMN) Vol. 5, No. 2, April 2013, [IVSL].

2.Saidi Ben Boubaker Ourida , “Implementation of an Intrusion Detection System”, International Journal of Computer Science Issues(IJCSI) ,Vol. 9, Issue 3,No. 1, May 2012, ISSN (Online): 1694-0814, www.IJCSI.org.

3.Garuba M., Liu C., Fraites D., "Intrusion Techniques: Comparative Study of Network Intrusion Detection Systems", IEEE Computer Society, Fifth International Conference on Information Technology, pp. 592-598, 2008.

4.Huang L. and Hwang M. “Study of Intrusion Detection Systems”, Journal Of Electronic Science And Technology, Vol. 10, No. 3, September 2012.

5.Zhou Q. and Zhao Y. , “ The Design and Implementation of Intrusion Detection System based on Data Mining Technology”, Journal of Applied Sciences, Engineering and Technology 5(14): 3824-3829, 2013 ISSN: 2040-7459; e-ISSN: 2040-7467 © Maxwell Scientific Organization.

6.Ahmad I., Abdulah A. B, Alghamdi A. S, Alnfajan K. and Hussain M., “Feature Subset Selection for Network Intrusion Detection Mechanism Using Genetic Eigen Vectors”, 2011 International Conference on Telecommunication Technology and Applications, Proc .of CSIT vol.5 (2011) © (2011) IACSIT Press, Singapore.

7.Reddy E. K. , Reddy V. N., Rajulu P. G., “A Study of Intrusion Detection in Data Mining”, Proceedings of the World Congress on Engineering 2011 Vol III WCE 2011, July 6 - 8, 2011, London, U.K.

8.Suebsing A., Hiransakolwong N. , "Euclidean-based Feature Selection for Network Intrusion Detection", 2009 International Conference on Machine Learning and Computing IPCSIT vol.3 (2011) © (2011) IACSIT Press, Singapore.

9.Bensefia H. and Ghoualmi N., "A New Approach for Adaptive Intrusion Detection", 2011 Seventh International Conference on Computational Intelligence and Security, 2011.

10.Vaarandi R., "Real-Time Classification of IDS Alerts with Data Mining Techniques", MILCOM'09 Proceedings of the 28th IEEE conference on Military communications pp.1786-1792.

11.Vaarandi R., and Podinš K., "Network IDS Alert Classification with Frequent Itemset Mining and Data Clustering", The 2010 International IEEE Conference on Network and Service Management, pp. 451-456.

12.Mohammad M. N., Sulaiman N. and Muhsin O. A., "A Novel Intrusion Detection System by using Intelligent Data Mining in Weka Environment", Published by Elsevier Ltd. Procedia Computer Science 3, pp. 1237–1242.

13.Guojun Z., Liping C. and Weitao H., "The Design of Cooperative Intrusion Detection System", IEEE Computer Society, 2011 Seventh International Conference on Computational Intelligence and Security, pp. 764-766.

14.Al-Janabi S. T., and Saeed H. A., "A Neural Network Based Anomaly Intrusion Detection System", IEEE Computer Society, 2011 Developments in E-systems Engineering, pp. 221-226.

15.Haldar N. A., Abulaish M. and Pasha S. A., "An Activity Pattern Based Wireless Intrusion Detection System", IEEE Computer Society, 2012 Ninth International Conference on Information Technology-New Generations, pp. 846-847.

Efficient Computation of Resonant Frequency of Rectangular Microstrip Antenna using a Neural Network Model with Two Stage Training

Guru Pyari Jangid*, Gur Mauj Saran Srivastava and Ashok Jangid

Department of Physics and Computer Science
Dayalbagh Educational Institute (Deemed University)
Dayalbagh Agra 282005, INDIA

Abstract

Artificial neural networks (ANNs) with two stage training have been proposed for efficient computation of resonant frequency of rectangular microstrip antenna. In the proposed approach in first stage the ANN model trained with empirical relation of resonant frequency with structural and substrate parameters of antenna then in second stage the model has been trained with the actual experimental data. The proposed approach has been validated using experimental published data and compared with results of other models published in different research papers. The result shows that the proposed approach is more accurate than the models developed using experimental data only. The results of the two stage training are in very good agreement with the measurements, and better accuracy than other ANN models developed using experimental data only.

Keywords: *Artificial neural networks (ANNs), computer aided design (CAD), learning algorithms, microstrip antenna, microwave device modeling.*

1. Introduction

During the past four decades, there has been a spectacular progress in microwave technology based microstrip devices and its application to both military and civilian areas primarily due to their simplicity, light weight, low profile, conformability, reproducibility, low manufacturing cost, reliability, ease in fabrication and integration with solid-state devices [1-3]. In military applications, it has played a key role in radar and electronics warfare (EW) systems and for the civil purposes, microwave has greatly helped in the expansion of mobile and satellite communication facilities. The emergent commercial civil and defence market of wireless communication devices over the past decade has led to explosion of interest and opportunities for design and developments of different types of microwave components. The wireless industry emphasizes on the development of components or systems (group of components) in shortest possible time and at low development cost. Modern industrial developments in the design of microwave components suggest the development of fully integrated subsystems that can be fabricated in large numbers. This places the demand of computer-aided-design (CAD) tools for the development of required microwave components and systems. The main objective of microwave CAD is faster and accurate development of components or systems while maintaining their efficiency.

Efforts to lower the cost and reduce the weight and volume of monolithic microwave and millimeter wave integrated circuits (MMIC's) have resulted in high-density circuits where a large number of interconnects are used. With this increased complexity and higher operating frequencies of microwave and millimeter wave devices an accurate and efficient design procedure is required to carry out the device synthesis. In present scenario methods of designing generally used are based on electromagnetic theory but there are some drawbacks of these methods like quasi-static models are valid only at low frequencies or over very short range of frequency and not available for all the devices. Full-wave characterization can lead to accurate results, but at much higher computational expense and they are very time consuming which prevents their use in practical interactive CAD. Many new EM simulation tools are being developed by industries to automate the design process. Some of them are embedding search methods like conjugate-gradient method, quasi-newton method, etc. for optimizing the design parameters. The drawback of these methods is that they require initial guess which should be close to the optimum design otherwise they reach up to local minima. Soft-computing algorithms [4] are reliable alternatives to these methods for getting optimum designs.

In the present paper, an efficient computation of resonant frequency of rectangular microstrip antenna has been presented using new two stage training on ANN model. The ANN model is used with four structural parameters of

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

antenna that are thickness of substrata, relative dielectric constant, width and length of patch as input to compute resonant frequency as output. In the proposed two stage training, an ANN model is trained in two stages. In first stage ANN model is trained with the data generated by empirical formula and in second stage with experimental published data to compute resonant frequency of real antennas. Results shows that the proposed approach give better accuracy in compression with the conventional approaches where models are trained with small data sets.

2. Problem Geometry

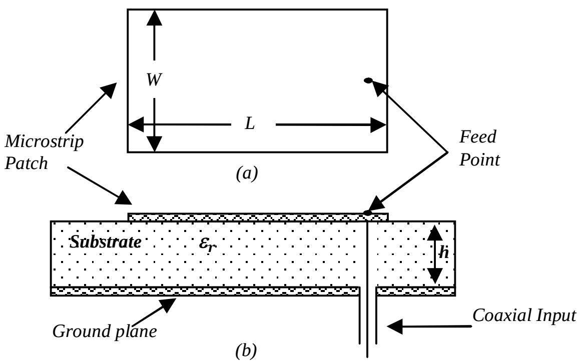


Fig. 1 Geometry of printed rectangular microstrip antenna (a) Top view of microstrip patch, (b) Side view of microstrip antenna with coaxial feed.

Problem geometry has been illustrated in Figure 1**.** A rectangular microstrip antenna of width W and length L printed on substrate of thickness h and relative dielectric constant ε_r. The resonant frequency for different modes of resonance of the printed antenna can be calculated from [1,3].

$$f_{mn} = \frac{c}{2\sqrt{\varepsilon_{eff}}}\sqrt{\left(\frac{m}{L_e}\right)^2 + \left(\frac{n}{W_e}\right)^2} \tag{1}$$

where, f_{mn} is resonant frequency corresponding to mode m and n both belongs to set of positive integers including zero. In the present study only fundamental TM_{10} mode is considered for which *m* is one and *n* is zero. c is speed of electromagnetic wave in free space, L_e and W_e are effective length and width of microstrip patch respectively and ε_{eff} is effective dielectric constant defined as

$$\varepsilon_{eff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\sqrt{\left(1 + 12\frac{h}{W}\right)} \tag{2}$$

The effective length L_e is sum of length of patch L and edge extension length ΔL because of fringing effect can be defined as

$$L_e = L + 2\Delta L \tag{3}$$

and edge extension length ΔL,

$$\Delta L = 0.412h\frac{(\varepsilon_{eff} + 0.3)(W/h + 0.264)}{(\varepsilon_{eff} - 0.258)(W/h + 0.8)} \tag{4}$$

Similarly, one can define effective width W_e.

3. Artificial Neural Networks

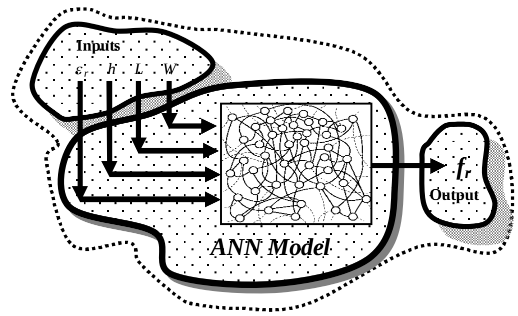


Fig. 2 ANN model with four input parameters and resonant frequency as output.

An Artificial Neural Network (ANN) [5] is a mathematical model which typically emulates certain features of real neural networks found in animal brains. ANN is used in wide variety of problems like information processing, pattern recognition, clustering, classification, image processing and system modeling among others [5, 6]. An ANN models designed in a way in which the animal brain performs a particular task or function of interest with given set of inputs. An artificial neural network has a build in capacity to learn from its environment by undergoing a training session by adjusting its adaptive parameters. According to Hayking [5] a neural network is massive parallel distributed processer that has a nature tendency for storing experiential knowledge and making it available for use. ANN works like an animal brain in two ways, 1) It acquires knowledge through learning and 2) Knowledge is gain or memorized by the means of strength of inter neuron synaptic weights. To realize or design specific function using ANN, large numbers of small processing units known as neurons are used which act as building block of any ANN models. Knowledge is gained by changing the strength of inter neuron synaptic weights and the rule which governs the weight change is known as learning rule or algorithm.

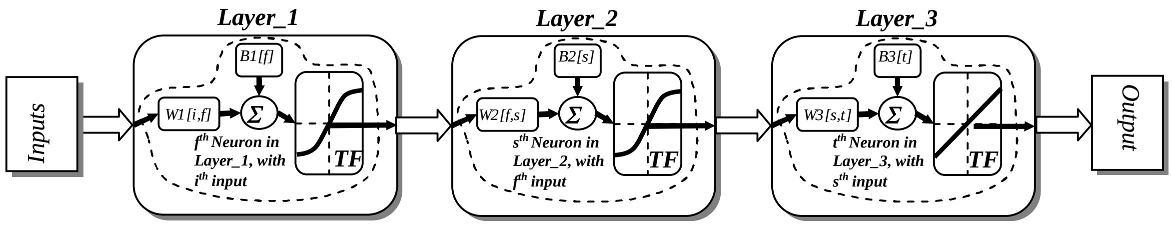


Fig. 3 Final ANN model configuration with two hidden layers and output layer.

4. Two Stage Training

Training is the next step after ANN network is structured for a particular function. In general ANN networks are initialized with random synaptic weight matrix before start of training. There are two main approaches used for training of ANN networks, supervised and unsupervised [5,6]. Supervised approach has been used in present study. In supervised approach network is provided desired output corresponding to the some specific input and these inputs-outputs data pair know as training data pairs or sets. In case of unsupervised approach network has to make its own sense of the inputs without external help. ANN models in supervised learning compares output of models with desired output to compute error or deviation in output, this error is propagated back through the network to adjust synaptic weight matrix in such a way that the error minimizes.

ANN models map input set to output set, these models act as black box which work as function but inside of network is not very clear. Accuracy of model in generating outputs for unseen or test data (data other than used during training) depends largely on training set. To have accurate model training data should be adequate, accurate and uniformly distributed over input range. But in the case of modeling of system where available data is inadequate in size it is very difficult to have accurate ANN models and with small size of training data there are chance of over-learning [5] in which model generate good results corresponding to training data but high errors on test or unseen data. In this paper a new approach has been proposed for training of ANN network when available training sets are of small size. The new approach is named as two stage training. In two stage training, in place of using small size training set directly on ANN model for training, the training is carried out in two stages. In first stage ANN model is trained with data generated from empirical formulas available for system, then the trained ANN model in first stage is used in second stage training to train ANN model second time with available small actual data. Using proposed approach accuracy can be increase with avoiding over-learning problem. Three step algorithm used in the paper is given below:

Algorithm: Two stage training

Step1: Structured ANN model initialize with random synaptic weight matrix.
Step2: Training of model of step1 with data generated from empirical formula.
Step3: Training of trained model in step2 with available experimental or actual data.

In the present study Levenberg-Marquardt algorithm [7,8] is used in step1 and step 2 for training of ANN models. Levenberg-Marquardt algorithm provides a numerical solution to nonlinear function minimizing problems and it is fast and gives stable convergence.

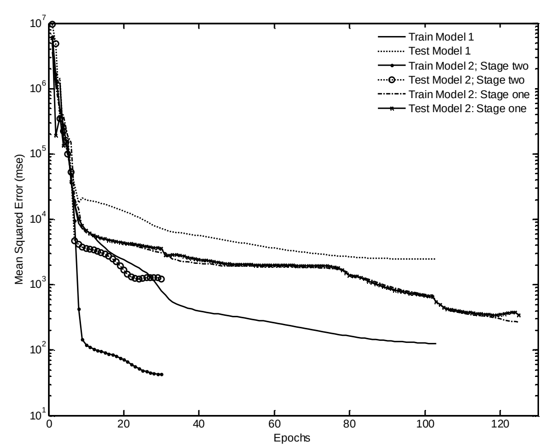


Fig. 4 Performance of ANN models during training.

5. Result and Discussion

For calculating the resonant frequency of rectangular microstrip antenna ANN model is with four inputs and one output as illustrated in Figure 2. Two inputs are substrate parameters that is *dielectric constant* ε_r and thickness of the substrate *h* and other two inputs are parameter of printed microstrip patch length L and width W.

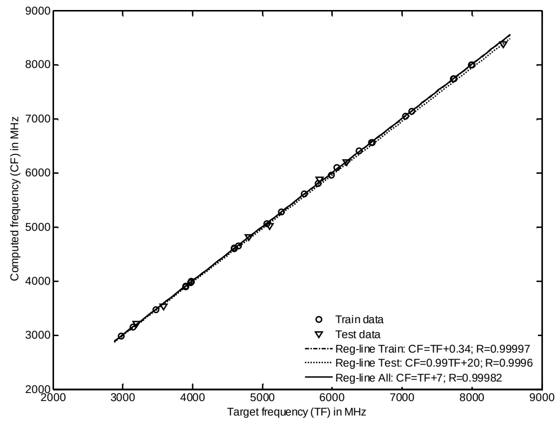


Fig 5. Comparison of computed and target frequency for *Model_1*.

The output of the model is resonant frequency of rectangular microstrip antenna. Hidden layers and the number of neurons in the ANN model has been decided by trial and error approach since there is no straight forward way to determine optimal number of hidden layers and numbers of neurons in corresponding layers. After testing several configurations it is found the most fitting network was $4 \times 4 \times 3 \times 1$. It means 4, 4, 3 and 1 number of neurons in input, first hidden, second hidden and output layers respectively. For hidden layers and output layer the tangent sigmoid and linear activation function was used respectively.

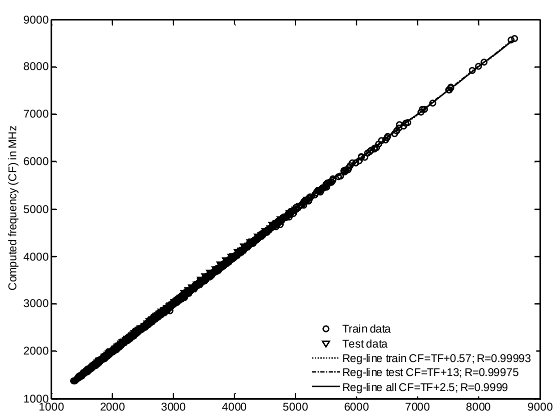


Fig 6. Comparison of computed and target frequency generated from empirical formula for *Model_2* in training stage one.

The final structure of ANN model for the present problem with three layers is shown in Figure 3. In Figure 3 *W1*, *W2*, *W3* are synaptic weight matrix and *B1*, *B2*, *B3* are biasing weight matrix corresponding to three layers. Initially synaptic weight matrix of the ANN model is randomized in range 0 to 1 and this model is referred as *Model_0* untrained model. Then the ANN *Model_0* is trained with two ways, first way referred as *Model_1* in which the ANN model is directly trained with 26 out of 33 sets of experimental data [9,10] given in Table 1. The *Model_1* is conventional way of use of ANN model in microwave device analysis and design [6]. For training Levenberg-Marquardt algorithm is used. *Model_1* is trained with 103 epochs, performance in terms of mean squared error (mse) of *Model_1* during training with epochs is given in Figure 4. In Figure 5 comparison of target (measured or desired) frequency (TF) with computed frequency (CF) and regression (Reg) line equations for training, testing and all data (training+testing) are plotted. In the same figure regression line equations and correlation coefficients R is given, all this proves goodness of model.

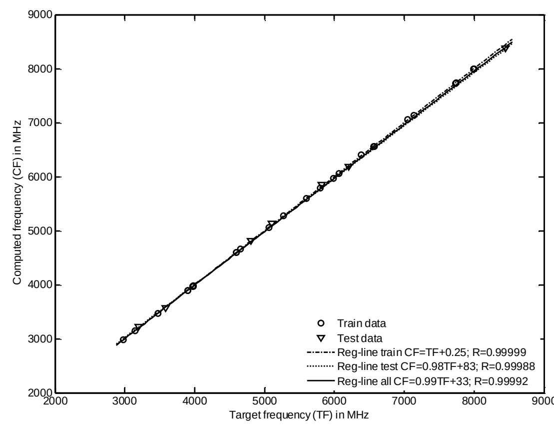


Fig 7. Comparison of computed and target frequency for experimental data [9, 10] for *Model_2* in training stage two.

An innovative training approach has been proposed, named two stage training and the model trained using this approach is referred as *Model_2*. In *Model_2*, in first stage training *Model_0* with same initial synaptic weight matrix which is been used for *Model_1* is trained with the ten thousand set of data generated from empirical formula given in equation (1). For ten thousand sets of data generated by combinations of 10 equally space values of each input and the range of inputs are *dielectric constant* $2.22 \le \varepsilon_r \le 10.20$, thickness of the substrate $0.017cm \le h \le 1.281cm$, length of microstrip $1.0cm \le L \le 3.5cm$ and width $0.776cm \le W \le 2.0cm$. *First stage training is carried out using* Levenberg-Marquardt learning algorithm with 125 epochs performance during training is shown in Figure 4 and comparison of target frequency (TF) with computed frequency (CF) is given in Figure 6. Then the trained ANN model from first stage training is used for second stage training in which 26 out of 33 sets of experimental data

Table 1: Comparison of two stage training with other methods for resonant frequency and the sum of absolute error between experimental and computed frequencies of rectangular microstrip antenna.

MSA No.	Input parameters W (cm)	L (cm)	h (cm)	ε_r	Target Frequency [9],[10]	Proposed two stage training *Model_2*	Using single stage training *Model_1*	Resonant frequencies computed from different models [11] [12]	[13]	[14]	[1]	[2]	[15]	[16]	[17]	[18]	[10]	[19]
1	0.850	1.290	0.017	2.22	7740	7740.73	7738.98	7804	7697	7750	7791	7635	7737	7763	7720	7717	412	7765
2*	0.790	1.185	0.017	2.22	8450	8378.15	8378.23	8496	8369	8431	8478	8298	8417	8446	8396	8389	488	8451
3	2.000	2.500	0.079	2.22	3970	3970.51	3969.54	4027	3898	3949	3983	3838	3951	3950	3917	3887	510	3977
4	1.063	1.183	0.079	2.55	7730	7728.54	7731.01	7940	7442	7605	7733	7322	7763	7639	7551	7376	1610	7730
5	0.910	1.000	0.127	10.2	4600	4599.59	4599.79	4697	4254	4407	4641	4455	4979	4729	4614	4430	113	4618
6	1.720	1.860	0.157	2.33	5060	5059.57	5057.64	5283	4865	4989	5070	4741	5101	4958	4924	4797	1621	5077
7*	1.810	1.960	0.157	2.33	4805	4820.33	4810.33	5014	4635	4749	4824	4520	4846	4724	4688	4573	1460	4830
8	1.270	1.350	0.163	2.55	6560	6563.85	6563.71	6958	6220	6421	6566	6067	6729	6382	6357	6114	2550	6563
9	1.500	1.621	0.163	2.55	5600	5600.13	5605.94	5795	5270	5424	5535	5158	5625	5414	5374	5194	1769	5535
10*	1.337	1.412	0.200	2.55	6200	6186.56	6198.19	6653	5845	6053	6201	5682	6413	5987	5988	5735	2860	6193
11	1.120	1.200	0.242	2.55	7050	7056.47	7046.91	7828	6566	6867	7052	6320	7504	6682	6769	6433	4792	7030
12	1.403	1.485	0.252	2.55	5800	5793.09	5799.13	6325	5435	5653	5801	5259	6078	5552	5586	5326	3259	5787
13	1.530	1.630	0.300	2.50	5270	5276.32	5269.99	5820	4943	5155	5287	4762	5572	5030	5081	4842	3383	5273
14	0.905	1.018	0.300	2.50	7990	7993.55	7991.80	9319	7334	7813	7981	6917	8885	7339	7570	6822	8674	8101
15	1.170	1.280	0.300	2.50	6570	6562.13	6554.02	7412	6070	6390	6550	5794	7076	6135	6264	5951	5486	6543
16*	1.375	1.580	0.476	2.55	5100	5138.77	5025.30	5945	4667	4993	5092	4407	5693	4678	4830	4338	5437	5193
17	0.776	1.080	0.330	2.55	8000	7991.98	7999.21	8698	6845	7546	7519	6464	8447	6889	7160	6367	8067	7948
18	0.790	1.255	0.400	2.55	7134	7138.02	7129.65	7485	5870	6601	6484	5525	7342	5904	6179	5452	7242	7169
19	0.987	1.450	0.450	2.55	6070	6063.27	6095.75	6478	5092	5660	5606	4803	6317	5125	5341	4735	6103	6026
20*	1.000	1.520	0.476	2.55	5820	5851.25	5884.36	6180	4855	5423	5352	4576	6042	4886	5100	4513	5875	5817
21	0.814	1.440	0.476	2.55	6380	6403.24	6402.49	6523	5101	5823	5660	4784	6453	5122	5396	4729	6546	6515
22	0.790	1.620	0.550	2.55	5990	5972.68	5953.36	5798	4539	5264	5063	4239	5804	4550	4830	4196	5976	6064
23	1.200	1.970	0.626	2.55	4660	4664.26	4652.85	4768	3746	4227	4141	3526	4689	3770	3949	3479	4600	4613
24	0.783	2.300	0.854	2.55	4600	4598.97	4614.29	4084	3201	3824	3615	2938	4209	3168	3446	2921	4603	4550
25*	1.256	2.756	0.952	2.55	3580	3576.74	3531.49	3408	2668	3115	2983	2485	3430	2670	2845	2461	3574	3628
26	0.974	2.620	0.952	2.55	3980	3973.78	3976.60	3585	2808	3335	3162	2590	3668	2790	3015	2572	3955	3956
27	1.020	2.640	0.952	2.55	3900	3895.89	3900.10	3558	2785	3299	3133	2573	3629	2771	2987	2555	3895	3907
28	0.883	2.676	1.000	2.55	3980	3987.09	3989.67	3510	2753	3294	3112	2522	3626	2721	2966	2509	3982	3922
29	0.777	2.835	1.100	2.55	3900	3896.19	3887.43	3313	2608	3147	2964	2364	3473	2554	2823	2356	3903	3747
30	0.920	3.130	1.200	2.55	3470	3470.13	3467.71	3001	2358	2838	2675	2146	3129	2317	2549	2137	3493	3381
31*	1.030	3.380	1.281	2.55	3200	3223.64	3213.48	2779	2183	2623	2474	1992	2889	2151	2357	1983	3197	3123
32	1.265	3.500	1.281	2.55	2980	2977.00	2980.51	2684	2102	2502	2370	1936	2752	2086	2259	1924	2982	2972
33	1.080	3.400	1.281	2.55	3150	3151.93	3153.58	2763	2168	2600	2453	1982	2863	2139	2338	1972	3160	3096
		Absolute Errors (in MHz):				327.09	460.00	13136	24097	11539	12322	30669	8468	22572	18148	30504	56698	1393

* Test data set and unmark MSA no. are training data set.

[9,10] given in Table 1 is used with Levenberg-Marquardt learning algorithm for 29 epochs for optimal learning and avoiding over-learning. Figure 4 shows performance and Figure 7 with regression (Reg-) line equations and correlation coefficients R compares computed frequency (CF) with target frequency (TF).

The efficiency of two stage training over single stage training and other conventional methods with published results [11] comparisons are made in Table 1. In Table 1 the proposed approach is compared with results from conventional methods [1, 2], [10], [12]-[19]. It is clear that in the proposed method sum of absolute error is smallest. It is also proved from the Table 1 that accuracy in computation of resonant frequency of microstrip antenna increase by considerable amount in case of two stage training in comparison with single stage training for same initial synaptic weight matrix. Goodness of models can be compared from Figure 5 and 7, correlation coefficients R between target frequency (*TF*) and computed frequency (*CF*) in case of *Model_1* based on single stage training is 0.99997 for training data, 0.9996 for test data and 0.99982 for all 33 (training + test) data set where as in case of *Model_2* based on two stage training correlation coefficients is more closer to one, 0.99999 for training data, 0.99988 for test data and 0.99992 for all 33 (training + test) data set. From above results it is clear that two stage training provides high degree of accuracy in computing resonant frequency and superior to single stage training and other conventional approaches.

6. Conclusions

Resonant frequency of rectangular microstrip antenna has been computed using novel two stage training approach on ANN model. ANN model is used as black box to generate resonant frequency as output corresponding to four input height of the substrate with dielectric constant and dimensions of rectangular patch antenna. In the proposed approach in first stage the ANN model trained with empirical relation of resonant frequency with structural and substrate parameters of antenna then in second stage the model has been trained with the actual experimental data. The results of the proposed approach are in very good agreement with the measurements and shows better accuracy than other traditional approaches. The proposed approach offers an accurate and efficient alternative method for the calculation of resonant frequency of antenna. This approach is not limited to the rectangular microstrip antenna it can be easily applied to other antenna and microwave circuit problems. The high-speed real-time computation feature of the proposed approach recommends its use in computer-aided design programs. It is expected that the hybrid approach will find potential application area in electromagnetic engineering and device designs.

Acknowledgments
The authors are grateful to Professor V. G Das, Director of the institute and Professor G. S. Tyagi, Head Department of Physics and Computers Science for providing necessary facilities, continuous cooperation and encouragement.
Dr. Ashok Jangid is also thankful to UGC-BSR, Start-Up-Grant for providing financial support.

References

[1] I. J. Bahl and P. Bhartia, Microstrip Antennas. Dedham, MA: Artech House, 1980.
[2] J. R. James, P. S. Hall and C. Wood, Microstrip Antennas-Theory and Design. London, U.K: Peregrinus, 1981.
[3] R. Garg, P. Bharti, I. Bahl, and A. Ittipiboon, Microstrip antenna design handbook. London: Aartech House, 2001.
[4] L. A. Zadeh, in Proceedings of Second International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, 1992, pp. XIII-XIV.
[5] S. Haykin, Neural networks - a comprehensive foundation. New Delhi : Prentice Hall of India, 2004.
[6] Q. J. Zhang and K. C. Gupta, Neural Networks for RF and Microwave Design. Boston, MA: Artech House, 2000.
[7] K. Levenberg, A method for the solution of certain problems in least squares, Quarterly of Applied Mathematics, vol. 5 (1944), pp. 164-168.
[8] D. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, SIAM Journal on Applied Mathematics, vol. 11 (1963), no. 2, pp. 431-441.
[9] M. Kara, The resonant frequency of rectangular microstrip antenna elements with various substrate thicknesses, Microwave Opt. Technol. Lett., vol. 11 (1996), pp. 55–59.
[10] M. Kara, Closed-form expressions for the resonant frequency of rectangular microstrip antenna elements with thick substrates, Microwave Opt. Technol. Lett., vol. 12 (1996), pp. 131–136.
[11] K. Guney and N. Sarikaya, A hybrid method based on combining artificial neural network and fuzzy inference system for simultaneous computation of resonant frequencies of rectangular, circular and triangular microstrip antennas, IEEE Trans. Antennas Propag., vol. 55 (2007), no.3, pp. 659-668.
[12] J. Q. Howell, Microstrip antennas, IEEE Trans. Antennas Propag, vol. 23 (1975), pp. 90-93.
[13] E. O. Hammerstad, Equations for microstrip circuits design, In Proc. 5th Eur. Microw. Conf., Hamburg, Germany, 1975, pp. 268–272.
[14] K. R. Carver, Practical analytical techniques for the microstrip antenna, In Pro. Workshop Printed Circuit Antenna Tech, Las Cruces, NM, 1979, pp. 7.1-7.20.
[15] D. L. Sengupta, Approximate expression for the resonant frequency of a rectangular patch antenna, Electron. Lett., vol. 19 (1983), pp. 834–835
[16] R. Garg and S. A. Long, Resonant frequency of electrically thick rectangular microstrip antennas, Electron. Lett., vol. 23 (1987), pp. 1149–1151.
[17] W. C. Chew and Q. Liu, Resonance frequency of a rectangular microstrip patch, IEEE Trans. Antennas Propag., vol. 36 (1988), pp. 1045–1056.
[18] K. Guney, A new edge extension expression for the resonant frequency of electrically thick rectangular microstrip antennas, Int. J. Electron., vol. 75 (1993), pp. 767–770.
[19] K .Guney, A new edge extension expression for the resonant frequency of rectangular microstrip antennas with thin and thick substrates, J. Commun. Tech. Electron., vol. 49 (2004), pp. 49–53.

A Model to Find Outliers in Mixed-Attribute Datasets using Mixed Attribute Outlier Factor

M. Krishna Murthy, A. Govardhan, Lakshmi SreenivasaReddy D

[1] Research scholar, ANU, Guntur, India

[2] Professor of CSE, JNTUH, Hyderabad, India

[3] Department of CSE, Rise Gandhi Groups of Institutions, Ongole, India.

Abstract

Outliers are records in real datasets which have abnormal behavior comparing with other records in datasets. Finding outliers in numerical dataset is easy. Many methods are available for numerical datasets. Number of methods is also available for categorical datasets. But very less number of methods is available for mixed attribute Datasets. In all the available methods, the concept of frequent pattern mining is used. Finding different frequent patterns from datasets for the categorical attributes is a cumbersome process. In proposed model, Mixed Attribute Outlier Factor (MAOF) is presented. Which is a simple technique and it requires only one scan of dataset. MAOF is derived based on Attribute Value frequency (AVF) for Categorical part of dataset and cosine factor for continuous part which is derived from the mean record to the remaining numerical data points in the dataset. Average of these two factors will give the MAOF score. This model has been applied on Bank dataset which is a real dataset taken from UCI ML repository [10]. This method shows the good results.

Keywords: ***Data mining, Outlier detection, Oteyscores, ODMADscore, MAOF score.***

.

1. Introduction

Most of the real datasets contain mixed type of attributes. Finding outliers of this type of datasets is very useful to model the data. Outlier analysis is an important task in many fields like medicine, bank, and networks. Existing systems concentrated on finding frequent patterns. Outlier factors are found by the frequent patterns gives us a very high complexity. There are so many methods derived from frequent mining concept. Some methodologies are derived based on Appriori property to reduce the complexity [14]. By this approach forming the number of subsets for each record and scanning the dataset for all these subsets for frequency is a problem. Even utilization of Appriori concept does not in prove efficiency. The proposed model solves all these problems. This proposed method utilized the concept of attribute value frequency for categorical part of dataset and cosine vector product concept for numerical part of records in datasets.

2. Terminology

Different terminologies are required for existing and proposed model about frequency, support and input number of outliers etc is given in Table.1 below.

Table.1. Terminology used in this paper.

Term	Description
K	Target number of outliers
N	Number of objects in Dataset
M	Number of Attributes in Dataset
xi	ith object in Dataset ranging from 1 to n
Aj	jth Attribute ranging from 1 to m
D(Aj)	Domain of distinct values of jth attribute
xij	cell value in ith object which takes from domain dj of jth attribute Aj
D	Dataset
V	Set of all distinct values in Dataset D
P	Set of all combinations of distinct attribute values, where each attribute occurs only once in any combination
I	Item set

F	Frequent Item set
IF	Infrequent item set.
f(xij)	Frequency of xij value
FS(xi)	Set of frequent Item sets of xi object
IFS(xi)	Set of infrequent Item sets of xi object
Minsup	Minimum support of frequent item set
Support(I)	Support of Item set I

3. Existing approaches for mixed attribute datasets based on Item frequency

3.1 Otey Score Algorithm

In this approach anomaly score is computed by partitioning the entire mixed dataset in to two parts. First part contains categorical subspace and the second part contains numerical subspace. Outlier factor of categorical part is denoted as Score1 (xi) and outlier score of numerical part is denoted by Score2 (xi). This approach is described based on links between attributes.
This approach is derived as below:
Let V= set of all distinct categorical values included in Dataset
C= Set of all combinations (itemsets I) of distinct attribute values
i.e $P = FS \cup IFS$ (1)
Where FS = Set of Frequent Itemsets such that sup (Itemset)$\geq$user defined threshold value.
IFS= Set of infrequent Itemsets
Sup (I) = support of itemset I
Now the outlier score of the categorical part is defined as below:

$$Score1(C(xi)) = \sum_{I \in FS(x_i)} \frac{1}{|I|} \quad (2)$$

Let CI is the covariance of the itemset I,
CIij is the covariance of I in 'i' and 'j' attributes from numerical part
CxiIij is the covariance of I in 'i' and 'j' attributes from numerical part for the object xi

$$C_{ij}^{xiI} = (x_i - \mu_i^I)\, x\, (x_i - \mu_i^I) \quad (3)$$

The violation score of an object xi is defined as below:

$$V_\Gamma(x_i) = \sum_i \sum_j v_\Gamma(x_{ij}) \quad (4)$$

$$v_\Gamma(x_{ij}) = \begin{cases} 1 & if\ |C_{ij}^{xiI} - C_{ij}^{I}| \leq \Gamma \\ 0 & otherwise \end{cases} \quad (5)$$

$$\sigma_{C_{ij}^I}^2 = \frac{1}{\sup(I)-1} \sum_{\{xi/I \in xi\}}^{n} (C_{ij}^{xiI} - C_{ij}^{I})^2 \quad (6)$$

Where $\dfrac{C_{ij}^{xiI} - C_{ij}^{I}}{\sigma_{C_{ij}^I}}$ follows the normal distribution
Now the outlier score of xi in a dataset

$$Score2(x_i) = \sum_{I \subseteq x_i} \left(\frac{1}{|I|} / (C_1 \vee C_2) isTrue \right. \quad (7)$$

Where C1: sup (I) $\leq$ threshold value C2: sup (I)> δ, where δ is maximum violate score.
Based on these two scores we can find outlier factor of an object in mixed attribute dataset. Another approach of finding outlier factor for every object in mixed type of Dataset is defined by Koufakou et al in [3].

3.2 ODMAD Score

This algorithm also depends on two parts of mixed data. The first part is categorical subspace; second part is numerical subspace of the dataset.

$$Score1(xi) = \sum_{|IF(xi)|=1}^{MAXLEN} \frac{1}{\sup(IF(xi)) * |IF(xi)|} \quad (8)$$

MAXLEN = user entered maximum length of in frequent itemset,
Sup (IF (xi) = support of infrequent itemset in object xi,
$|IF(xi)|$ =length of infrequent itemset in object xi,

$$Score2((xi) = \frac{1}{|a \in xi^C|} \sum_{\forall a \in xi^C} COS(xi^N, \mu_a) \quad (9)$$

Where

$$\mathrm{COS}(xi^N, \mu_a) = \sum_{j=1}^{m_N} \frac{1}{\|x_{ij}^N\|} * \frac{\mu_{aj}}{\mu_a} \quad (10)$$

Here 'a' is a categorical value included in the object xi.
Based on the above scores the outlier factor is found in ODMAD.In both approaches finding frequent itemsets is a big problem. So we approached the below way.

3.3 Proposed Method

In this approach outlier factor is found with forming any frequent patterns in an object. Instead of this the attribute value frequency has been proposed. From the above two approaches number of scans of a dataset is required. Proposed method needs only on scan of the dataset. This proposed method finds again two scores, one is for

categorical part of dataset and other is for numerical part of the dataset. Score1 is defined like below:
Let there are 'm' categorical attributes and 'n' numerical attributes in a dataset.

$$Score_1(C(x_i)) = \sum_{j=1}^{m} \frac{\sup(x_{ij})}{|D|} \quad (11)$$

$$Score_2(N(x_i)) = COS(N(x_i) = \frac{\langle \mu_{xiN}, x_N \rangle}{\|\mu_{xiN}\| * \|x_N\|} \quad (12)$$

Here μ_{xiN} is a vector of means of all attributes in Numerical part of Dataset.

x_{iN} is the vector of all attribute values in the numerical part of the ith object.

MAOF factor can be defined as

$$MAOFscore(x_i) = \frac{Score_1(C(x_i)) + Score_2(N(x_i))}{2} \quad (13)$$

4. Experimental Results

Experiments are conducted on 1342, 1298, 1279, 1200 records respectively from 1-in-2, 1-in-5, 1-in-8, 1-in-10 samples are taken from Bank dataset which is taken from UCI machine Repository [10]. Different sample are selected from Bank date with two class values. These samples are selected like that one sample from each two records, one sample from each 5 records, one sample from each 8 records, and one sample from each 10 record from "yes" class records. These are mixed-up with normal records. Then 45 records are created randomly with much variation and mixed up with the above said samples. All these operations are conducted by Clementine 11.1 tool. Our model found these created records from each sample as given in the below Table.1.1

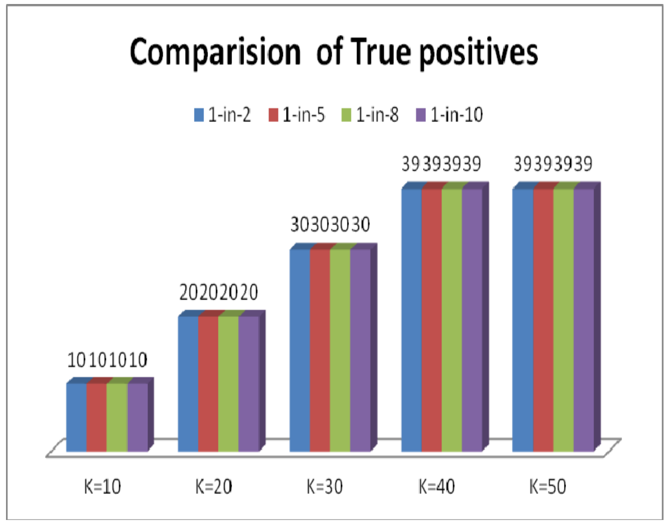


Fig.1 Comparison of true positives found in different samples for different inputs

Table.1 Comparison of the number of outliers found for different input 'K'

Sample	Number of true and false outliers are found for different K values given as input									
	K=10		K=20		K=30		K=40		K=50	
	True	False	True	False	True	False	True	False	True	False
1-in-2	10	0	20	0	30	0	39	1	39	11
1-in-5	10	0	20	0	30	0	39	1	39	11
1-in-8	10	0	20	0	30	0	39	1	39	11
1-in-10	10	0	20	0	30	0	39	1	39	11

From the Table 1 it is shows that for inputs k=10, k=20, k=30 this model found exact number of true positives and for k=40 and for k=50 it has found 39 true positive out of 45 outliers which are included in Bank Dataset. The first score computes how much an object deviates from the others in categorical part and the score computes cosine product which calculates the similarity between mean for numerical part and an object in the Numerical part of each object. These results are not compared with others because the approach is entirely different with the existing ones.

5. Conclusion and Future Work

This model has been developed on Attribute Value Frequencies for categorical data and cosine dot product for numerical dataset. These two scores give the factors in the range of 0 to 1 and its average again gives the value in the range of 0 to 1. In future we investigate the numerical score by correlation factor between mean of the attribute values in numerical part and the numerical part in each object.

References

[1] Anna Koufakou, Michael Georgiopoulos, “A fast outlier detection strategy for distributed high-dimensional data sets with mixed attributes” Data mining and Knowledge Discovery , Volume 20, 2010, pp.259-289.

[2] M.E. Otey, A. Ghoting, and S. Parthasarathy. “Fast Distributed Outlier Detection in Mixed-Attribute Data Sets." Data Mining and Knowledge Discovery, 12, 2006, 203-228.

[3] A. Koufakou, E.G. Ortiz, M. Georgiopoulos, G.C. Anagnostopoulos, and K.M.Reynolds. “A Scalable and Efficient Outlier Detection Strategy for Categorical Data.” IEEE International Conference on Tools with Artificial Intelligence ICTAI, 2007, pp. 210-217.

[4] Z. He, X. Xu, S. Deng, D. Calvanese, G. De Giacomo, and M. Lenzerini. “A Fast Greedy Algorithm for Outlier Mining.” Proceedings of 10th Pacific-Asia Conference on Knowledge and Data Discovery, 2006, pp. 567-576.

[5] Z. He, X. Xu, J.Z. Huang, and SC Deng.“FP-Outlier: Frequent Pattern Based Outlier Detection.” Computer Science and Information System, valume2, 2005, pp.103-118.

[6] Z. He, X. Xu, J.Z. Huang, and SC Deng. “FP-Outlier: Frequent Pattern Based Outlier Detection.” Computer Science and Information System, Volume2, No.1, 2005, pp103-118.

[7] J. Han, J. Pei, Y. Yin, and R. Mao. “Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach.” Data Mining and Knowledge Discovery, Volume8, No1, 2004, pp53-87.

[8] S.D. Bay and M. Schwabacher. “Mining distance-based outliers in near linear time with randomization and a simple pruning rule.” Proceedings of the ACMSIGKDD International conference on Knowledge Discovery and Data Mining, 2003, pp. 29-38.

[9] C. Borgelt. “Efficient Implementations of Apriori and Eclat.” Workshop of Frequent Item Set Mining Implementations FIMI, 2003.

[10] L. Wei, W. Qian, A. Zhou, W. Jin, and J.X. Yu. “HOT: Hyper graph-Based Outlier Test for Categorical Data.” Proceedings Pacific-Asia Conference PAKDD, 2003, pp. 399.

[11] T. Calders and B. Goethals. “Mining All Non-Derivable Frequent Itemsets.” Proceedings PKDD International Conference Principles of Data Mining and Knowledge Discovery, 2002, pp.74-85.

[12] F. Pan, G. Cong, A.K.H. Tung, J. Yang, and M.J. Zaki. “CARPENTER: Finding closed patterns in long biological datasets.” Proceedings ACM SIGKDD International Conference on Knowledge Discovery and Data mining, 2003, pp. 637-642.

[13] He, Z., Deng, S., Xu, X., “A Fast Greedy algorithm for Outlier mining” Proc. of PAKDD, 2006.

[14] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[15] P. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining: Pearson Addison-Wesley, 2005

[16] E. Knorr, R. Ng, and V. Tucakov, "Distance-based outliers: Algorithms and applications," VLDB Journal, 2000.

[17] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density based local outliers," presented at ACM SIGMOD International Conference on Management of Data, 2000

[18] S. Papadimitriou, H. Kitawaga, P. Gibbons, and C. Faloutsos, "LOCI: Fast outlier detection using the local correlation integral," presented at International Conference on Data Engineering, 2003

[19] Z. He, X. Xu, J. Huang, and S. Deng, "FP-Outlier: Frequent Pattern Based Outlier Detection”, Computer Science and Information System (ComSIS'05)," 2005

[20] Shu Wu and Shengrui Wang, “Information-Theoretic Outlier Detection for Large-Scale Categorical Data, IEEE Transactions on Knowledge Engineering and Data Engineering,2011

[21] Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[22] LakshmiSreenivasaReddy.D, .B.RaveendraBabu and A.Govardhan, “Outlier Analysis of Categorical Data using NAVF”’, Informatica Economica vol 17, Cloud computing issue 1, 2013.

[23] LakshmiSreenivasaReddy.D, B.RaveendraBabu “Outlier Analysis of Categorical Data using FuzzyAVF”, presented at IEEE international conference ICCPCT-2013, pp 1259-1263.

[24] LakshmiSreenivasaReddy.D, B.RaveendraBabu and A.Govardhan, “A Novel Approach to Find Outliers in Categorical Dataset” presented at Elsevier AEMDS-2013

[25] R. Agrawal and R. Srikant. "Fast Algorithms for Mining Association Rules in Large Databases." Proceedings International Conference on Very Large Data Bases, 1994, pp. 487-499.

[26] LakshmiSreenivasaReddy.D, .B.RaveendraBabu and A.Govardhan, "A model for Improving Classifier Accuracy for Categorical data using Outlier Analysis", International Journal of Computers and Technology" Volume 7, 2013.

THE AUTHORS

Mr. Mudimbi.Krishna Murthy did His M.C.A in first Class in 2003 from MKU Madurai. He has 15 years of technical experience in Computer Science and Engineering at School of Information Technology (SIT), Jawaharlal Nehru Technological University Hyderabad, India. He has five research papers at international and national conferences. His area of research is Data Mining and Information Retrieval Systems.

Dr.A.Govardhan is presently a Professor of Computer Science & Engineering, Jawaharlal Nehru Technological University Hyderabad (JNTUH), India. He did his B.E (CSE) from Osmania University College of Engineering, Hyderabad in 1992, M.Tech from Jawaharlal Nehru University (JNU), New Delhi in 1994 and Ph.D from Jawaharlal Nehru Technological University, Hyderabad in 2003. He is a recipient of several International and National Awards including A.P. State Best Teacher Award, Bharat Seva Ratna Puraskar, CSI Chapter Patron Award, Bharat Jyoti Award and Mother Teresa Award for Outstanding Services, Achievements, Contributions for Meritorious Services, Outstanding Performance and Remarkable Role in the field of Education and Service to the Nation. He is a Chairman and Member on several Boards of Studies of various Universities. He is the Chairman of CSI Hyderabad Chapter. He is a Member on the Editorial Boards for Eight International Journals. He is Member of several Advisory Boards and Committee Member for several International and National Conferences. He has guided 15 Ph.D theses and he has published 152 papers at International/National Journals/Conferences including IEEE, ACM, Springer and Elsevier. He has delivered more than 35 Keynote addresses and invited lectures. He served as Principal, Head of the Department and Students' Advisor. He is a member in several Professional and Service oriented bodies. His areas of research include Databases, Data Warehousing & Mining and Information Retrieval Systems.

Mr.LakshmiSreenivasareddy.D obtained his Masters degree from Jawaharlal Nehru Technological University Hyderabad (JNTU). He is pursuing his Ph.D in Computer Science and Engineering from JNTUH, Hyderabad. He is currently heading the Department of Computer Science & Engineering, RISE Gandhi Groups of Institutions Ongole. He has 10 years of teaching experience. He has six research papers at international conferences and journals including IEEE and Elsevier. His area of interest is Data Warehousing & Mining and Information Retrieval Systems.

Using Discrete Wavelet Transformation To Enhance Underwater Image

Ali A. Yassin[1], Rana M. Ghadban[2], Salah F. Saleh[1], Hikmat Z. Neima[2]

[1] Basrah University, Education College, Computer Department, Basrah, Iraq

[2] Basrah University, Science College, Computer Department, Basrah, Iraq

Abstract

The underwater image suffers from two main problems. The first is light absorption and the second is the losing colors. These two problems cause losing the basic features of images. A new active approach is presented in this paper to solve the mentioned problems, which depends *Discrete Wavelet Transform* (DWT), *Hue saturation value* color space (HSV), and Slide Stretching. Firstly the underwater image is transformed to the HSV color space, then DWT is applied on the layer H of this image which divides this layer to the four parts according to the importance of the data. Secondly, the part which has much data (called *Low-Low* (LL)) is used with the slide stretching to increase the clarity of color (lost color compensation), on the other hand it provides a suitable solution for the light absorption. The proposed approach has accomplished high processing performance time through focusing on the important part of image by using the DWT. Besides, we did an enhancement in the image that had lost its data and at the same time there are no sources of the image to compare with it. Along with, we noted that the amount of compensation up to its highest level. We gained high ratio of the image enhancement which succeeds up to 98.08%. In addition, the performance time decreases and the compensation ratio of the enhanced image is increases.

Keywords: *DWT, HSV, Underwater Image, Enhancement.*

1. Introduction

The attention of image processing is directed into two fundamental fields: change for the better in figurative information for human visual system as connotation; and manipulating of digital image for storage in outsource such as cloud computing environment, transmission via network, and representation as automatically manner in computers. Image enhancement is a big-name and inevitable grade in several uses of digital image-processing, it has many applications and techniques which are used to get better detail of processed an image. We know that water is denser than air. In fact, it is about 800 times denser merely when the light enters the water. It interacts with the molecules of the water. The atoms await to cause a decrease of light, waste of contrast, changes in color, prevalence, and others negative impacts. The effect of this operation causes losing the light, idling of contrast, changes in color, diffusion and other water impacts. These major problems are light absorption and idling of colors. Absorption impacts several conditions such as the position on the map of the Earth, daily timing, and any season of the year; clarity, and hollowness of the water to acquire the image. The amount of the light reflects upside depending on the height of the sun's light, strictly and even more, under normal circumstances of the sea. In the case of a stormy sea, a lot of light will be distracted. Here the sea can do as of mirror to reflect a lot of light. Depths play a major role in influencing the lighting. Scientific base proves that the loss of light is increased to half whenever we went to the rear more. (In other words, in 10% there will be 50%.,.. , and 12,5% As in a lot of light 30 m , etc). [1, 2] The other dilemma is the dispersion of light. Water molecules react with light by overlapping certain wavelengths. Proportionally, the absorption of light with the wavelength is inversely proportionate. Whereas the short wavelength is moved in the water as long form of colors which are transmitted in batches. The first batches are the red and orange colors, and then it starts with the following colors: yellow, green, and purple, to emerge. The final color is blue. The color that is quickly missing is red, where it falls down at 50 meters. The ratio of loss is

90% in the level of 5 meters. The main scale of idling the color is the distance. Therefore, there are various ratios of them. Fig.1 views the losing of colors [2]. Underwater image grapples with finite range, non uniform lighting, decreasing contrast and hidden color because of specific conversion properties of light in the sea. A significant factor that must be processed is the light alleviating. Light alleviating detects the clarity distance to about 20 meters in visible water. As for the stained water, it is located about 5 meters or less [3, 4]. The problem of lack of lighting is caused by absorption and dissipation. In addition to that, it obtains excitement whenever we move deep in the sea. Absorption deletes the light's energy whereas the scattering inflects the steering of the light's path. Absorption and Scattering infections are suitable to the sea water itself and to other elements such as small observable floating molecules. Another issue in the underwater milieu is the little color differences [4, 5, 6]. This issue causes the color falling to rely on the length of wave of the colors [6, 7]. Enhancement of true images addresses a more complex process not only because of the embedded dimension of the data. However, it refers to the inserted hardness of color impression. Furthermore, underwater image enhancement may need refinement of color counterpoise, color contrast, color brightness in a truth image. The scientific area is witnessing a significant wide development especially in the field of informatics. The underwater image is miserable because of major ordeals such as light absorption and lacking of color when it is acquired at detected depths. As a result of the tremendous scientific advances in the field of information processing, the digital processing of the image was a large shared of it. The digital image processing field is one of the most developing fields of the computer world. The anxiety of most researchers in this field has been to enhance the quality of the images or to calculate some information that intends to obtain lost information throughout complicated provisos such as sight. Image processing in some implementations need an enhancement operation to vivid the image. Image enhancement is considered a significant assignment in the image processing domain. It develops the quality of the images for the human visual system [8, 9]. Removing blurriness, noise, scaling up contrast, and uncovering details are examples of enhancement processes. In other words, it compensates the wastage of the image's information. The admission of information is due to accidental circumstances through image acquiring. The noble goal of the proposed work is to process any picture to manufacture a well image better than the original. The complicity of the image enhancement relies on the ambiance of the image. One of the most complication is the natural of water [10, 11]. Additionally, the source of difficulty in this type of processing is not finding any destination image for compare with it. Therefore, we depended on histogram scale to know enhanced ratio in brightness and contrast in images. Additionally, we used correlation measure to access real value of all components of image. This paper received the gains through the implementation of wavelet transformation for analyzing and de-noising the image and then getting rid of underwater image challenges. In our proposed scheme, repaired color, increase intensity and brightness of underwater image enhancement based on DWT. The mechanism work of our approach, hue is repaired in order to restore color distortion. The reconstructed color is achieved by applying DWT on hue layer. The adaptive luminance enhancement is used to increase the true color and offer solution of light problem, this process is performed via the saturation and intensity stretching of HSV. Our paper is detailed underwater image enhancement and we have used the scale approaches for explaining mechanism of loses color and absorption light. The varieties of approaches can degrade an underwater image into components, which can be applied to know quality of improve contrast; brightness and substitution of lose formation in the image. We are depend on image histogram and correlation as scales to present work, due to it does not find any source image such as previous works. Therefore, Any person can notes amount ratio of enhancement these scales. The underwater image enhancement by using wavelet is quite method to solve problems in this type of digital image.

The organization of this paper includes a series of sections that display the following form . The section 2 addresses and views a summary of the related work. In Section 3, we introduce the primitives and requirements of our work. Section 4 contains the mechanism proposed in this paper. Experimental Results and Analysis have limited in the section 5. The conclusion is referred in the section 6.

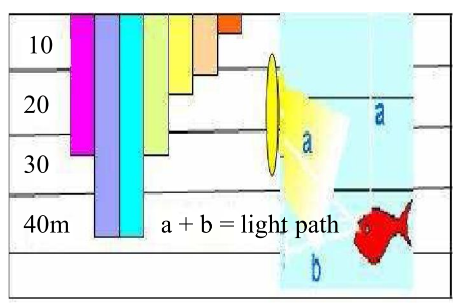


Fig. 1 Loss of Colors

2. Related Works

Often, the data of the image captured is exposed to extreme difficulties. We must, at the outset, address a prior preliminary step. In addition to that, we need to analyze it. In effect, the problems attached of the images underwater is caused by the overlapping of light with the region of absorption. Eventually, the light will disperse. It is activated through the bottom the contrast, thinning luminosity, and limited visibility. There is a need for processing image enhancement impacts. During damping, due to redress of light and restoring, a better balance between the color's components of the color image occurs. Some ways to improve the image retroactive are discovered and are based on responsive phenomenon light impairment [12]. Tristan et al., has suggested a new approach based on the promotion of the video frame sequence to detect different types of fish in the nature of underwater. They applied the equation of the histogram to duplicate the contrast so it can be clearer At the same time, it is used for the median filter to remove the noise. We present a method promoted by slide stretching with DWT. Our proposed scheme does not require median filter because it obtained enhancement ratio is better. The edge detection algorithm was used to disclose Sobel background [13]. Stephane et al., proposed an automated algorithm to reprocess underwater images. The turmoil has decreased under the water, and increased in picture quality. It consists of many sequential phases of processing, which corrects the illumination autonomous non-uniform, lowers noise, increase contrast, and improves color. They used the wavelet transformation to analyze the image. Also, they extended the contrasting for the variation in color and graphed the equation to correct the colors [5]. Our work does not like as above methods, our method based on HSV color system and we selected important layer data (LL) to represent in slide stretching which increase accuracy and performance. Jyoti et al., provided a means of image processing to improve diversity in slow motion underwater, earth, and space images. They proposed, having polished noise that extends on the conversely, they can change the level of brightness for better contrast by using the Histogram Equation (HE) [12]. Our suggested method does not need this service, due to having the privilege gains ability to detail with all types diversity of brightness. Kashif et al., proposed a new approach which depends mainly on slide stretching. This method is a double. First, they use contrast stretching from the RGB method to neutralize color contrast in the image. In addition to that, they applied the stretching of the HSI into the other layers which are intensity and saturation. This increases the true color and presents a solution to the issue of lighting [6]. Liu Chao et al., offered to affect the murky water that can be disabled to detect the clarity of the original vision of the image [14]. Our proposed uses all layers of HSV in easy manner during merge DWT with slide stretching. At last, we got good results in time processing and accuracy of enhancement. The DWT is very quite method to reduce time series. Gang Song et al. suggested an adaptive truth image enhancement depend on human visual features in HSV color system. The truth images enhanced by their algorithm have many features such as richer color, more visible details, and better visual impact. This algorithm is relied on mathematical mean and variance computation. We know the mathematic mean filter leads to miss significant image details like edges and sharpness [15]. The proposal approach enjoys splendid performance in color compensation, regular contrast, brightness balancing and computational cost. G. Padmavathi et al. submitted a set of techniques for the nomination of the image which are deployed in earlier works as a preprocessing of the image. They have been using natural filters to improve the quality of the image, suppression the noise, preserve edges in the image, and increase the smoothness of the image. They compared the experimental results of the current method with three more popular filters: the homomorphic filter, the filter of spread of anisotropic, and the removing noise wavelet by an applied average filter. It is used for underwater image as prior processing [16]. Lastly, Stephane Bazeille et al., reduces confusion under the water, and increases in the quality of the image. This manner contains many of the successive phases of independent processing, nonstandard illumination valid to remove noise, enhance contract and settings to stretch the colors. The evaluating of the performance nomination will be used in the severity of the edge detection standard [5]. The subsequent paper proposes a great way to resolve the difficulties in underwater image. It uses the HSV color system instead of the RGB color system. The HSV is partitioned into three regions: hue, saturation, and value. Equally important, everyone of them, is in a separate stratum. This method allows you to process each of the strata alone by using the discreet wavelet transformation. We concentrate on the sensitive data of image during the use of the transformation of the wavelet and the stratum of hue, respectively. Consensually, the proposed manner offers processing with parts of the image, not as a whole. This is a very important factor in fast, performance and accuracy (See fig.9).

3. Primitives and Requirements

In general, the digital image is a two dimensional matrix. It contains a set of rows and columns. Each of the elements is called a pixel. There are three main types of images: color or true image, gray level image , and monochrome. Color image can be modeled into three main layers. It depends on the basic colors red, green and blue so each layer is different from the other in terms of contract and light intensity [17, 18]. The underwater image enhancing field has addressed enormous interesting within the last years, viewing significant achievements. Now, an increasing attend in naval research has instigated scholar or researchers from different specialties to ship in the deep

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

of underwater world. The underwater image has the same features of the normal images but the most different which is made the underwater image differs from the other images is its environment where, it depends on the stability or agitation of the sea or ocean as well as the reflecting of the sun light effective these issues cause the capturing of image is more difficult as well as the mentioned problems above.

2.1. Underwater image enhancement based on Color System

The underwater image enhancement can be assorted into two classes depending on the color system:

1) **The RGB Color System:** The RGB color system is the most popular systems. It is used to encode the colors. An effective feature is quantization of similar values per element in an image. The output underwater image enhancement in RGB color system not always has improved optimally; Moreover, it has impairment in assimilating rapid brightness impacting. For this reason, we consider transforming an underwater image from RGB color system to HSV color system. Fig.2 shows different types of images.

(a) Binary image (b) Gray-scale (c) Color image

Fig. 2 Views different types of images

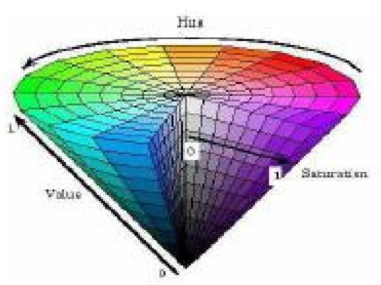


Fig. 3 HSV color system

2) **The HSV Color System:** HSV color system has the ability distinguished between color and severity and it can rebuild images better than is use with the RGB color system. We notice from Fig.3, that the system of HSV is a conical form. The hue stratum ranges from 0 to 1.0 so that it is dedicated to an angle for each color of the same band. The colors here are a closed loop. It starts with red and then tags the following chain of colors: yellow, green, cyan, blue, and magenta [18]. The saturation layer is expressed as the distance which starts from the center of the circle. We notice that the high percentage of saturated colors is on the borders of the cone, while small portions of gray tones are located in the center of the cone. Elements of this layer's are restricted in scope from 0 to 1. Moreover, all components are limited in the range [0,1] in the value layer, the layer that is specialized with brightness. Hence, it produces a more brilliant color. Fig.3 displays the HSV color system [19]. We illustrate this method by the following equations:-

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \quad (1)$$

Each normalized H,S and V elements are then obtained by

$$h = \cos^{-1}\left(\frac{05\ (r-g)+(r-b)}{((r-g)\ ^{2+(r-b)+(g-b))^{\frac{1}{2}}}}\right),\ h \in [0,\pi] \quad \text{for} \quad b \leq g \quad (2)$$

$$h = 2\pi - \cos^{-1}\left(\frac{05\ (r-g)+(r-b)}{((r-g)\ ^{2+(r-b)+(g-b))^{\frac{1}{2}}}}\right),\ \mathrm{h} \in [0,\pi] \text{ for } \mathrm{b} > g \quad (3)$$

$$s = 1 - 3\mathrm{Min}(\mathrm{r, g, b})\ ;\ s \in [0,1] \quad (4)$$

$$\mathrm{v} = \frac{(R+G+B)}{32\ 55};\quad \mathrm{v} \in [0,1] \quad (5)$$

From Fig.3 and Eq.(2) and (3), we note that the calculation of the values of the most important layer (H) , in this system by creating the linear transformation which starts from top to bottom. When you want to make a comparison between the original image and the image of H, the highest values are associated with a shadow of deep blue. The values in the opposite side are connected with a shadow of deep red [19, 20]. The degree of purity is related to a layer of color saturation. Color values in the saturation layer's are very high. Whereas a large degree of convergence is connected to the white shadow. We are noting blending colors like blue, red, and yellow, which resides in the kernel layer saturation. In addition to those intermingled colors, there is a shade of gray. The value stratum is a quasi synonym for brightness. The close link between the original image and the layer value is that the values match the original bright image with counter parts in the layer V [8, 19, 20].

2.2. Discrete Wavelet Transform

The discrete wavelet transformation is used to analyze the two dimensions image where it has the ability to divide the image into four main areas. The distributed data depends on the degree of the important data. The data area is called the low-low (LL) sub band. Occasionally, it is named as an approximation area.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

The rest of the regions are known as the detail. The present sub bands are as follow: high-low (HL), low-high (LH) and the high-high (HH). In brief, they are indicated as (DWT) [14, 16]. The wavelet analysis is permitted in using the long periods of time where the ambition is to get more accurate information from low frequency, and high frequency from the shorter zones [9, 21, 22]. The important data reside in the low frequency because it gives you the opportunity to get a signal of an identity. When you go to the high frequency content, it is characterized by high accuracy. It can be observed by the wavelet analysis for Lena's image (see Fig.5). The following figure, which is figure (4), displays the first level of the DWT [21, 22]. The tactical followed by the separation of the different characteristics of the signal relies on the method of gathering the energy of a few elements. This is called mechanism sub band coding. This class of DWT refers to the place of analysis via the approximation area at level j in four zones. Specifically, level j + 1 consists of approximation and the details. The details are distributed in three zones: horizontal, vertical, and diagonal [17, 19 , 23].

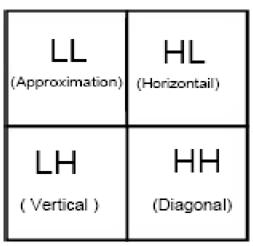


Fig. 4 The first level of the DWT


Fig. 5 The original image and 1-level wavelet decomposition of Lena's image

4. Our Proposed Scheme

Our proposed scheme focuses on providing an appropriate solution to the main problems of the underwater images. We had referred it in the previous parts of this paper. This technique deals with each layer of the HSV system separately. It chooses the most effective class (H), by Eq.(7) to be processed in the DWT. And then it determines the data area (LL) that emerged from the previous processing. The operation of compensation for the lost data are by Eq.(6). The remaining two layers are treated to restore balance in the intensity and illuminate the image contrast by Eq.(6). It will then be returned to enhance the stratum (LL) to the spatial transform. Then it will integrate again with its peers to step out into a new image in the area of HSV. Lastly, the outcome image is converted to space (RGB) which represents the image improved. The stretching function is used to scale the linear function that is applied to the values of pixels in the digital image. Each pixel is processed depending on the following equations:

$$P_i = \begin{cases} s, v + 01 & if \quad P_i \in [\, s, v \,] \\ DWT(h\,) & if \quad P_i \in H \end{cases} \qquad (6)$$

$$P_0 = \frac{(P_i - Low) \times (MAx - Low)}{(High - Low)} + \text{Min} \qquad (7)$$

Where:
Each h, s, and v rely on Eq. (3-7).
Po :is the enhanced pixel value[18].
Pi :is the pixel value which is applied to Eq.(6).
Min: is minimum value of the scope required.
Max: is the maximum value of scope required.
Low: is the low value of the pixels existing in the image.
High: is the highest value of the pixels which exist in the image.
DWT : discreet wavelet transformation.

We can divide the proposed work into four stages as follow:-The first stage is exerted to overhaul splits of the images. Individual part of the genuine data is enhanced for different color images, resulting in an eminent reduction in enhancement time. This approach implements the DWT. Primarily; it moves the true image from the pattern of the RGB color system into the HSV color system, because the HSV Color system also leads to offer best solutions to the absorption light by using brightness and intensity parameters. The second stage, it performs the DWT on the layer that is most influential which is termed hue in the HSV image. DWT fragments the image data into four data parts. One part of these data are a very heavyweight information. The other parts are for the detailed information. The part of important data will be functioned in the enhancement process. It is stretched to substitute the debiting of colors and rebalance of contrast. We can describe the above stages in more detailed manner. In this approach, the hue layer is represented as image to enter in wavelet domain. DWT are used to partially enhance underwater images. The fundamental idea behind DWT using is very convenient: when acquiring at underwater images and their wavelet domain will generates distributions of frequencies, it seems to contain the most energy (important information) and it locates in the lower frequency bands. One can profit this by splitting an underwater image into twofold: a low-frequency (L) and a high frequency (H) applying convenient filters which called low-pass and high-pass. We know an underwater image which is represented a two-dimensional array, this splitting has to be achieved in both trends leading in 4 subbands; they are named LL, HL, LH, and HH. The important information is distributed in the LL-subband; it fills a small scale edition of the underwater original image. The edge information is offered in other subbands. After the conversion of the image into wavelet domain the LL part will be enhanced. In current process, only important part of image is enhanced whereas the other parts are transmitted without enhanced. To prove the performance of this approach, Fig. (9) explains performance of this approach. The two layers, saturation and value, are also stretched to solve color problem and it contribute to increase intensity and brightness. The arithmetical model for saturation and value components enhancement is obtained by Eq. (6, 7). Where S and V are represent original saturation and value components respectively. As result, P_i is the enhanced saturation and value layers consecutively. The 0.1 is the addition stretch coefficient which detects the adaptive degree of saturation and value enhancement. The value of addition stretch coefficient was arrived at a value of 0.1 after conducting several experiments. This is happen in stage three. Lastly, the four parts are unified back again to gain the image in the wavelet phase. In the spatial pattern, the image will be generated by utilizing the inverse discrete wavelet transform. And then return from HSV color system to RGB. The proposed approach, It dealt with important data area by applying the wavelet transformation and the HSV color system. The stretching was a deep impact that we expressed in Eq.(6) and (7). The stretching operation was performed to rebalance the brightness and the contrast of the image. This is done by applying it on the layers S and V consecutively. Fig. 6 shows how the proposed scheme a worked.

5. Experimental Results and Analysis

In this part, we will address to view the experimental results that took place on a set of images. Our results show the effectiveness of the proposed method. Good results are evident through the seeing experiments on a set of images as can be seen, a histogram of the original image and the resulting image. So, it can reflect the compensation of the lost data during the acquired operation of the image. The evaluation process for the enhanced images will depend on two scales. The reason is due to the strength of the work that we have presented here. The first factor depends on the conviction of specialists in this field to retrieval of lost data through what we observe from the histogram. The second is the calculated amount of error and the amount of compensation to the input images during correlation scale [18,19].

5.1. Histogram

A histogram image is the mathematical standard of the digital image. Furthermore, it helps to understand the distribution of the graphic image. The histogram image means the process of distribution density of brightness and the contrast in the gray-level image [17, 22, 23]. This method is applied in our approach as measure, any person can note enhanced ratio during this scale. Fig.7 refers to the image of the histogram method.

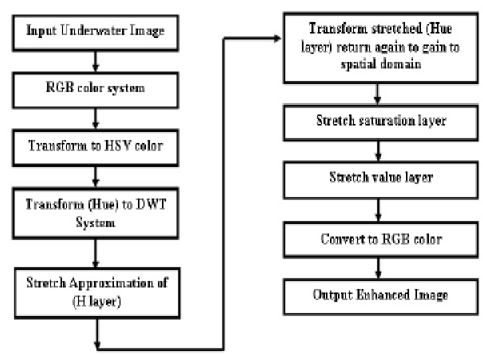
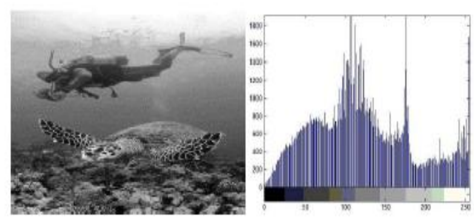


Fig. 6 Demonstration our proposed scheme



(a) Gray-scale image (b) The distributed levels of brightness and contrast

Fig. 7 Histogram of image

5.2. Correlation

Correlation (Crr) is a measure of sameness between the image before and after processing. The exemplary outcomes are near the one which can be identified the following equation [18, 19]:

$$C_{rr} = \frac{\sum_{r=1}^{N}\sum_{c=1}^{M} \left(I_n(r,c) - \overline{I_n}\right)\left(I_0(r,c) - \overline{I_{n0}}\right)}{\sqrt{\left[\sum_{r=1}^{N}\sum_{c=1}^{M} \left(I_n(r,c) - \overline{I_n}\right)^2\right]\left[\sum_{r=1}^{N}\sum_{c=1}^{M} \left(I_n(r,c) - \overline{I_n}\right)^2\right]}} \tag{8}$$

Where:

$I_n(r, c)$: The digital value of the pixel in the (r, c) of image before processing.

I_n : It is referring of the image before processing that

$$\overline{I_n} = \frac{1}{M \times N} \sum_{r=1}^{N}\sum_{c=1}^{M}\left(\overline{I_n}\,(r,c)\right) \tag{9}$$

$I_0(r,c)$: The digital value of the pixel in the (r, c) of image before processing.

$$\overline{I_0} = \frac{1}{M \times N} \sum_{r=1}^{N} \sum_{c=1}^{M} (\ \overline{I_0}\ (r,c)) \quad (10)$$

M: height of the image. N: width of the image. r and c: row and column numbers. The average of these three correlation's are used to generate the Crr of the recreate color image in RGB system. The color correlation equation is as follows:

$$Crr_{RGB} = \frac{Crr_{Red} + Crr_{Green} + Crr_{Blue}}{3} \quad (11)$$

are closeness for each color layer and Crr_{Red} and Crr_{Green}, Crr_{Blue} where Crr computed by Eq.(10). We can computing the substituting ratio by the following equation:

$$Sub_{RGB} = Crr_{RGB} - (1 - Crr_{RGB}) \quad (12)$$

The proposal scheme offers mechanism for achieved good balance for main components of the image such as color, brightness and contrast. we use R layer to substitute color loss, while we use both of s and V layers for control the diversity contrast ratio and balancing luminance in images, addition to that, the mechanism work of DWT function as an orthonormal transformation is to decrease the high-dimensional time series into a much more integrate data consideration, with full information saved within its operands. Also we observed that the proposed method has provided balanced solutions for all the contents of the image. The empirical findings view that color underwater images enhanced by the presented approach are pronounced, more visible and more brilliant than original image. The performance of the presented approach is obtained by running color enhancement performance, brightness enhancement performance, contrast brightness enhancement performance, histogram, correlation ratio, and substituting ratio. Fig.8 (a, e, i) show the source underwater images of size 512*512 pixels. Fig.8 (b, f, j) view the histogram of source images . Fig.8 (c, g, k) shows the underwater image enhanced based on our scheme. Fig.8 (b, f, j) views the histogram of target images, where, we can be distinguished into two categories according to proposed approach. Table(1) shows experimental results of underwater images in Fig.(8). In our scheme, it is also proposed to improve the level of performance by performing DWT in H layer which explained in Fig. (9) for images in Fig. (8). The Fig. (9) measures the time per amount of image data processed which explains robustness of the proposed technique which we obtained on a high level of performance. The proposed technique achieved a fantastic balanced mechanism between the image enhancement and processing time. This seems evident in Table (1) and Fig. (9) which is inferred from a set of images that we have chosen in this paper that in which we tested a wide variety of.

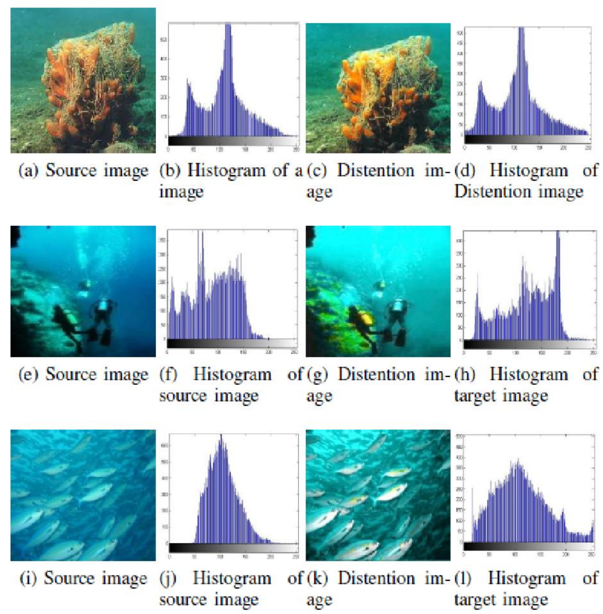


Fig. 8: Results of underwater image enhancement by using our proposed scheme. Source images (a, e and i). Images (c, g, and k) represented the resulting images.

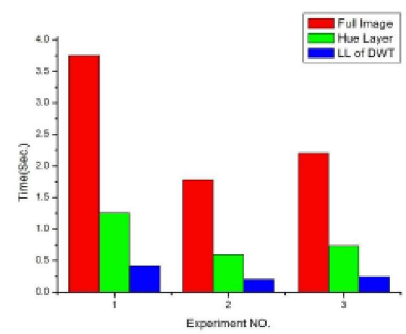


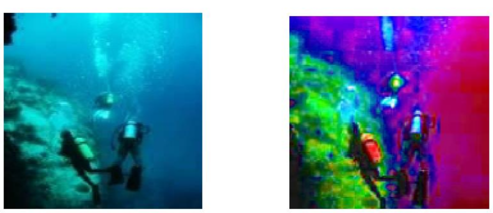
Fig. 9: Time versus enhancement method for source images (a, e, and i) in Fig.8

Table 1: shows the amount of error images and amount of compensation the after processing

source images	Error	Correlation	Min	Max	substituting
a	0.0260	0.9740	0	1	0.9480
e	0.0097	0.9903	0.3	1	0.9806
i	0.0252	0.9748	0.2	1	0.9496

5. Conclusions

There is no doubt that the image processing is a source of the most important sources of scientific impact in many sciences terms. Therefore we see it has been entered strongly in all areas. In this paper we address an important gap, all professionals in the field of marine science and image enhancement can benefit from this. Our proposed scheme aims to achieve a good balance among the missing elements of the underwater image. During that time, we focused on three important components in the images, as reflecting in important data, lighting intensity, and contrast. This is so the tasks can be distributed on this basis. The process of compensating the missing data is the task of the DWT. The balance of brightness and the contrast is due to the stretched process. We used the wavelet transform with the HSV color space to focus on the major areas in the image of precision and processing speed. On the other hand, if we raise the discrete wavelet transform of photos to be addressed, we realize its impact on improving the image as in Fig.10. Out of the results, any person can observes that: 1) the compensation between the original image and the resulting-image is very high; 2) the running time is decreased. The running time is cleared(in case LL) as shown in Fig.9 for source images in Fig.8. The proposed method can be developed to work on most accurate levels of the image. And using methods of classification on the image to be processed. The work here has a dual function; 3) we can develop approach to work in several fields such as enhancement of satellite image and medical images; 4) the primal occupation as classifier where the data classification based on the colors that affect the image; 5) the secondary function as an evaluator compares the levels of light intensity and the contrast between the image and colors before and after treatment.



(a) Source image (b) Target image

Fig. 10: Stretching colors without wavelet

References

[1] M.C. Motwani, M.C.Gadiya, and R.C.Motwani, "survey of mage denoising techniques", In Proc. of Global Signal Processing Expo (GSPx'04), Canda, 2004, pp. 27-30.

[2] K. Lebart, C. Smith, E. Trucco, and D. M. Lane, "Automatic indexing of underwater survey video: algorithm and benchmarking method," IEEE J. Ocean. Eng.", vol. 28, no. 4, 2003, pp. 673–686.

[3] I. Vasilescu, C. Detwiler, and D. Rus, "Color-accurate underwater imaging using perceptual adaptive illumination", Autonomous Robots , vol.31, no. 2-3, 2011, pp. 285-296.

[4] H. Raju, M.C. Ravishankar, M. Rameshbabu, and R. Chandran,"Color Image Enhancement Using Multiscale Retinex with Modified Color Restoration Technique", In Proc. of 2nd International Conference on Emerging Applications of Information Technology (EAIT'11), IEEE Computer Society Conference, 2011, pp.93 - 97.

[5] A. Yamashita, M. Fujii, and T. Kaneko, " Color registration of underwater images for underwater sensing with consideration of light attenuation", In Proc. Of IEEE International Conference on Robotics and Automation (RA'07), 2007, pp. 4570-4575.

[6] K. Iqbal, R. Abdul Salam, and A. Osman, "Underwater Image Enhancement Using an Integrated Colour Model", IAENG International Journal of Computer Science, vol. 32, no. 2, 2007, pp. 239-244.

[7] A. Arnold-Bos, J.P Malkasse, and G. Kervern, "A Preprocessing Framework for Automatic Underwater Images Denoising", In Proc. of European Conference on Propagation and Systems, 2005, pp. 1-8.

[8] C. Perer and H. Richard, "Improved Single Image Dehazing Using Geometry", In Proc. of Digital Image Computing: Techniques and Applications (DICTA'09), 2009, pp. 103-110.

[9] C.J. Prabhakar and P.U Praveen Kumar, "An Image Based Technique For Enhancement Of Underwater Images", International Journal of Machine Intelligence, vol. 3, no. 4, 2011, pp. 217-224.

[10] L. Velho, A. C.Frery, and J. Gomes, "Image Processing for Computer Graphics and Vision", London, England: Springer-Verlag London Ltd., 2009.

[11] F. Petit, A. S.C.Laize, and P.Carre, "Underwater Enhancement By Attenuation Inverstonwith Quanternions", In Proc. of IEEE International conference on Acoustics, Speach, and Signal Processing (ICASSP'09), 2009, pp. 1177 - 1180.

[12] J. Singhai and P. Rawat, "Image Enhancement Method for Underwater, Ground and Satellite Images Using Brightness Preserving Histogram Equalization with Maximum Entropy", In Proc. of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'07), 2007, pp. 507-512.

[13] T. J. Lambert, "Digital Enhancement Techniques for Underwater Video Image Sequences", M.Sc. Thesis, 2007, University of Tasmania, Australia, 2005.

[14] L. Chao and M. Wang, "Removal of Water Scattering", In Proc. of the 2nd International Conference on Computer Engineering and Technology (CET'07), 2010, pp. 734-1467.

[15] G. Song and X. Qiao, "Adaptive Color Image Enhancement based on Human Visual Properties", In Proc. of the3rd IEEE Conference on Industrial Electronics and Applications (ICIEA'08), 2008, pp. 1892-1895.

[16] G. Padmavathi, P. Subashini, M.M Kumar, and S.K. Thakur, "Comparison of Filters used for Underwater Image Pre-Processing", International Journal of Computer Science and Network Security, vol.10, no.1, 2010, pp. 58-65.

[17] R. Gonzalez, R.E. Woods, and S.L. Edidins, "Digital image processing and using matlab processing", Pearson Prentice-Hall, New Jersey,USA, 2004.

[18] Z. Weiqing, Z. Min, W. Junwei, H. Haiyun, Y. Bo, X. Lijun, and Z. Liang, "Signal Processing for High Speed Underwater Acoustic Transmission of Image", In Proc. of Oceans-Europe, 2007, pp. 1-6.

[19] G. Blanchet and M. Charbit, "Digital Signal and Image Processing using MATLAB",Wiley- ISTE Ltd., 2006.
[20] Y. Yoo, A. Ortega, and B. Yu, "Image subband coding using context based classification and adaptive quantization," IEEE Trans. Image Processing, vol. 8, 199, pp. 1702–1715.
[21] S. Grace Chang, B. Yu, and M. Vetterli, "Adaptive Wavelet Thresholding for Image Denoising and Compression", IEEE Trans. Image Processing, vol. 9, no. 9, 2000, pp. 1532-1547.
[22] R. Claypoole, G. Davis,W. Sweldens, and R. Baraniuk, "Nonlinear wavelet transform for image coding", In Proc. Of the 31st Asilomar Conference on Signals, Systems, and Computers, USA, 1998, pp.662-66.
[23] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation", IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp.888-905.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Enhanced Technique for Data Cleaning in Text File

Arup Kumar Bhattacharjee[1], Atanu Mallick[2], Arnab Dey[3], Sananda Bandyopadhyay[4]

Dept. of MCA, RCC Institute of Information Technology, West Bengal University of Technology
Kolkata, West Bengal, India
www.rcciit.in

Abstract

Data cleaning is a process of correcting or removing of erroneous data caused by contradictions, disparities, keying mistakes, missing bits, etc to create consistent and reliable information. Text files are used to store simple information and which can be also deceptive in terms of dirty data. In this paper we have provided a solution to cleanup simple text file using some data cleaning processes. Though we use text files so often but there is no such robust method exist to clean up text files. As data cleaning plays a crucial role for decision management which is depend on high quality data. So we have implemented a set of methods to clean text files. Here we use text files to store data in tabular format and our system checks whether there exist any error and finally try to correct or remove the errors according to different algorithms.

Keywords: *ETL, Data Dictionary, Metaphone, Date Validation Rules, Gender Validation Rules.*

1. Introduction

At the beginning of our cleaning process, we fetch data from text file and then we apply several algorithms to rectify the erroneous record and after modification we put back the corrected data to the same source text file and may be kept in newly created text files as per user request. For implementing the process we have used **ETL** model (Extract, transformation and load) [1], [6], [7].

Extract - The process of fetching data from external sources (Text files).

Transform – In this process, several rules are applied on the fetched data for validation.

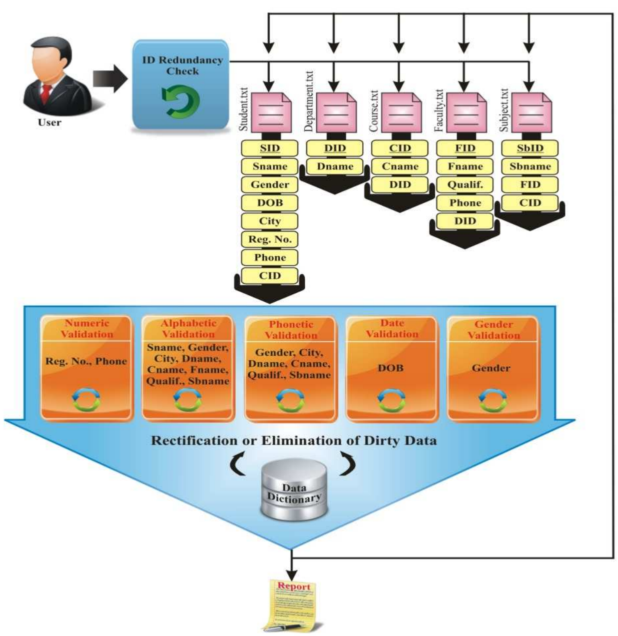
Load – The process of putting back the transformed data to a target location (May be source text file or other text file).

We categorize different type of error that can be occurred due to various reasons. And use respective rules to correct the data. Cleaning processes use various data dictionary (text file format) to match with the nearest correct data and to replace the erroneous data with the correct one. Finally, a report consist of detailed information about the rectified data along with the percentage of modification is being generated by the system.

As high quality data is essential for accurate data analysis and decision making, this data cleaning process ensures users to get correct and quality data. In our project we have used simple text file for keeping the information as it reduces the overhead of maintaining the storage and cost complexity of other database packages and making the system portable.

2. System Architecture

(Fig 1) describes the overall system working principal. Our system provides a user interface (UI) where user gives requested input. Here we have considered college information system as a demo process. First of all, system validates the ID field *(At entry level we are checking the redundancy and pre-defined format of Id fields. If the ID is redundant or empty, the system will request user for a unique ID. The entry will not be submitted to the input text file until user gives a valid ID, If user input an ID of improper format, system will try to rectify it into pre-defined pattern)* [1]. The data are extracted from the input text files *(student.txt, course.txt, department.txt, faculty.txt, subject.txt)* based on attributes and system categorizes them to process through some functions namely Numeric Validation, Alphabetic Validation, Metaphone Phonetic Validation, Date Validation and Gender Validation. These functions use different Data Dictionaries for valid data reference and the system finally generates a report containing all modifications to the original files.



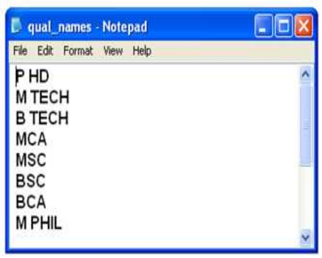
. **Fig 1: System Architecture [1]**

3. Data Dictionary

We have included some Data Dictionaries (DD) for some specific fields like city, qualification, course name, department name, subject name etc to rectify and replace with specified form or most nearest form for the erroneous data. To incorporate this transformation we first generate the phonetic code of the erroneous data and match with each phonetic code of each and every value in corresponding data dictionary. Two examples of such Data Dictionary are shown below.

DEPARTMENT DD

QUALIFICATION DD



4. Taxonomy of Errors

Here we have classified the types of error that can occur in the input text file, those are:-

(1) Numeric Value in alphabetic token [1] *(e.g. Name, Gender, City)*
(2) Alphabets in Numeric token [1] *(e.g. Phone no, registration no, Date)*
(3) Phonetic error *(e.g. City, Course Name)*
(4) Invalid or Redundant ID pattern [1] *(e.g. SID, CID, DID)*
(5) Misspelling *(e.g. City)*
(6) Invalid date *(e.g. DOB)*
(7) Invalid gender [1], [2]

5. Rules and Algorithms

(A) ID Correction Rules: In this ID correction rules we assume that the id will be always of 4 characters *(e.g. S007 for student ID, C276 for Course ID).*

Step 1: Eliminate all alphabets from the given string id.
Step 2: If length of id equals to 1 or 2 then we will add "00", "0" in front of the number respectively. In all other cases we will take only first three digits.
Step 3: After that, for the case of Student, Department, Course, Faculty id we will add S, D, C, F respectively in front of their ids. (e.g. for Student ID - S001, S002. For faculty -F001, F002. For Course - C001, C002 etc)
Step 4: After transformation if cause any redundancy, request user for unique one.

(B) Alphabetic Rules:

Step 1: Extract each character from the given string.
Step 2: Check whether the character is an alphabet or not. If it is digit, then check whether it matches with 0, 5, $, &, @ , i, I, l.
Step 3: If matches with the above digits, then we will transform those digits into the resembling character. If it is 0(Zero) then transform into O, if 5 or $ or & then transform into S, if @ then transforms into A, if ! or 1 or l or | then transform into I.
Step 4: In case of all other characters remove them all from the given string.

(C) Numeric Rules:

Step 1: Extract each character from the given string.
Step 2: Check whether the character is numeric or not. If it is alphabet, then check whether it matches with o, O, i, I, l, !, s, S..
Step 3: If matches with the above alphabets, we will transform those digits into the resembling digit. If it is 'o' or 'O' then transform into 0(Zero), if 's' or 'S' then transform into 5, if '!' or 'I' or 'i' or '|' then transform into 1(one).
Step 4: In case of all other character remove them all from the given string.
Step 5: In case of Phone number first we check the length of the number.
- **(1)** If Phone No length less than 9 then initialize Null to the string.
- **(2)** If Phone No length equals to 9 then append 0 in the end of the string.
- **(3)** If Phone No length greater than 10 then return the First 10 digit.
- **(4)** If Phone No length equals to 10 then take the whole string.

(D) Metaphone Phonetic Rules:

1. Drop duplicate adjacent letters, except for C.
2. If the word begins with 'KN', 'GN', 'PN', 'AE', 'WR', drop the first letter.
3. MB → B only if MB at the end of word.
4. CIA→ X; CH → X; SCH→K; C[IEY] → S; Otherwise C→K
5. DG[EIY] → J Otherwise D→T
6. Drop 'G' if followed by 'H' and 'H' is not at the end or before a vowel. GN -> N; GNED→NED and is at the end.

7. 'G' transforms to 'J' if before 'I', 'E', or 'Y', and it is not in 'GG'. Otherwise, 'G' transforms to 'K'.
8. Drop 'H' if after vowel and not before a vowel.
9. CK → K
10. PH → F
11. Q → K
12. SH → X; SIO → X; SIA → X
13. TI[AO] → XI[AO] ; TH → O ; TCH → CH
14. V → F
15. 'WH' transforms to 'W' if at the beginning. Drop 'W' if not followed by a vowel.
16. Drop 'Y' if not followed by a vowel.
17. Z → S
18. Drop all vowels unless it is in the beginning.
(Here → implies transformation from left side to the right hand side)

(E) Date Validation Rules:

We have specified our pre defined date format as dd/mm/yy. We assume year is in the last portion of our input always and the age of the student is greater than 12. Following steps summarize most of the rules from the original implementation in our project.

Step 1: Input Date.
Step 2: The system will take only the input string containing minimum 6 digits and it will check whether the use has given delimiters or not. Otherwise system will return null.
Step 3: We parsed the input string into three parts- day, month, year respectively; which is separated by delimiter.
Step 4: If length of the year string is equals to 4 then we extract last 2 characters and if the length of the year string is equals to 2 then we extract the year string.
Step 5: We convert the year string into integer, and check whether the age is greater than 12 or not. If the year is in between 1900 and 1999 then it will execute following steps otherwise it will return nothing.
Step 6: If user gives the month input using name of the month instead of numbers, the system will convert the month in system defined format even if the month string starts with or ends with few characters. Those rules are given below

Months	Starts With	Ends With	Replace With
January	Ja	nuary	01
February	F	ruary	02
March	Mar	ch	03
April	Ap	il	04
May	M[aeiou]y	ay	05
June	Jun	ne	06
July	Jul	uly	07
August	Au	ust	08
September	S	tember	09
October	O	ober	10
November	N	vember	11
December	D	cember	12

Step 7: Then we will check for different condition for the day and month field. It can be in dd/mm/yy format or mm/dd/yy format.
Step 8: Here if the day is greater than 31 then first we are reversing the number and then check whether the number is greater than 31 or not. If not, then we will consider the value. (*Our system considers that, it may be possible that the user can give wrong input for certain reasons*)
Step 9: Check for the April, June, September, November month. Those months cannot exceed 30 days if user gives wrong input as 31. Then our system will consider as typing mistake and will give value of 13.
Step 10: Check for February month. And if the year is leap year then only the month can have 29 days. For normal year the system won't take day as 29.
Step 11: After all checking, the system will give input in dd/mm/yy format. If none of the rule satisfies the user input then system will return null.
Step 12: End.

(F) Gender Validation Rules:

1. First go through Alphabetic validation.
2. If the return string starts with 'm' or 'M' then convert the string to MALE.
3. If the string starts with 'f' or 'F' then convert the string to FEMALE.

.

6. Class Models

We have modularized our project through some classes for ease of understanding. Our system has two set of classes.
1. Classes that represents the input text files *(Student, Department, course, Faculty, Subject)*.
2. Validation Classes *(Date, Gender, Numeric, Alphabetic, Phonetic)*.

Whenever system retrieves a tuple from a text file, a new object of corresponding class gets created, populating each property with the respective field values of that tuple. Now related validation objects *(instantiated validation classes)* takes this object as input and validate properties and finally return back the object with rectified field values which can be inserted into output text files.

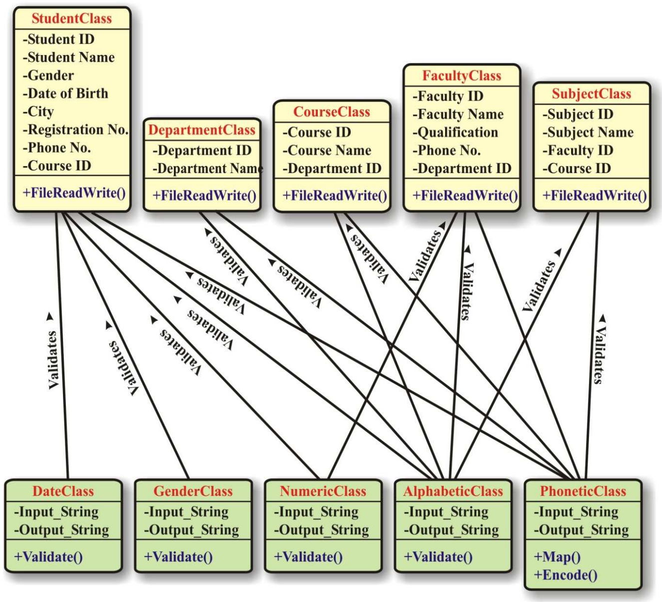
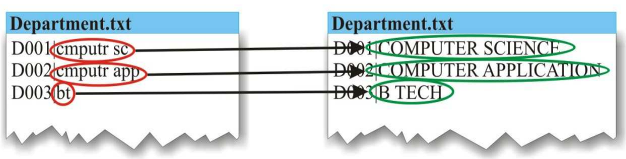


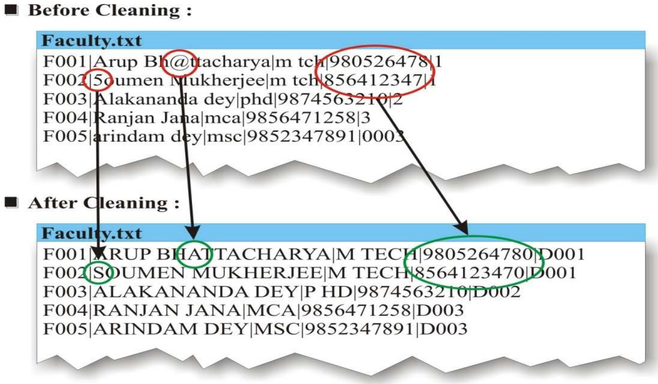
Fig 2: Class Models

7. Sample Outputs

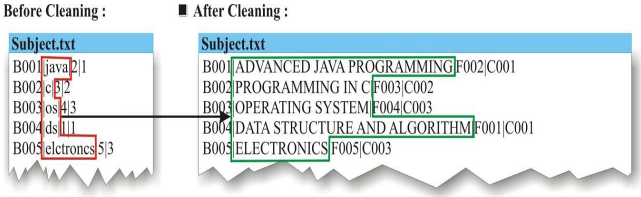
■ **Before Cleaning :** ■ **After Cleaning :**



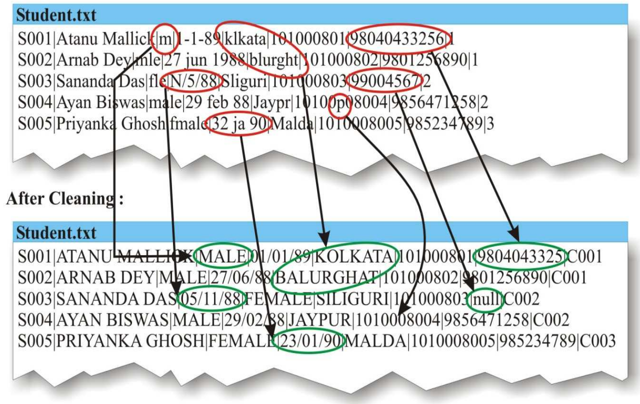
■ **Before Cleaning :**



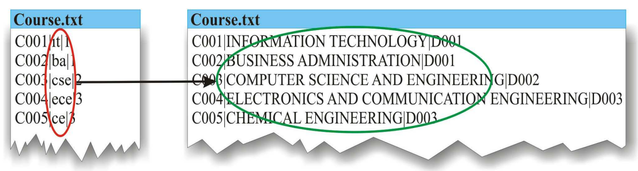
■ **Before Cleaning :** ■ **After Cleaning :**



■ **Before Cleaning :**



■ **Before Cleaning :** ■ **After Cleaning :**



8. Future Scope

Here we have used text file in tabular format, it is possible to correct errors without the tabular structure. In this paper we have used Metaphone (phonetic algorithm) for pronunciation of word; it can be implemented by double Metaphone and Metaphone 3 algorithm which are latest algorithms. Here we have improved the knowledge base according to our requirements, anyone can change that for their own suitable condition. Our algorithms for ID, Alphabetic, Numeric, Date, Gender and Phonetic validation can be improved or replaced depending upon situation arises.

9. Conclusion

Our data cleaning framework preserves the quality error free data in text file. Data accuracy is very hard to achieve through data-cleansing in the general case, because it requires accessing an external source (Data Dictionary) of data that contains the true value. So our approach is quietly based on this predefined knowledge base which can be improved further with a best possible outcome of true value for the erroneous data. As text files are used massively it can flaw a decision making process. So by enhancing the process of data cleaning in text file we can resolve the situation and can be later used as analytics software for decision making or removing dirty data in text file.

References

[1] Arup Kumar Bhattacharjee, Atanu Mallick, Arnab Dey and Sananda Bandyopadhyay, “Data Cleaning in Text File”, IOSR Journal of Computer Engineering (IOSR-JCE), ISSN: 2278-0661, Volume 9, Issue 2 (Jan. - Feb. 2013).

[2] R. Cody, “Data cleaning 101,” Proceedings for the Twenty-Seventh SAS User Group International Conference. Cary, NC: SAS Institute Inc,2000.

[3] Dr. Mortadha M. Hamad and Alaa Abdulkhar Jihad, “An Enhanced Technique to Clean Data in the Data Warehouse”. Computer Science Department. University of Anbar, Ramadi, Iraq.

[4] Hasimah Hj Mohamed, Tee Leong Kheng, Chee Collin and Ong Siong Lee, “E-Clean: A Data Cleaning Framework for Patient Data”. School of Computer Sciences. University Sains Malaysia Penang, Malaysia.

[5] Arindam Paul, Varuni Ganesan, Jagat Sesh Challa and Yashvardhan Sharma, “HADCLEAN: A Hybrid Approach to Data Cleaning in Data Warehouses”. Department of Computer Science & Information Systems . Birla Institute of Technology & Science, Pilani, Rajasthan, India – 333031.

[6] Erhard, Rahm and Hong Hai Do. “Data Cleaning: Problems and Current Approaches”. University of Leipzig, Germany.

[7] Srivatsa Maddodi, Girija V. Attigeri and Dr. Karunakar A. K, “Data Deduplication Techniques and Analysis”. Manipal Institute of Technology, Manipal, India.

[8] R. Kimball and J. Caserta, “The Data Warehouse ETL Toolkit”. Wiley,2004.

[9] V. Raman and J. M. Hellerstein, "Potter‟s Wheel: An Interactive Framework for Data Transformation and Cleaning.," in Proceedings of the 27th VLDB Conference, Roma, Italy, 2001.

[10] K. Kukich, “Techniques for Automatically Correcting Words in Text”, ACM Computing Surveys, vol. 24, no. 4, pp.377-439, 1992.

[11] R. Bheemavaram, J. Zhang and W. N. Li, “Efficient Algorithms for Grouping Data to Improve Data Quality”, roceedings of the 2006 International Conference on Information & Knowledge Engineering (IKE 2006), CSREA Press, Las Vegas, Nevada, USA, pp. 149-154, 2006.

Author Biography

Arup Kumar Bhattacharjee received his MCA Degree from University of Kalyani and his M.Tech from West Bengal University of Technology. He has contributed to 15 books and coauthored 2 publications. He is an Assistant Professor of Computer Application at RCC Institute of Information Technology which is affiliated to West Bengal University of Technology in Kolkata, West Bengal. His research interests include Software Engineering, Object Technology and Parallel Computing.

Atanu Mallick has completed his Bachelors in Computer Science (*Hons.*) from Surendranath College under Calcutta University in Kolkata, West Bengal. Currently he is pursuing his Masters in Computer Application from RCC Institute of Information Technology which is affiliated to West Bengal University of Technology in Kolkata, West Bengal, India.

Arnab Dey received his Bachelors degree in Computer Application from Pailan College of Management & Technology, Kolkata under West Bengal University of Technology. Currently he is pursuing his Masters in Computer Application from RCC Institute of Information Technology which is affiliated to West Bengal University of Technology in Kolkata, West Bengal.

Sananda Bandyopadhyay has completed her Bachelors in Computer Application from Techno India (salt lake) under West Bengal University Of Technology. Now she is pursuing her Masters in Computer Application from RCC Institute of Information Technology which is affiliated to West Bengal University of Technology in Kolkata, West Bengal.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

A Multimodal Approach for Face and Ear Biometric System

Gandhimathi Amirthalingam[1], Radhamani. G[2]

**[1] Research Scholar, Department of Computer Science, Bharathiar University
Coimbatore, India**

**[2] Department of Computer Science, Dr. G. R. Damodaran College of Science
Coimbatore, India**

Abstract

Multi modal biometric system is one of the major areas of study identified with large applications in recognition system. Single modal biometric systems have to challenge with a variety of problems such as noisy data, Intra-class variations, non-universality, spoof attacks, and unacceptable error rates. Some of these limitations can be solved with multi modal biometric systems. The major purpose of the study is to review and analyze the prime works in multimodal biometric system and its efficiency in recognition rate. The proposed framework of the multimodal biometric system using face and ear is given. This paper also discusses the levels of fusion that are possible and understand the types of challenges focused by prior research work in this area.

Keywords: *Face recognition, Ear recognition, biometric recognition, multi-modal recognition.*

1. Introduction

Biometric techniques are being used increasingly as a hedge against identity theft. The premise is that a biometric is a measurable physical features or behavioral trait and is a more trustworthy indicator of identity than bequest systems such as passwords and PINs [3]. Biometrics first came to renown in 1879 when Alphonse Bertillon (1853–1914), a French Criminologist, introduced his anthropometrical signalment or Bertillonage system for identifying criminals [22]. A method of identification based on anthropometry of different parts of the human body had developed including head, ear, fingers etc., the size of which remain constant throughout life after attaining full growth [3]. However, greater accuracy and robustness is desired in biometric identification.

A method of identifying or verifying the identity of an individual person or subject based on the physiological and behavioral characteristics is biometric recognition. Physiological biometrics is based on data derived from direct dimension of a part of the human body [6]. It involves fingerprint, iris-scan, DNA, retina scan, hand geometry, and facial recognition. Behavioral biometrics is based on data derived from an action taken by a person or individuals behavioral characteristic. Behavioral biometrics characteristics involve voice recognition, keystroke-scan, and signature-scan. Any physiological or behavioral characteristic of human can be used as a biometric characteristic as long as it is Universal, Unique, Collectable and Permanent [22].

Biometrics recognition features can be either passive or active. The recognition of Face and ear feature are Passive biometrics. Users participation is not require. It can be analyzed and successful even without any explicit action of the part of the user. But Active biometrics like fingerprint, retina scanning, signature recognition, DNA etc. however, do require some voluntary action by the user and will not work if one reject participating in the process. Biometric-based personal recognition systems can be classified into two main categories: Verification and Identification. Biometric verification ("one-to-one matching") compares the registered template of identity against an input image of an unknown person, whether the person is claims to be. Biometric Identification ("one-to-many matching") compares the input image of an unknown person against all records in a registered template [6]. The system identifies the individual from the database gallery. This category is usually associated with law enforcement applications.

Biometrics is a fast growing technology which can be useful in criminal justice system like mug-shot, post-event analysis, forensics. It provides security to prevent unauthorized access to ATMs, computer networks, cellular phones, email authentication on multimedia workstations, PDA, medical records management and distance learning. The voice biometric can be used during transactions conducted via telephone and internet commerce and banking. Retinal patterns of an individual provide medical information about diabetes or high blood pressure. In automobiles, keys can be replaced with key-less entry devices by the fingerprint biometric system. Face biometric is used in smart card applications [10]. The face-print can be stored in a smart card, bar code or magnetic stripe. Active biometrics like iris, fingerprint, and retina are most widely used and well-known biometrics. The passive biometric, face recognition is used in forensic applications such as terrorist identification, corpse identification etc. The other biometric applications such as social security, national ID card, border control and passport control.

2. Issues in Unimodal Biometric

The successful installation of biometric systems in different civilian applications does not entail that biometrics is a completely solved problem. Single modal biometric traits have plenty of error rates and they may not achieve the desired performance requirements. There are many factors which degrades the recognition performance. Researchers are addressing to enhance the usability of biometric system. Some of the issues imposed by single modal biometric are given below:

2.1 Noise in sensed data

Biometric system has different sensed data. The sensed data might be noisy or deformed. A voice altered by cold, iris recognition with wearing glasses. Finger print with a scar, might be too oily, dry, wet or damaged temporally or permanently. Face sensed weaknesses due to variations in light, pose or illumination. Gait sensed with fluctuation in body weights. Noisy sensed biometric data may be false matched with templates in database resulting in a false rejection.

2.2 Distinctiveness

Biometric trait is expected to vary significantly across two individuals. The characteristics of the individuals are represents with the large inter-class similarity in the feature sets. The information content (number of distinguishable patterns) in two of the most commonly used representations of hand geometry and face are only the order of and, respectively [13].

2.3 Nonuniversality

Problems regarding the quality or consistency of the capture of biometric data may not necessarily due to a fault or error in the sensor [20]. About 4% of the population may have poor quality fingerprints, due to scars or cuts and it shows erroneous result. Intra-class variations, the biometric data acquired from an person during testing may be different from template data during enrollment. The users are incorrectly interacting with the sensor. This may affect the matching result.

2.4 Spoof attacks

Biometric traits of the legitimate user are enrolled in the template database; an imposter may attempt to spoof the sensed data of the biometric system when the traits such as signature [25] and voice [2] are used. The fingerprint traits can also be spoof with the artificial fingers/finger print to thwart a fingerprint verification system [24].

3. Multimodal Biometric System

Multi-modal biometrics increase accuracy by considering other highly specific biological traits to limit the number of applicant for an identity. This system is expected to be more reliable due to the presence of multiple, independent trait and not easy to forge multiple biometric trait. Variety of biometric scenarios is depending on the traits, feature sets and sensors applied. Some of the scenarios are multiple sensors, multiple algorithms, repeated instances, multiple modalities. Multimodal system functions in three different modes. In Serial mode, the output of the one modality is used to reduce the number of possible identities before the next modality is used. In parallel mode, sensed data from multiple modalities are used concurrently. In hierarchical mode, individual modality is combined in a hierarchy structure.

The performance of the multimodal system is expressed in terms of matching errors and image acquisition errors. Matching errors consist of false match rate (FMR), in which an impostor’s sample matches a legitimate user’s template, and False Non Match Rate (FNMR), in which a legitimate user’s sample does not match his/her own template. Image acquisition errors consists of Failure to Enroll (FTE) which is defined as a user that is unable to successfully enroll in a biometric system, and Failure to Acquire (FTA) is a user that is unable to provide a good quality biometric trait at verification.

4. Level of Fusions

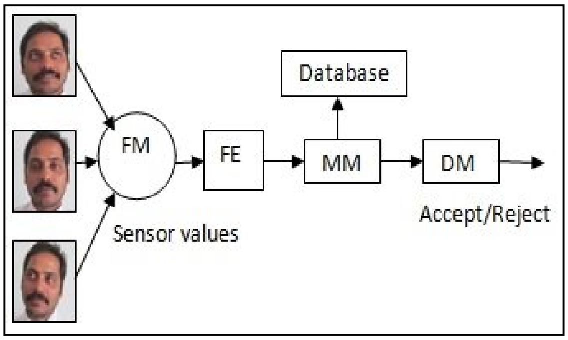
The Biometric fusion is the technique to integrate the classification results from each biometric channel. Multimodal biometric fusion combines the aspect from various biometric features to improve the strengths and reduce the limitations of the individual aspects. The efficiency of the fusion scheme greatly influences the accuracy of a multimodal biometric system. The various levels of fusion are:

4.1 Sensor level fusion

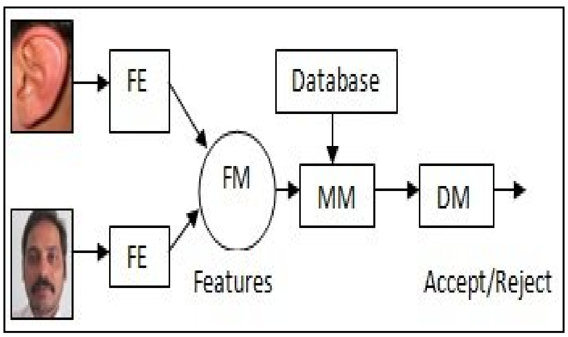
The raw data obtained from multiple sensors can be practiced and merged to generate new biometric data from which trait can be extracted. Biometric traits from different sensors like fingerprint, video camera, iris scanner, digital signature etc, are fused to form biometric trait to process. Sensing a speech signal concurrently with two various microphones may be fused and then be subjected to feature

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

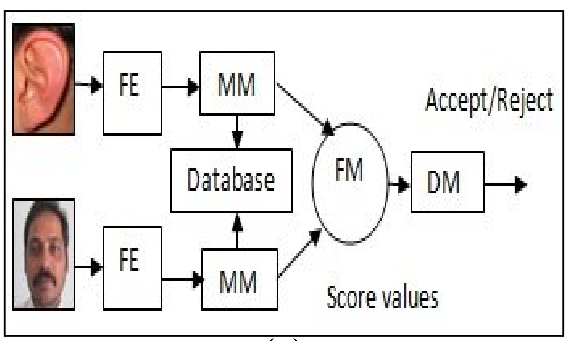
extraction and matching. Sensor level is projected to improve the recognition accuracy; it remains possible problems related with unimodal biometric system because of incompatibility of data from various modalities [12].



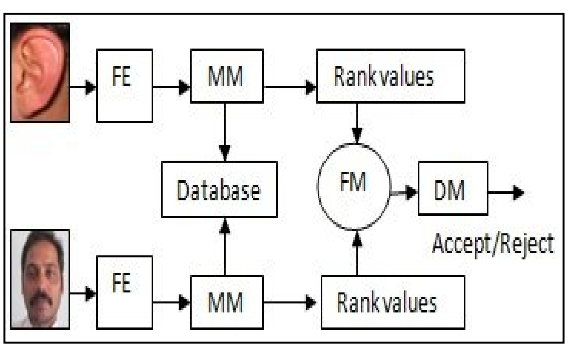
(a)



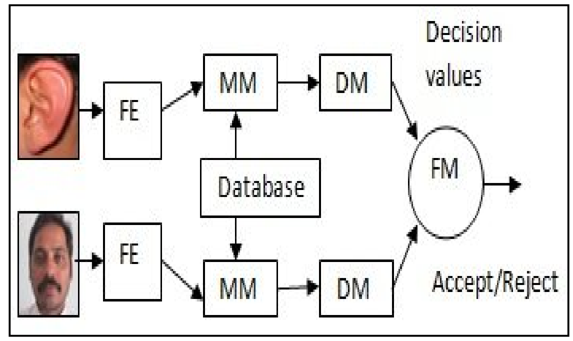
(b)



(c)



(d)



(e)

Fig 1. Level of Fusions: (a) Sensor Level fusion, (b) Feature Level fusion, (c) Match Score Level fusion, (d) Rank Level fusion, and (e) Decision Level fusion. FE: Feature Extraction Module; MM: Matching Module; DM: Decision-Making Module; FM: Fusion Module.

4.2 Feature level fusion

The feature sets are extracted from different biometric channels can be fused using specific fusion algorithm to form a composite feature set. The feature collections of different modalities agree to extract a minimal feature set from the high-dimensional feature vector. The feature vectors extracted from the face and ear modalities can be fused is an example of multimodal system. The feature level fusion is the extraction of correlated feature from the different modalities and in course identifies a prominent set of features that can improve recognition accuracy [7]. The feature level fusion is likely to achieve superior result in comparison with score level and decision level fusion.

4.3 Match score level fusion

Feature vectors are generated separately for each modality. Extracted feature vectors compared with the templates residing in the database individually for each biometric trait to generate match scores. Depending on the accuracy of each biometric channel, output set of match scores which are fused to create composite matching score. As an example, face and hand modalities match score may be combined by the use of simple sum rule in order to obtain a new match score which is then sent to the decision module [4].

4.4 Rank level fusion

Rank level fusion is a new fusion approach where each classifier associates a rank with every enrolled identity. Fusion involves consolidating the rank output by individual biometric subsystems and determining a new rank that would support in establishing the final decision. However, these fusions have one weakness. In multimodal biometric, more different identities output from two or three matching modules which are designed to appear some identities of only one matcher. In this case, the rank level fusion shows the risk of wrong results [11].

4.5 Decision level fusion

In multimodal biometric system, the final decision is based on the separate decision of different modalities using techniques such as majority voting, behavior knowledge space, weighted voting, AND rule, OR rule. Decision level fusion is least powerful due to availability of inadequate information and limits the basis for enhancing the system accuracy.

5. Prior Research Work

The prior works of the researches in multimodal biometric system are reviewed. Features of the face or other parts of the human have dissimilar properties for different sensors. Each parameter of the biometric can be characterized as better or worse depending on the data of the individual is acquired for identification purposes. The important features of multi-modal biometric studies are summarized in Table 1.

Muhammad Imran Razzak et al. [18] combined the face and finger veins, in which multilevel score level fusion is performed to increase the robustness of the authentication system. The score level fusion of client specific linear discriminant analysis (CSLDA) for fusion of face and finger veins result is performed. CLSDA uses the PCA and LDA to generate a client specific template. The score of face and finger veins are combined using weighted Fuzzy fusion. This system is efficient in reducing the FAR 0.05 and increasing GAR 91.4.

A human recognition method combined face and speech information in order to improve the problem of single biometric authentication are proposed by Mohamed Soltane et al. [16]. Gaussian mixture modal (GMM) is the main tool used in text-independent speaker recognition, in which it can be trained using the Expectation Maximization (EM) and Figueiredo-Jain (FJ) algorithms for score level data fusion. The use of finite GMM based Expectation Maximization (EM) estimated algorithm for score level data fusion is proposed. Extracted face and extracted audio is fused to achieve recognition rate. Face speech biometric EER is reduced to 0.087.

A multi-biometric system using lip movement and gestures is proposed by Piotr Dalka et al. [19]. . Lip gesture recognition is performed by an artificial neural network (ANN) approach. ANN contains parameters like no gesture, mouth opening, forming puckered lips, sticking out the tongue and all gestures. The experiment used 6120 image frames. The entire feature vector for ANN contains lip region only. ANN is trained with a resilient back propagation algorithm (RPROP). The result shows that the recognition rate is 93.7%.

A multi-biometric system using face and ear is presented by A.A. Darwish et al. [1]. PCA deccorrelate data by finding the eigen vectors of the covariance matrix. MIT, ORL and Yale databases are used for implementation. The individual face and ear images are normalized and preprocessed and then transformed to the PCA space. The system performance is 92.24% with FAR of 10% , FRR of 6.1%, Because of high accuracy and security, it concluded that the fusion of face and ear is a good technique.

The schedule extraction of local 3D features (L3DF) from ear and face biometrics and their arrangement at the feature and score levels for identification has been presented by S.M.S. Islam *et al.* [23]. 3D features removed from ear and frontal face information are fused at feature level. Scores from L3DF and iterative closest point algorithm were fused at matching level by means of a weighted sum rule. This system achieved recognition and verification (at 0.001 FAR) rates of 99.0% and 99.4%, respectively, with neutral and 96.8% and 97.1% with non-neutral facial expressions.

The multimodal system of face and ear at feature level fusion by Sparse Representation (SR) are proposed by Zengxi Huang *et al.* [27]. SR-based classification methods used in classification phase were Sparse Representation based Classification (SRC) and Robust Sparse Coding (RSC). Finally, they have obtained a group of SR-based multimodal recognition techniques, together with Multimodal SRC with feature Weighting (MSRCW) and Multimodal RSC with feature Weighting (MRSCW).

A novel kernel-based feature fusion algorithm method in combination of face and ear is proposed by Xu Xiaona et al. [26]. Combining with KPCA or KFDA algorithm, the feature fusion method were presented and applied to multimodal biometrics based on fusion of ear and profile face biometrics. This system defines the Average rule, Product rule, Weighted-sum rule in kernel-based fusion feature method and USTB database is analyzed. The experimental shows that the recognition rate of KPCA is 94.52% and KFDA is 96.84%, and this method is efficient for feature fusion level.

H. Mahoor et al. [15] proposed with a 2D face and 3D ear fusion at the match scores level using weighted sum technique. Active Shape Model is used to extract a set of facial landmarks from frontal facial images. For the ear recognition, a set of frames is extracted from a video clip and ear region in each frame is restructured in 3D using Shape from Shading (SFS) algorithm. The resulting 3D ear

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

models are aligned using the iterative closest point (ICP) algorithm. The experiment performed on a database of 402 subjects. The performance of the system is increased to 100%; FAR 0.01%, EER of the multi-modal system is .01%.

M. Kawulok et al. [14] presents a face and eyes using multi-level ellipse detector combined with a SVM verifier. The main contribution is in increasing the accuracy of eye detection in high-quality images. The authors show that the detection error propagation considerably influences the face recognition performance. With the proposed improvements, face recognition increase the rate by 0.5 % for FERET and 7.7% for AR database compared with the publicly available implementation of the well established Viola-Jones face and eye detector.

Linlin shen et al. [23] proposed improve the accuracy by integrating multiple modal biometrics i.e face and palmprint. The both face and palmprint feature are represent by feature code, namely FPCode. FPCode uses fixed length 1/10 bits coding scheme that is very efficient in matching, and at the same time achieves higher accuracy than other fusion methods available. This approach compares with the Gabor+PCA and Gabor+KDRC. Experimental results show that both feature level and decision level strategies achieve much performance with the accuracy of 91.52% and 91.63%.

There are different ways of integrating different modality and they are depends on the number of samples, multiple matches, multiple snapshots, multiple sensors and the number of biometric features in the context of multi-biometric studies. Ear feature provide better biometric performance. Ear undergoes very slight changes from the childhood to adulthood. Due to ears semi-rigid shape and robustness against change over time, the ear has become an increasingly popular biometric feature. It has been shown that combining individual biometric methods face and ear into multi-biometric systems improves recognition.

6. Multiple Biometrics using Face and Ear

Multimodal biometrics based on the fusion of two different biometric modalities face and ear; provide a new approach of non- invasive biometric authentication. There are several inspirations to choose face and ear for a multi-modal biometric recognition. During image acquisition, ear and face data can be captured using conventional cameras. The data collection for face and ear does not require participation or cooperation from the user. The traits face and ear are in close physical proximity to each other. Both biometric features are jointly present in an image or video captured of a user's head and are both available to a biometric system. In prior literature work, the fusion of face and ear biometric shows the good performance in accuracy and recognition.

7. Proposed Framework

Multimodal Face and ear are combined to increase the robustness of the recognition system. The proposed model includes a training phase and recognition phase. In training phase, the samples of data on which the system needs to recognize are trained. Figure 2 present the proposed framework of multimodal recognition system. Figure 3 shows the sample face and ear image tested in the proposed system. Two different modalities face and ear are applied here. Recognition phase comprise with preprocessing, feature extraction and authorization. The input face and ear images are preprocessed, to reduce or eliminates some of the variations in input image. It enhances the image to improve the recognition performance of the system. Shape and texture features are extracted from the input images. A shape feature is to extract the shape of face and ear by using modified region growing algorithm. Texture feature are extracted by LGXP Technique. The authorization is done with the fuzzy vault. For decoding, the constructed face and ear image is combined with the stored fuzzy vault to generate the final key. The performance will be evaluated with the False Matching Rate (FMR), False Non Matching Rate (FNMR) and Genuine Acceptance Rate (GAR).

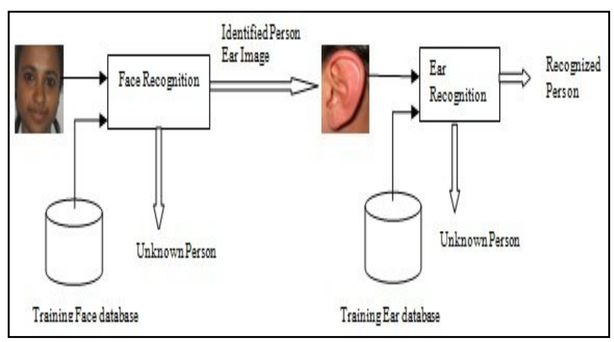


Fig. 2 Proposed framework of the system

Fig. 3 Sample Face and Ear image

TABLE 1: MULTI - MODAL BIOMETRIC STUDIES

Source (year)	***Databases***	***Biometric sources***	***Technique Adopted***	***Performance of Classification in percentage***	***No. of subjects***
Muhammad Imran Razzak(10') [18]	CAIRO	Face + Finger Veins	client specific linear discriminant analysis (CSLDA)	FAR 0.05% and GAR 91.4%	35 subjects,
Mohamed Soltane(10') [19]	UYVY. AVI 640 x 480, 15.00 fps	Face + Speech	Gaussian mixture modal (GMM)	EER: Face 0.44935, Speech 0.00269, Face + Speech (fusion) 0.08728	30 subjects
Piotr Dalka(10') [16]	Faces are recorded using web camera	Lip movement + Gestures	Artificial Neural Network (ANN)	Recognition Rate 93.7%	176 subjects
Darwish (09') [1]	MIT, Yale	Face + Ear	Principal Component Analysis (PCA)	Accuracy of 92.24% with FAR of 10% and FRR of 6.1%	MIT – 40 subjects ORL – 15 YALE-10.
S.M.S. Islam (13') [23]	UWA, UND-FRGC, UND-F and FRGC V2	Face + Ear	L3DF, Iterative closet point	FAR 0.001 Recognition: 96.8% Verification: 97.1%	UWA – 56 subjects UND-FRGC : 326 UND-F and FRGC V2: 100
Zengxi Huang (13') [27]	MD I: Yale B and USTB. MD II : AR and USTB	Face + Ear	Sparse Representation based Classification (SRC), Robust Sparse Coding (RSC)	MDI MSRCW : 95.732% MRSCW:97.86% MD II MSRCW: 98.39% MRSCW: 99.0%	MD I : 38 MD II: 79
Xu Xiaona(09') [26]	USTB database	Face + Ear	KPCA, Kernel Fisher Discriminant Analysis (KFDA)	Recognion Rate fusion KPCA 94.52%, KFDA 96.84%	79 subjects
M.H. Mahoor(09') [15]	West Virginia University database	2D Face + 3D Ear	Weighted sum technique	EER .01%, FAR .01%, Rank one identification 100%,	402 subjects
M. Kawulok (12')[14]	FERET, AR database	Face + Eye	multi-level ellipse detector combined with a SVM verifier	Increase the recognition rate by 0.5% for FERET and 7.7% for AR.	FERET: 3657 images AR: 3313 images
Linlin shen(11') [23]	AR, PolyU database	Face + Palmprint	FPCODE	Feature level fusion : 91.52% Decision level fusion : 91.63%	AR : 119 subjects PolyU : 386 palms

8. Comparison with Other Biometrics

Gait: Gait is a behavioral biometric. Gait is not supposed to be very distinctive, but is sufficiently discriminatory to allow verification in some low-security applications. It may not remain invariant, especially over a long period of time, due to fluctuations in body weight, major injuries involving joints or brain.

Iris: Iris is much smaller than the ear, a high resolution camera device is required in order to acquire image of acceptable quality. In general, the capturing sensor device is usually placed far from the subject. Iris recognition also can fail when the subject wear glasses.

Fingerprint: Fingerprint biometric system entails the use of specially designed sensors and computational resources which maybe too expensive for large scale deployment, especially when operating in the identification mode. Fingerprints of a small fraction of the population may be unsuitable for automatic identification because of genetic factors, aging, environmental, occupational reasons. Manual workers may have a large number of cuts and marks on their fingerprints that keep changing.

Voice: The voice of a person changes over time due to age, health conditions and emotional state, etc. Voice is also not very unique and may not be appropriate for large-scale identification. A disadvantage of voice-based recognition is that speech features are sensitive to a number of factors such as background noise.

Keystroke: This behavioral biometric is not expected to be unique to each individual. The keystroke dynamics may vary depends on the health condition. It is expect to observe large variations in typical typing patterns. The keystrokes of a person using a system could be monitored quietly as that person is keying in information.

Palmprint: The palmprints scanners need to capture a large area; they are more expensive than the fingerprint sensors. The physical size of a palmprint based system is large, and it cannot be embedded in certain devices.

Signature: The signature of a person is to be a characteristic of that individual. Signatures require contact with the writing instrument and an effort on the part of the user, which have been accepted in government, legal, and commercial transactions as a method of verification. It changes over a period of time and is influenced by physical and emotional conditions of the signatories. Signatures of some people vary significantly. Professional forgers may be able to reproduce signatures that fool the system.

8.1 Face biometric

Face recognition has potential applications in security control, surveillance, office automation, prevention of fraud, video indexing, automatic personalization of environments, etc. [21]. Face recognition is passive and non-intrusive unlike other active biometric techniques such as those using fingerprints, speech and signature [17]. There are two main categories of face recognition systems: First, Face detection and normalization, the face image database contains one image per person. System identifies a person and returns a list of names that most likely matches the query face image. Secondly, Face identification, System identifies a person from a smaller face databases so that they can gain entry to a particular resource. The face recognition techniques can be modified and used for gender classification. The high accuracy of the biometric system is to identify faces in real time under different facial expressions, hairstyle, and image background.

8.2 Ear Biometric

Ear is a new class of human biometrics for physiological identification with uniqueness and permanence. Ear has information rich anatomical feature and unaffected by ageing. Its location on the side of the head makes extraction easier. Ear biometric is convenient in collecting data comparison to other technologies like retina, iris, fingerprint [5]. The combination of ear and face show high recognition results. The ear features and ear identification were using in forensic for more than 10 years [1, 8]. In the absence of fingerprint, due to lack of expression and less effect of aging, the ear biometric is suggested for the identification. The recognition is similar to face recognition and it consists of image acquisition, preprocessing, feature extraction, model training and template matching.

7. Conclusion

Multimodal biometric systems address numerous problems observed in single modal biometric systems. The complex methods employed to find a good combination of multiple biometric modality and various level of fusion applied to get the best possible recognition result are discussed in this paper. The prior work has shown the performance evaluation of the multimodal system under the different trait combination scheme, identification rate and databases. The combination of face and ear modality are suggested and the proposed framework of the biometric system is given. In this paper, table 1 claims that multi-biometrics improve over a single biometric system and uncorrelated modalities are used to achieve performance in multimodal system.

References

[1] A..A. Darwish, R. Abd Elghafar and A. Fawzi Ali, “Multimodal Face and Ear Images”, Journal of Computer Science, 2009, Vol. 5 (5), pp. 374-379.

[2] A. Eriksson and P. Wretling, “How flexible is the human voice? A case study of mimicry,” in Proc. Eur. Conf. Speech Technology, Rhodes, 1997, pp. 1043–1046.

[3] A. K. Jain, A. Ross and S. Prabhakar, “An introduction to biometric recognition”. IEEE Transactions on Circuits and Systems for Video Technology, 2004, Vol. 14, pp. 4–20.

[4] A. Ross and A. K. Jain, “Information fusion in biometrics,” Pattern Recogn. Lett., 2003,Vol. 24, No. 13, pp. 2115–2125.

[5] Charles Schmitt, Allan Porterfield, Sean Maher, David Knowles, “Human Identification from Video: A Summary of Multimodal Approaches”, Institute for Homeland security solutions, Jun 2010.

[6] Gandimathi. A, Radhamani. G, “Comparison on Multi-modal Biometric Recognition Method”, CIIT, Springer Publications, 2011, pp. 355-357.

[7] J. Thiran, F. Marques, and H. Bourlard, Multimodal Signal Processing, Elsevier Ltd, 2010.

[8] Kyong I. Chang, “New multi-biometric approaches for improved person identification”, Ph.D. thesis, December 2004.

[9] Linlin Shen, Li Bai, and Zzhen Ji,” FPCODE: An efficient approach for multi-modal biometrics”, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 25, No. 2 (2011), pp. 273-286.

[10] L. Puente Rodríguez, A. García Crespo, M. J. Poza Lara, B. Ruiz Mezcua, “Study of Different Fusion Techniques for Multimodal Biometric Authentication”, Proc. IEEE CS International Conference on Wireless & Mobile Computing, Networking & Communication, 2008, pp. 666-671.

[11] Md. Maruf Monwar, Marina L. Gavrilova, “Multimodal Biometric System Using Rank-level Fusion Approach”, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 2009, Vol.30, No.4, pp. 867-878.

[12] M. Faundez-Zanuy, "Data fusion in biometrics," IEEE Aerospace and Electronic Systems Magazine, 2005, Vol. 20, pp.34-38.

[13] M. Golfarelli, D. Maio, and D. Maltoni, “On the error-reject tradeoff in biometric verification systems,” IEEE Trans. Pattern Anal. Machine Intell., 1997, Vol. 19, pp. 786–796.

[14] M. Kawulok, J. Szymanek, “Precise multi-level face detector for advanced analysis of facial images”, IET Image Process., 2012, Vol. 6, Iss. 2, pp. 95–103l.

[15] Mohammad H. Mahoor , Steven Cadavid, and Mohamed Abdel-Mottaleb, “Multi-modal Ear and Face Modeling and Recognition”, Proc. IEEE 16th International Conference on Image Processing, 2009, pp. 4137-4140.

[16] Mohamed Soltane, Noureddine Doghmane, Noureddine Guersi, “Face and Speech Based Multi-Modal Biometric Authentication”, International Journal of Advanced Science and Technology, 2010, Vol. 21(8), pp. 41-46.

[17] M. Turk and A. Pentland, “Eigenfaces for recognition”, Journal of Cognitive Neuroscience, 1991,Vol.3(1), pp.71-86.

[18] Muhammad Imran Razzak1, Rubiyah Yusof and Marzuki Khalid,”Multimodal face and finger veins biometric authentication”, Scientific Research and Essays, 2010, Vol. 5(17), pp. 2529-2534.

[19] Piotr Dalka, Andrzej Czyzewski, ”Human-Computer Interface Based on Visual Lip Movement and Gesture Recognition”, International Journal of Computer Science and Applications, 2010, Vol. 7(3), pp. 124 - 139, 2010.

[20] Prof. V. M. Mane and Prof. (Dr.) D. V. Jadhav ,” Review of Multimodal Biometrics: Applications, challenges and Research Areas “,International Journal of Biometrics and Bioinformatics (IJBB), Volume 3, Issue 5, pp. 90-96.

[21] Rabia Jafri and Hamid R. Arabnia, “A Survey of Face Recognition Techniques”, Journal of Information Processing Systems, 2009, Vol.5, No.2, pp. 41-68.

[22] Ruma Purkait, “Ear Biometric: An Aid to Personal Identification”, Anthropologist Special, 2007, Vol. 3, pp. 215-218.

[23] S.M.S. Islam, R. Davies, M. Bennamoun, R.A. Owens and A.S. Mian, “Multibiometric human recognition using 3D ear and face features”, Pattern Recognition, 2013, Vol. 46, No. 3, pp. 613-627.

[24] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, “Impact of artificial gummy fingers on fingerprint systems,” Proc. SPIE, 2002, Vol. 4677, pp. 275–289.

[25] W. R. Harrison, Suspect Documents, Their Scientific Examination.Chicago, IL: Nelson-Hall, 1981.

[26] Xu Xiaona, Pan Xiuqin, Zhao Yue, Pu Qiumei, “Research on Kernel-Based Feature Fusion Algorithm in Multimodal Recognition”, IEEE CS International Conference on Information Technology and Computer Science, pp.3-6, 2009.

[27] Zengxi Huang, Yiguang Liu, Chunguang Li, Menglong Yang and Liping Chen, “A robust face and ear based multimodal biometric system using sparse representation”, Pattern Recognition, 2013, Vol. 46, No. 8, pp.2156–2168.

Gandhimathi Amirthalingam received her Bachelor’s degree in Electronics from Bharathiar University. Then she obtained her Master’s degree from University of Madras in the year 2000 and M.Phil from Madurai Kamaraj University in the year 2007. She is pursuing Ph.D in Computer Science from the Bharathiar University. Her specializations include Image Processing, Computer vision, Multimedia Computing. Her current research interests are Biometric Security, Pattern recognition. She is a student member of the IEEE and IEEE Computer society.

Radhamani. G is Professor and Director in Department of Computer Science, Dr. G.R.Damodaran College of Science, India. She received her PhD (Computer Engineering) from Multimedia University, Malaysia and M.Sc., M.Phil (Computer Science) degree from PSG College of Technology, India. She served in MMU, Malaysia from 2001 to 2006 and Muscat from 2009 to 2012. She has published papers in International Journals and Conferences. She is a Senior Member of IEEE and CSI. Her research interests are Databases, Computer Security and Mobile Computing, Image Processing, Pattern Recognition.

A Vision Based Approach for Web Data Extraction Using Enhanced Cocitation Algorithm

R.Vijay[1], Dr. K. Prasadh [2]

**[1]Research Scholar, M.S University
Tamil Nadu, India**

**[2]VK college of Engineering and Technology
Parippally, Kerala.**

Abstract

Normally, the World Wide Web maintains a set of databases which can store several data records retrieved by web query interface. The information maintained in web is hidden in the database that can be retrieved through dynamic script pages are termed as deep web content. These forms of deep web contents are normally accessed by the web queries, but, extracting the structured data from web database involves complexity. To address the issue, Wei Liu et. al., presented programming language independent vision based approach that use the visual features of deep web pages for web data extraction. The vision based approach also includes the process of extraction of data record and data item. But the unsolved issues in Liu's vision based approach is that it not only process the deep web pages in one data region of the web page but also consumes additional time to extract the visual information of web pages. To address the demerit present in ViDE, a novel technique called vision based approach for deep web data extraction is presented. In this work, we describe a framework that processes the deep web pages present in multi data regions. The framework uses enhanced co-citation algorithm that, instead of developing a new set of APIs for the extraction of visual information, the algorithm retrieve the visual information of the deep web pages directly from the web database. Empirical studies with large set of database for web data extraction demonstrate that the performance of the proposed vision based approach [VBEC] are capable of offering high precision while enabling efficient and accurate recall value of similar queries with better time consumption compared to other extraction processes.

Keywords: ***Deep web data, vision based approach, multi data regions, co-citation algorithm, visual features, and web data extraction.***

1. Introduction

Nowadays, there is a tremendous increase of usage in World Wide Web and has secured to one million searchable information sources. These searchable information sources includes both Web databases and search engine. By relocating queries to search interfaces of these information sources, practical information from them is readily accesses. For instance, the processed information is returned to the requested pages using several techniques consisting of data records, each of which are again related with the entity for example a document or a book. Data records are significantly revealed carefully on Web browsers to alleviate the utilization of human users. On the other hand, to construct the data records which are taken out from the machine process able, which is required in different applications for instance deep web crawling and meta-searching, it is highly required to be taken out from the response set of web pages.

The existing deep web page problems have come up with several solutions which are highly supported by examining the HTML source files of the response set of web pages. Even though they can realize sensibly high accuracies in the provided experimental outcomes, the present studies of this crisis have numerous limitations; HTML-based techniques suffer from the subsequent problems:

HTML itself is still growing and when novel versions or marks emerge, the earlier solutions have to be processed frequently to settle in to novel provision and original tags.

- Earlier versions on HTML-based techniques, only measured the HTML file that do not include JavaScript and CSS. As more and more web pages use more complex scripts, the applicability of the presented resolutions become poorer.
- If HTML is restored by a novel language in the prospect, then earlier resolutions have to be adjusted significantly or even discarded, and other techniques must be presented to put up the novel language.
- Finally, conventional performance procedures, precision and recall, do not completely discard the excellence of the extraction.

There are previously some parts that examine design construction of web pages. They attempt to electively symbolize and recognize the construction of web pages, which are substantial formation independent. It is recognized that web pages are employed to distribute information for humans to peruse, and not considered for computers to haul out information repeatedly. Based on such deliberation, in this paper we propose a technique to haul out data records mechanically supported with visual illustration of web pages. Our VBEC approach follows three-steps to realize this objective.

2. Literature Review

Numerous techniques have been discussed in the literature for extracting out the information from Web pages. The most primitive techniques are the instruction manual approaches in which languages were considered to help out programmer in building wrappers to recognize and take out all the preferred data items/fields. In recent times [1], presented a new vision based technique for the extraction of web page by developing a set of API.

The crisis of extracting out the data records on the response web pages revisited from web databases or search engines remained a critical task. World Wide Web has created a demanding crisis in taking out the relevant data [2]. The huge number of techniques has been presented to deal with this crisis, but all of them contain intrinsic limitations since they are Web-page-programming-language reliant or sovereign [3]. The paper [4] studied the crisis of taking out the data records on the response web pages revisited from web databases or search engines. A new and language self-sufficient system is presented to resolve the data extraction crisis.

With the intention of develop the efficiency and decrease physical efforts, most current researches center on regular approaches in place of corporeal or semiautomatic ones. The automatic data alignment technique in [5] proposes a clustering technique to achieve alignment supported with five features of data items, as well as font of text. On the other hand, this technique is chiefly text-based and tag-structure-based, while the resented technique is mainly visual-information-based. So the efficiency of this approach is very less.

A search engine revisited consequence page might hold search results that are prearranged into numerous dynamically measured sections in reply to a user query. Moreover, such a outcome page frequently contains information unrelated to the query, such as information connected to the hosting location of the search engine. In [6], the author presented a technique to robotically produce wrappers for taking out search result records from all active sections on outcome pages revisited by search engines.

Dynamic web page has a huge quantity of pages, high-value data and elevated modularity construction. Along with these feature, the paper [7] developed routine web information taking out scheme based on page clustering. Web pages are modeled by closing data values to reconstructed templates. Physical data extraction from semi supervised web pages is a complex job. The paper [8] focused on revise of different automatic web data extraction methods. OLERA [9] is semi supervised information taking out tool [10] where user can produce extraction rules consistent with training pages. But the information charge does not provide appropriate applicable web pages based on users' query. The above discussed issues are resolved in this work by implementing the vision based approach for web page extraction using enhanced co-citation algorithm under multi-data region.

3. Proposed Methodology

In this paper, the process of deep web data extraction from multi data regions is done effectively by adapting the enhanced co-citation algorithm. The proposed vision based approach for web data extraction is processed under three different phases. The first phase describes the process of obtaining visual representation of the deep web pages from the web database using enhanced co-citation algorithm. The second phase elaborates the process of retrieving data records from the web pages from multi data regions. Finally, the third phase describes the process of extracting data items from the data records and forms an efficient web data page. The architecture diagram of the vision based approach is shown in Fig. 1.

From the figure (Fig. 1), it is being noted that the process of VBEC is clearly explained with the step by step procedure. At first, the web pages are extracted from the web database using enhanced co-citation algorithm. Second, once the process of extraction of web pages is accomplished, next the data records are extracted from the multi- data regions and are further processed. Based on users' queries, the third and final step involved in VBEC is that the exact information is retrieved from the web database for an ease of use.

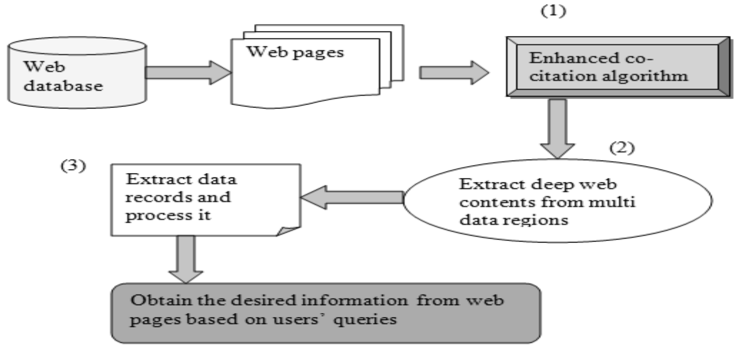


Fig. 1 Architecture diagram of VBEC

3.1 Extraction of web pages using enhanced co-citation algorithm

Web page consists of several set of information. The information present in the web page involves a text, image, video etc. Visual information of the web pages is normally related with the location, size and typeset of the web pages. The extraction of visual information of the web page is retrieved using enhanced co-citation algorithm.

The enhanced co-citation algorithm is processed based on defining some notations. Consider a set of two web pages as A and B. Let us further assume that page A is a parent of page B as page A might refer the web page B in its page. But page B has already derived from another web page which could be referred as B'. Now, these two different set of pages B and B' are said to be co-cited only if both share the same parent web page A. With this process, the degree of co-citation is measured. But some of the drawbacks related to co-citation algorithm were, the main problems like navigation links and retrieving duplicate set of pages. To address these kinds of issues related to co-citation algorithm, an enhanced co-citation algorithm is presented to retrieve the visual information of the web page directly from the web database.

The enhanced co-citation algorithm follows two strategies to extract the visual information of web pages from web database with respect to a users' query:

- Content-based, and
- Link-based

The content-based technique extracts the textual content of the users' query links and its siblings. The link-based technique utilizes only the link construction among the web pages collected for the enhanced Co-citation algorithm.

As stated above, the collection of web pages from the web database should detain the degree of union related with a topic pointed to by the links inside that group of web pages. If the topic of the link that is being focused on is restricted, then the parent webpage are expressed by the average similarity of the topics of the web page siblings based on users' query. The retrieved two topics are similar only if the sim (t1, t2) =1.

For the specified topic and sim functions, the notion of obtaining the content based Web Page extraction WP with URL u is obtained by identifying the similarities among the query URL's topic and the focus of the other siblings. It is specified as,

$$\text{Content based (WP)} = \frac{\sum \text{sim(topic(u), topic(wp))}}{|WP| - 1} \quad \ldots\ldots\ldots\ldots (1)$$

The content based enhanced co-cited algorithm retrieves the set of web pages to symbolize the average similarity of its similar web pages based on users' query with the obtained URL. Higher values point out stronger union with the siblings of WP concerning the topic of the query URL u. If the web page does not have includes any title tag, then the content based web page retrieval might be a complex one. In that case, a link based co-citation is utilized for extraction of web pages.

Let WD be a web database with a set of web pages wp1,w p2, . . . , wpM (as well as the query URL). The occurrence of overlap of the parent web pages are identified as P and Q, which is a sign of the union among the parent web pages on their children. We suggest that the parent web page P is measured even if more parents concur on more siblings of P. P consists of the utmost focal point in the case when all the parents related to the web pages are similar. This overlapping of web pages is normally accessed by the link based web data extraction which is expressed as,

$$\text{Link based (WP)} = \frac{\sum \frac{wp1 \cap wp2}{wp1 \cap wp2}}{|WP|} \quad \ldots\ldots\ldots\ldots (2)$$

This link-based view of web page extraction is computationally low-cost because the mandatory data is previously evaluated by the enhanced Co-citation algorithm. The pseudo code below describes the process of enhanced co-citation algorithm

Input: Web database WD, web pages WP
Identify the queries sent by the user
Identify the topic (t) of the query
    Perform content based web page extraction
If (WP has no title tag) **do**
    Perform link based web page extraction
End if
For (each parent web page p) **do**
    Customize the collected topics of the user query
End For
Rank sibling web pages based on its degree
Return (web pages based on highest degree)
End

With the queries, the web pages are efficiently extracted directly from the database based on following the subsequent steps describes in the above pseudo code.

3.2 Data record extraction

The main objective of Data record extraction is to determine the border line of data records and remove them from deep Web pages. Let us assume that the perfect record extractor should realize the subsequent properties and sees to that the following assumption are satisfied:

1) All data records present in multi data region are extracted and

2) For every extracted data record, no data item should be neglected and no erroneous data item be incorporated.

The figure (Fig. 2) below describes the general case of multi-data region of web database.

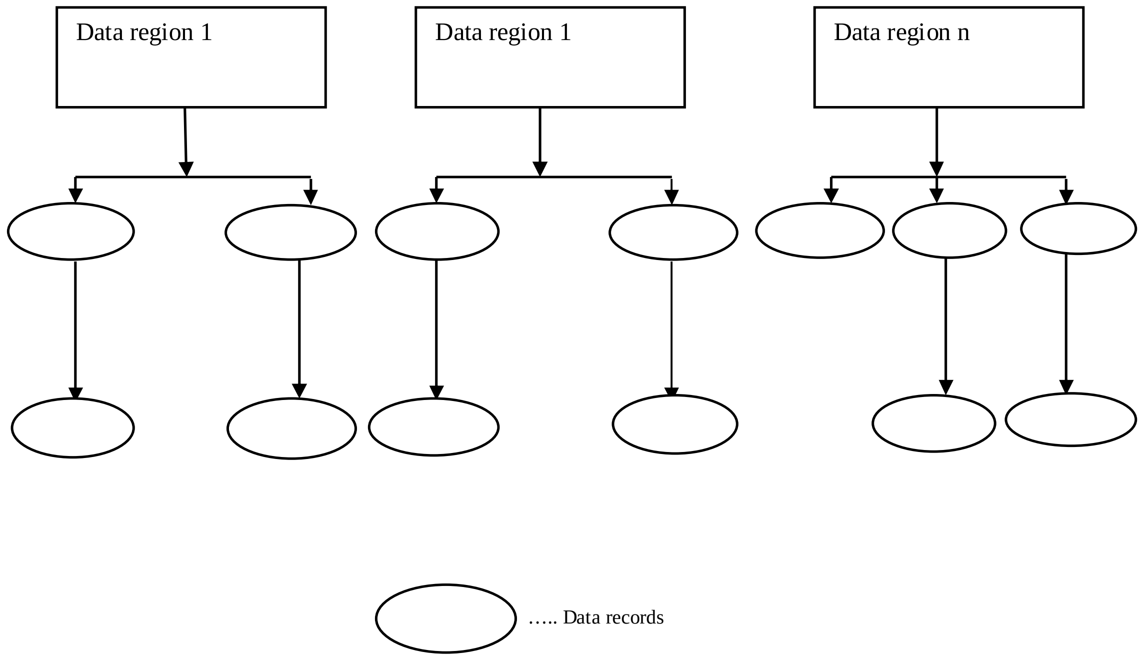


Fig. 2 Multi data region

Instead of retrieving the data records from the deep Web page straightforwardly, the data regions in the database are first established and then pull out data records from the data region. The data records are normally the prime content on the deep Web pages and the data region is centrally positioned on these web pages. The data region corresponds to a page P in the enhanced co-citation algorithm. We situate the multiple set of data regions by deciding the position of the web page that satisfies the two enhanced co-citation techniques.

In order to pull out data records from the multi-data region precisely, two facts are measured. First, there might be web pages Pi that do not fit in to any data record, for instance the geometric information. Then such web page(s) that do not fit into any data record are removed. Second, one data record might communicate to one or more web pages in the co-citation method, and the total number of web pages in which one data record hold is not permanent. In this case, the data record used more than once is referred for communication but not during the processing stage, that avoid redundancy.

3.3 Data item extraction

A data record is observed as the depiction of its equivalent data objects, which has a collection of data items and some fixed template texts. In genuine applications, these prearranged data records are gathered (often in relational tables) at data item rank and the data items of the similar semantic must be positioned beneath the similar column. When commencing the enhanced co-citation algorithm, every data record has been twisted into a series of data items during data record segmentation. Data item arrangement explain the way to support the data items of the similar by keeping the arrangement in semantic manner for the data items in every data record.

After the data regions are identified on a deep Web page WP, the data record is extracted from the respective web page and the space connecting two data records are identified. For the location data region in a new page, each data record is discovered by the visual comparison with the collected visual information using enhanced co-citation algorithm. The pseudo code below describes the process of enhanced co-citation algorithm for web data extraction in multi data region.

Input: Web database WD, Data records DR, Data items DI, Users U
Identify the queries Q sent by the user
For each query **do**
    Perform enhance co-citation algorithm described in section 3.1
End
For each extracted webpage P **do**
    Identify data region DR
    For each DR **do**
        Identify set of records DR
        Select DR which is more relevant

Identify DI present in the appropriate DR
End For
For each users' query **do**
Sort the Web pages P relevant to the user query related information
End for
End

The above algorithm describes the entire process of enhanced co-citation algorithm with vision based approach for web page extraction. The next section describes the experimental evaluation to estimate the performance of the proposed technique.

4. Experimental Evaluation

An experimental evaluation is done to estimate the performance of the vision based approach [VBEC] using enhanced co-citation algorithm. The VBEC experiments are done on a Pentium 4 1.9 GH, 512 MB PC. A large set of web database is utilized which consists of more than 10,000 entries of web database. These web databases are categorized into several domains. Several set of users submit their queries to extract the relevant information from the web database.

For each Web database, three set of user queries are presented and collect five deep Web pages holding three data records at any rate. With the set of web database, the enhanced co-citation algorithm is implemented to the database to extract the web pages directly from the web database. With the set of web pages, the data records and its corresponding data items are retrieved based on user query related information. The performance of the vision based approach [VBEC] using enhanced co-citation algorithm is measured in terms of

i) Precision
ii) Recall
iii) Time consumption

Precision measures the ratio of web data records extracted using enhanced co-citation algorithm based on the relevancy factor that is relevant to the search based on user query related information.

Recall measures the ratio of data records extracted relevant to the query that are successfully retrieved using enhanced co-citation algorithm.

Time Consumption refers to the time consumed to extract the web pages using enhanced co-citation algorithm.

Time Consumption (TC) = (Query Count)i * (Cycle Time)i

Where time consumption for a specific user is evaluated by the products of queries to be executed for n users (i=1,2,…,N) and cycle time for the specific ith user (i = 1,2,….,N).

5. Results and Discussion

In this work, we have seen that the proposed VBEC approach efficiently extracts the web pages directly from the web database presents in multi-data region. For diverse set of web pages, the users' required data records are extracted and processed with the data items. After extracting the data records, the users' query related information is processed. The below table and graph describes the performance of the proposed vision based approach [VBEC] using enhanced co-citation algorithm.

Table 1: No. of entries in database vs. precision

No. of entries in database	Precision (%)	
	Proposed VBEC	Existing ViDE
100	85	75
200	97	77
300	89	80
400	90	85
500	91	90.5
600	92	91

The precision is made out in Table 1 based on the number of entries in the web database. The precision of the VBEC approach is compared with the existing ViDE approach.

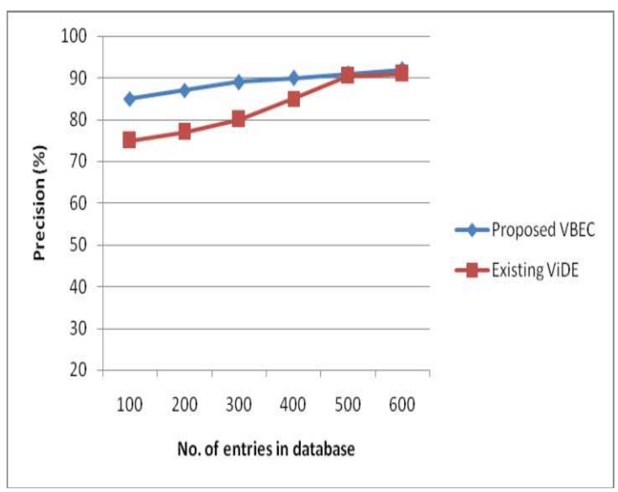


Fig. 3 No. of entries in database vs. precision

Fig. 3 describes the precision value of the extracted web pages based on the number of entries in the web database. The precision value is measured based on the rate at which the retrieved web pages are relevant to the users' queries. Compared to the existing ViDE approach, the VBEC approach provides better precision outcome since VBEC extracts the web pages directly from the web database using enhance co-citation algorithm. The enhanced co-citation algorithm efficiently filters the web pages and provides the desired outcome to the user. But in the existing ViDE approach, precision rate become less because of the presence of noise in the extracted web pages. So, the variance in the precision rate is 5-10% high in the proposed VBEC approach.

Table 2: No. of entries in database vs. recall

No. of entries in database	Recall (%)	
	Proposed VBEC	Existing ViDE
100	87	77
200	89	79.5
300	90	80
400	92	85
500	93	88.5
600	94	89

The recall is made out in Table 2 based on the number of entries in the web database. The recall of the VBEC approach is compared with the existing ViDE approach.

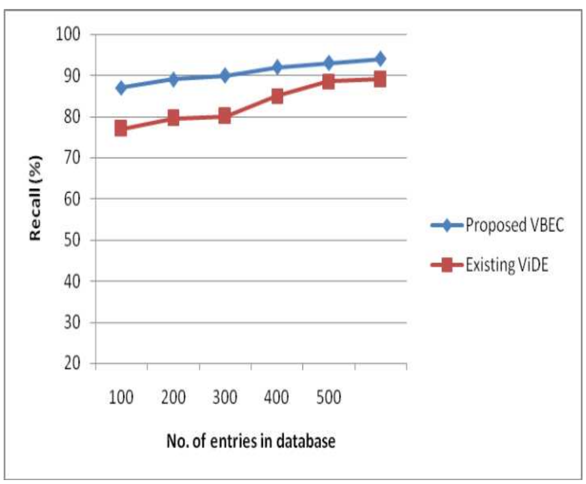


Fig. 4 No. of entries in database vs. recall

Fig. 4 describes the recall value of the extracted web pages based on the number of entries in the web database. The recall value is measured based on the rate at which the number of related web pages retrieved to the users' queries. Compared to the existing ViDE approach, the VBEC approach provides a better recall rate outcome with a variance 5-10% since VBEC uses multi-data region from the web database using enhance co-citation algorithm. But in the existing ViDE approach, recall rate become less because of the presence of noise in the extracted web pages.

Table 3: No. of entries in database vs. time consumption

No. of entries in database	Time consumption (sec)	
	Proposed VBEC	Existing ViDE
100	10	20
200	16	26
300	20	35
400	23	41
500	27	48
600	30	53

The time consumption required to extract the relevant deep web pages based on the entries of the web database is illustrated in Table 3. The time consumed for VBEC approach is compared with the existing ViDE approach.

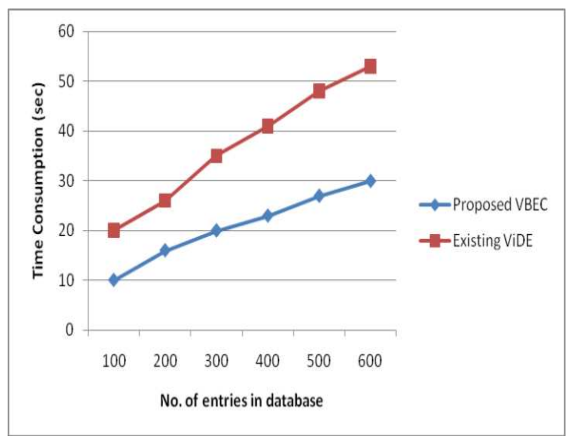


Fig. 5 No. of entries in database vs. time consumption

Fig. 5 describes the consumption of time required to extract the relevant deep web pages based on the entries of the web database. Time consumption is measured in terms of seconds. Compared to the existing ViDE approach, the proposed VBEC technique consumes less time to extract the web pages. Because the existing ViDE approach used IE as APIs for extracting the web pages from the web database. This consumes more time since for developing the IE tools for web page extraction. But in the proposed VBEC approach, the deep web pages are extracted directly from the database instead of developing the APIs. So, the proposed VBEC approach consumes less time for web page extraction. The variance in the time consumption is 40-50% less in the VBEC approach for web page data extraction.

Finally, it is being depicted that the proposed VBEC approach efficiently extract the user relevant information from the web database based on their query by adapting the enhanced co-citation algorithm, data record extraction and data item extraction process.

6. Conclusion

In VBEC approach, we have discussed the problem of efficient usage of web database. Especially, we addressed this issue by enhanced co-citation algorithm and extracting the deep web pages directly from the database instead of using APIs. With the VBEC approach, users achieved a great chance of extracting the abundant information from the web database effectively. So, the users obtain the desired information in the deep Web pages returned by Web databases based on their queries. In this paper, we focused on extracting the structured Web data, as well as data record extraction and data item extraction by adapting the enhanced co-citation algorithm. At first, the existing work limitations is analyzed and implement the enhanced co-citation algorithm with vision based approach to obtain the visual information of Web pages. Based on extracted web pages, a vision-based approach is presented to haul out structured data from deep Web pages. Experimental evaluation is efficiently conducted with the sample set of web database and processes the users' queries. Performance evaluation revealed that the proposed VBEC approach is better with 5-10% higher in precision rate, with 5-10 % increase in recall and time consumption reduced from 40-50% compared to the existing ViDE approach.

References

[1] Wei Liu, Xiaofeng Meng, Weiyi Meng," ViDE: A Vision-Based Approach for Deep Web Data Extraction", Ransactions On Knowledge And Data Engineering, Vol. 22, 2010

[2] D. Raghu, Sridhar Reddy, Raja Jacob, "Dynamic Vision-Based Approach in Web Data Extraction", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2, No. 6, 2011, pp. 2734-2736

[3] S.Vasantha Kumari M.K. Chandrasekaren ,"Deep Web Data Extraction Using A Vision based Approach", IJART, Vol. 2, No 2, 2012,

[4] W. Liu, X. Meng, and W. Meng, "Vision-Based Web Data Records Extraction," Proc. Int'l Workshop Web and Databases (WebDB '06), June 2006, pp. 20-25.

[5] Y. Lu, H. He, H. Zhao, W. Meng, and C.T. Yu, "Annotating Structured Data of the Deep Web," Proc. Int'l Conf. Data Eng. (ICDE), 2007, pp. 376-385

[6] H. Zhao, W. Meng, and C.T. Yu, "Automatic Extraction of Dynamic Record Sections from Search Engine Result Pages," Proc. Int'l Conf. Very Large Data Bases (VLDB), 2006, pp. 989-1000.

[7] Tianqi Yang , Taofen Qiu, "A method of automatic web information extraction based on page clustering", 9th World Congress on Intelligent Control and Automation (WCICA), 2011.

[8] Devika K, Subu Surendran, "An Overview of Web Data Extraction Techniques", International Journal of Scientific Engineering and Technology, Vol. 2, No. 4, 2013.

[9] K. Simon and G. Lausen, "ViPER: Augmenting Automatic Information Extraction with Visual Perceptions," Proc. International Conference on Information and Knowledge Management (CIKM), 2005.

[10] C.-H. Chang and S.-C. Kuo. OLERA: A semi-supervised approach for web data extraction with visual support. IEEE Intelligent Systems (SCI, EI), Vol. 19, No. 6, 2004, pp. 56–64.

A NEW METHOD TO SOLVE MULTI-OBJECTIVE NON-LINEAR FRACTIONAL PROGRAMMING PROBLEMS

Ramesh babu.R[1], Palanisamy.V[2], Divyaa devi.S[3]

[1] Alagappa University, Department of Computer science and Engineering,
Karaikudi, India - 630 003.

[2] Alagappa University, Department of Computer science and Engineering,
Karaikudi, India - 630 003.

[3] Anna University of Technology Madurai, Dindigul Campus,
Dindigul, India - 624 622.

Abstract

The multi-objective optimization method proposed in this paper does not convert multi-objective problem into a single objective problem. The given objective functions are treated as they are. A new type of transformation known as sum of ratios of objective functions is adopted. The contribution of decision variables to all the objective functions is determined in each iteration of the proposed algorithms. Solution that is extreme to all the objectives is determined.

1.0 Introduction

Existing methods to solve separable non-linear multi-objective optimization[11] problems transform the given objective functions in one way or other and convert them into single objective optimization problem and a compromise solution is obtained. When all the objectives are important, a compromise solution may not be optimal.

In this paper, a new method to solve large scale multi objective separable non-linear programming problems is presented. The proposed method solves the multi objective optimization problem by performing a new type of transformation on objective functions.

2.0 Linear Multi objective Optimization Problem

A separable non-linear multi objective optimization problem has a number of functions which are to be extremized. A seperable non-linear multi objective optimization problem with **t** number of objective functions, m number of constraints and n number of decision variables in its general form:

$$\text{Extremize} \quad Z(X) = \begin{Bmatrix} z_1(X) \\ z_2(X) \\ \cdots \\ z_t(X) \end{Bmatrix}$$

$$where \;\; z_i = \phi_i\left(x_1, x_2 ... x_n\right) + \phi_{i0} \qquad i = 1 ... t$$

Subject to the constraints,

$$\begin{aligned} &g_1\left(x_1, x_2 \ldots x_n\right) \le b_1 \\ &g_2\left(x_1, x_2 \ldots x_n\right) \le b_2 \\ &g_3\left(x_1, x_2 \ldots x_n\right) \le b_3 \\ &\quad : : : : : : : \\ &g_m\left(x_1, x_2 \ldots x_n\right) \le b_m \\ &and \;\; x_j \ge 0 \quad j = 1, 2, \ldots n \end{aligned}$$

Solution is a vector of n decision variables X=(x1, x2, ….xn) which extremizes multiple objectives.

2.1 Proposed Method

Proposed method finds solution to multi objective optimization problem by employing the idea of optimizing a sum of separable non-linear fractional functions[13]. Different orderings (permutation) of the given objective functions are formed. Problem is solved by taking different orderings of objective functions. Each ordering is used to form a sum of ratios objective function. Promising variables are identified, arranged and allowed to enter into basis for each ordering of objective functions. Promising variables are allowed to enter into basis only if they improve the objective fraction value compared to the value in previous iteration. Solution is found in each iteration. Finally, the solution corresponding to the ordering that gives the best ratio value is chosen as the optimal solution.

Let z_1, z_2, z_3 …. z_t be the objective functions. All the possible ordering of objective functions are formed. There will be ${}^tP_2$ number of different orderings. For example, with four objectives (t=4), there will be 24 unique orderings to form sum of objective fractions. Permutation set will contain 12 unique orderings after the removal of duplicate entries as shown below:

Let **tc** represent total number of permutations, i.e.

$tc=4p_2$

PN_set[tc][t]={(1,2,3,4),(1,2,4,3),(1,3,2,4),(1,3,4,2),

(1,4,2,3),(1,4,3,2),(2,1,4,3),(2,4,3,1),
(3,1,4,2),(3,4,2,1),(4,1,3,2),(2,3,4,1)}

A sum of ratio of 't/2' separable non-linear fractional functions[1,5,6,] with 't' number of objective functions. The sum of ratios[7] of objective fractions are formed for each ordering with objective functions as numerator and denominator part of the fractions as follows:

$$\frac{Z_1}{Z_2}+\frac{Z_3}{Z_4},\quad \frac{Z_1}{Z_2}+\frac{Z_4}{Z_3},\quad \frac{Z_1}{Z_3}+\frac{Z_2}{Z_4},\quad \frac{Z_1}{Z_3}+\frac{Z_4}{Z_2},$$

$$\frac{Z_1}{Z_4}+\frac{Z_2}{Z_3},\quad \frac{Z_1}{Z_4}+\frac{Z_3}{Z_2},\quad \frac{Z_2}{Z_1}+\frac{Z_4}{Z_3},\quad \frac{Z_2}{Z_4}+\frac{Z_3}{Z_1},$$

$$\frac{Z_3}{Z_1}+\frac{Z_4}{Z_2},\quad \frac{Z_3}{Z_4}+\frac{Z_2}{Z_1},\quad \frac{Z_4}{Z_1}+\frac{Z_3}{Z_2},\quad \frac{Z_2}{Z_3}+\frac{Z_4}{Z_1}$$

In general, the numerator and denominator parts of the objective fractions are formed using the recurrence relation:
For every **q** varying from **1** to **tc**

$$Z_{ni}=Z_{PN\_SET[q][2i-1]} \qquad Z_{di}=Z_{PN\_SET[q][2i]}$$

$$i=1,2\ldots t/2$$

2.2 Multi Objective separable non-linear Optimization Algorithm

The steps in multi objective separable non-linear optimization algorithm are as follows:

Step 1 Convert the separable non-linear function in to a linear function using the steps involved in piecewise linear approximation[12].

Step 2: Slack variables are added and the initial basic solution XB is determined. B matrix, CB and DB matrices are formed.

2a: Create the Permutation set PN_set with tc number of orderings where $tc=4P_2$ and t' is the number of objective functions.

2b: Let q=1

2c: Let Z_q = PN_set[q][o] where o = 1,2, … t.

2d: Exchange the objective rows in M^{-1} matrix as per the permutation PN_set[q].

2e: The optimality conditions of the solution are checked and promising variables are identified and arranged using the procedures 3.7.1 and 3.7.2. The set J is constructed. Let *l* be the number of elements in set J. Let p=1 and go to step 3a.

2f: Under non optimal conditions increment p, goto step 8 without updating the value of Z.

2g: Under optimal conditions, the feasibility is verified. The process is terminated if the solution is feasible.

Step 3a: Take the variable corresponding to the p^{th} element of the set J. Let the subscript of the selected variable be j.

For each of the decision variables x_j, the values of the net evaluations of the numerator and denominator are computed using the formula:

$$\begin{bmatrix} z_{n1}-c_1 \\ z_{d1}-d_1 \\ . \\ . \\ . \\ z_{nt}-c_t \\ z_{dt}-d_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \ldots\ldots & 0 & 0 & C_{B1}B^{-1} \\ 0 & 1 & 0 & \ldots\ldots & 0 & 0 & D_{B1}B^{-1} \\ . & & & & & & \\ . & & & & & & \\ . & & & & & & \\ 0 & 0 & 0 & \ldots\ldots & 1 & 0 & C_{Bt}B^{-1} \\ 0 & 0 & 0 & \ldots\ldots & 0 & 1 & D_{Bt}B^{-1} \end{bmatrix} \begin{bmatrix} -c_1 \\ -d_1 \\ . \\ . \\ . \\ -c_t \\ -d_t \\ P_j \end{bmatrix}$$

3b: Compute the ratios

$$R_i=\frac{(z_{ni}-c_i)}{(z_{di}-d_i)}\ \{i=1,2,3,..t\}$$

If all the ratios do not satisfy optimality conditions in, then go to step 8. If at least one ratio satisfies the optimality conditions, then x_j may still be promising.

3c: A new column α is constructed using the formula

$\alpha=(B^{-1}P_j)$.

The solution vector $(B^{-1}P_0)$ is also obtained.

Step 4a: Procedure to find leaving vector.

If the i^{th} constraint is an equality constraint and the i^{th} basic variable is not a decision variable, find

$$r_{1j} = \min_{i=1...m} \left\{ \frac{(B^{-1}P_0)_i}{\alpha_i} \; ; \alpha_i > 0 \right\}$$

If r_{1j} exists then the ith basic variable corresponding to r_{1j} is the leaving variable, and r_{1j} is the value of the leaving variable.

i.e., $x_j = r_{1j}$, Goto step 5, else goto step 4b.

Step 4b: If the i^{th} constraint is a lower bound constraint and the i^{th} variable is not a decision variable, find

$$r_{2j} = \max_{i=1...m} \left\{ \frac{(B^{-1}P_0)_i}{\alpha_i} ; \alpha_i > 0 \right\}$$

If r_{2j} exists then the i^{th} variable corresponding to r_{2j} is the leaving variable and $x_j = r_{2j}$ Go to Step 5 else go to Step 2c.

Step 4c: If the i^{th} basic variable is a decision variable or i^{th} constraint is an upper bound constraint, then find

$$\theta_i = \left\{ \frac{(B^{-1}P_0)_i}{\alpha_i} ; \alpha_i > 0 \right\}$$

If the i^{th} basic variable is the feasible slack variable, then find

$$\theta_i = \left\{ \frac{(B^{-1}P_0)_i}{\alpha_i} ; \text{if } \alpha_i < 0, (B^{-1}P_0)_i < 0 \right\}$$

Compute $r_{3j} = \min_{i=1...m} \{\theta_i\}$

If r_{3j} exists, then i^{th} variable corresponding to r_{3j} is the leaving variable and $x_j = r_{3j}$

Step 5: Compute improvement formula

$$u_j = \sum_{i=1}^{t} \frac{-(z_{ni} - c_i) * x_j + z_{ni}}{-(z_{di} - d_i) * x_j + z_{di}}$$

If value of u_j is greater than previous value of Z then go to step 6 else goto step 8.

Step 6: Computation of E matrix

The transformation matrix corresponding to the new entering and leaving variables can be obtained by using the product form of inverse. In the first step η vector is computed. Let j be the subscript of the entering variable and r be the column vector in which j^{th} variable enters. The column vector corresponding to the j^{th} vector is used to calculate η vector by using the relation.

$$\eta_1 = z_{n1} - c_1$$
$$\eta_2 = z_{d1} - d_1$$
$$\eta_3 = z_{n2} - c_2$$
$$\eta_4 = z_{d2} - d_2$$
$$\eta_n = z_{nt} - c_t$$
$$\eta_d = z_{dt} - d_t$$
$$\eta_{t+2} = \left[B^{-1}P_j \right]_i \; i = 1 .. m$$

Computation of η_{new}

Since the variable corresponding to the r^{th} row is the leaving variable $(r+2t)^{th}$ element in the η vector is the pivotal element. η_{new} can be obtained using the following relation:

$$\eta_{i\ new} = \frac{-\eta_{i\ old}}{\left(B^{-1}P_j\right)_r} \quad when \;\; i \neq r$$

$$\eta_{r\ new} = \frac{-\eta_{i\ old}}{\left(B^{-1}P_j\right)_r}$$

Replace the r^{th} column of (m+2t) x (m+2t) unit matrix by the η_{new} vector. The resulting matrix is the transformation matrix E.

Step 7: Computation of M^{-1}

Compute the inverse matrix for the next iteration as

$M^{-1}_{next} = E * M^{-1}_{current}$

Step 8: Let p = p + 1. If $p \le l$ go to step 3a else go to step 1f.

Step 9: Compute the optimal solution using the relation

$$\begin{bmatrix} z_1 \\ z_2 \\ .. \\ z_t \\ x_B \end{bmatrix} = M^{-1} \text{x} \begin{bmatrix} c_{01} \\ d_{01} \\ .. \\ c_{0t} \\ d_{0t} \\ P_0 \end{bmatrix}$$

Step 10: Compute value of Z for the current ordering of objective functions and save optimal solution and value of Z.

Step 11: If q < tc, Let q = q + 1 and go to step 2d.

Step 12: Arrange the Z value in ascending order and output the best solution set with maximum Z value.

2.3 Illustration of Multi-objective non-linear programming problem

Maximize

$$Z_1 = 5x_1^2 + 3x_2^2 + 1$$
$$Z_2 = 5x_1^2 + 2x_2^2 + 1$$
$$Z_3 = 4x_1^2 + 9x_2^2 + 2$$
$$Z_4 = 7x_1^2 + 5x_2^2 + 2$$

Subject to

$$3x_1^2 + 5x_2 \le 15$$
$$5x_1 + 2x_2^2 \le 10$$
$$where\ x_1, x_2 \ \ge 0$$

Phase I

Step 1: Conversion of Multi-objective Non-linear programming problem into sum of ratios linear fractional form. Find the maximum value t_j of the variable x_j which satisfies all the constraints.

$$t_1 = The\ \max imum\ value\ of$$
$$x_1 = \min\left\{\sqrt{\frac{15}{3}}, \frac{10}{5}\right\} = 2$$
$$t_2 = The\ \max imum\ value\ of$$
$$x_2 = \min\left\{\frac{15}{5}, \sqrt{\frac{10}{2}}\right\} = \sqrt{5}$$

Step 2: Let m_1=8 and m_2=8 where m1, m2 are the number of mesh points for the variables x_1 and x_2 respectively.

Then

$$\Delta\ for\ x_1 = \frac{t_1}{m_1} = \frac{2}{8} = \frac{1}{4}$$
$$and\ \Delta\ for\ x_2 = \frac{t_2}{m_2} = \frac{\sqrt{5}}{8}$$

The modified objective functions are

$$Z_1 = 0y_1 + \frac{5}{16}y_2 + \frac{5}{4}y_3 + \frac{45}{16}y_4 + 5y_5 + \frac{125}{16}y_6 + \frac{45}{4}y_7 + \frac{245}{16}y_8 + 20y_9 + 0y_{10} + \frac{15}{64}y_{11} + \frac{15}{16}y_{12} + \frac{135}{64}y_{13} + \frac{15}{4}y_{14} + \frac{375}{64}y_{15} + \frac{135}{16}y_{16} + \frac{735}{64}y_{17} + 15y_{18}$$

$$Z_2 = 0y_1 + \frac{5}{16}y_2 + \frac{5}{4}y_3 + \frac{45}{16}y_4 + 5y_5 + \frac{125}{16}y_6 + \frac{45}{4}y_7 + \frac{245}{16}y_8 + 20y_9 + 0y_{10} + \frac{5}{32}y_{11} + \frac{5}{8}y_{12} + \frac{45}{32}y_{13} + \frac{5}{2}y_{14} + \frac{125}{32}y_{15} + \frac{45}{8}y_{16} + \frac{245}{32}y_{17} + 10y_{18}$$

$$Z_3 = 0y_1 + \frac{1}{4}y_2 + y_3 + \frac{9}{4}y_4 + 4y_5 + \frac{25}{4}y_6 + 9y_7 + \frac{49}{4}y_8 + 16y_9 + 0y_{10} + \frac{45}{64}y_{11} + \frac{45}{16}y_{12} + \frac{405}{64}y_{13} + \frac{45}{4}y_{14} + \frac{1125}{64}y_{15} + \frac{405}{16}y_{16} + \frac{2205}{64}y_{17} + 45y_{18}$$

$$Z_4 = 0y_1 + \frac{7}{16}y_2 + \frac{7}{4}y_3 + \frac{63}{16}y_4 + 7y_5 + \frac{175}{16}y_6 + \frac{63}{4}y_7 + \frac{343}{16}y_8 + 28y_9 + 0y_{10} + \frac{25}{64}y_{11} + \frac{25}{16}y_{12} + \frac{225}{64}y_{13} + \frac{25}{4}y_{14} + \frac{625}{64}y_{15} + \frac{225}{16}y_{16} + \frac{1225}{64}y_{17} + 25y_{18}$$

The Constraints are

$$0y_1 + \frac{3}{16}y_2 + \frac{3}{4}y_3 + \frac{27}{16}y_4 + 3y_5 + \frac{75}{16}y_6 + \frac{27}{4}y_7 + \frac{147}{16}y_8 + 12y_9 + 0y_{10} + \frac{5\sqrt{5}}{8}y_{11} + \frac{5\sqrt{5}}{4}y_{12} + \frac{15\sqrt{5}}{8}y_{13} + \frac{5\sqrt{5}}{2}y_{14} + \frac{25\sqrt{5}}{8}y_{15} + \frac{15\sqrt{5}}{4}y_{16} + \frac{35\sqrt{5}}{8}y_{17} + 5\sqrt{5}y_{18} = 15$$

$$0y_1 + \frac{5}{4}y_2 + \frac{5}{2}y_3 + \frac{15}{4}y_4 + 5y_5 + \frac{25}{4}y_6 + \frac{15}{2}y_7 + \frac{35}{4}y_8 + 10y_9 + 0y_{10} + \frac{5}{32}y_{11} + \frac{5}{8}y_{12} + \frac{45}{32}y_{13} + \frac{5}{2}y_{14} + \frac{125}{32}y_{15} + \frac{45}{8}y_{16} + \frac{245}{32}y_{17} + 10y_{18} = 10$$

$$y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 + y_9 = 1$$

$$y_{10} + y_{11} + y_{12} + y_{13} + y_{14} + y_{15} + y_{16} + y_{17} + y_{18} = 1$$

In a multi-objective fractional problem various sum of ratios are formed taking any two objective function to form a ratio.

Problem 2.3 $\frac{Z_2}{Z_1}+\frac{Z_3}{Z_4}$ Initial Solution

	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13	y14	y15	y16	y17	y18	RHS
Z_2	0	$-\frac{5}{16}$	$-\frac{5}{4}$	$-\frac{45}{16}$	–5	$-\frac{125}{16}$	$-\frac{45}{4}$	$-\frac{245}{16}$	–20	0	$-\frac{5}{32}$	$-\frac{5}{8}$	$-\frac{45}{32}$	$-\frac{5}{2}$	$-\frac{125}{32}$	$-\frac{45}{8}$	$-\frac{245}{32}$	–10	1
Z_1	0	$-\frac{5}{16}$	$-\frac{5}{4}$	$-\frac{45}{16}$	–5	$-\frac{125}{16}$	$-\frac{45}{4}$	$-\frac{245}{16}$	–20	0	$-\frac{15}{64}$	$-\frac{15}{16}$	$-\frac{135}{64}$	$-\frac{15}{4}$	$-\frac{375}{64}$	$-\frac{135}{16}$	$-\frac{735}{64}$	–15	1
$\frac{Z_2}{Z_1}$	0	1	1	1	1	1	1	1	1	0	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	1
Z_3	0	$-\frac{1}{4}$	–1	$-\frac{9}{4}$	–4	$-\frac{25}{4}$	–9	$-\frac{49}{4}$	–16	0	$-\frac{45}{64}$	$-\frac{45}{16}$	$-\frac{405}{64}$	$-\frac{45}{4}$	$-\frac{1125}{64}$	$-\frac{405}{16}$	$-\frac{2205}{64}$	–45	2
Z_4	0	$-\frac{7}{16}$	$-\frac{7}{4}$	$-\frac{63}{16}$	–7	$-\frac{175}{16}$	$-\frac{63}{4}$	$-\frac{343}{16}$	–28	0	$-\frac{25}{64}$	$-\frac{25}{16}$	$-\frac{225}{64}$	$-\frac{25}{4}$	$-\frac{625}{64}$	$-\frac{225}{16}$	$-\frac{1225}{64}$	–25	2
$\frac{Z_3}{Z_4}$	0	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{4}{7}$	0	$\frac{9}{5}$	$\frac{9}{5}$	$\frac{9}{5}$	$\frac{9}{5}$	$\frac{9}{5}$	$\frac{9}{5}$	$\frac{9}{5}$	$\frac{9}{5}$	1
$\frac{Z_2}{Z_1}+\frac{Z_3}{Z_4}$	0	$\frac{11}{7}$	$\frac{11}{7}$	$\frac{11}{7}$	$\frac{11}{7}$	$\frac{11}{7}$	$\frac{11}{7}$	$\frac{11}{7}$	$\frac{11}{7}$	0	$\frac{37}{15}$	$\frac{37}{15}$	$\frac{37}{15}$	$\frac{37}{15}$	$\frac{37}{15}$	$\frac{37}{15}$	$\frac{37}{15}$	$\frac{37}{15}$	2
$C1$	0	$\frac{3}{16}$	$\frac{3}{4}$	$\frac{27}{16}$	3	$\frac{75}{16}$	$\frac{27}{4}$	$\frac{147}{16}$	12	0	$\frac{5\sqrt{5}}{8}$	$\frac{5\sqrt{5}}{4}$	$\frac{15\sqrt{5}}{8}$	$\frac{5\sqrt{5}}{2}$	$\frac{25\sqrt{5}}{8}$	$\frac{15\sqrt{5}}{4}$	$\frac{35\sqrt{5}}{8}$	$5\sqrt{5}$	15
$C2$	0	$\frac{5}{4}$	$\frac{5}{2}$	$\frac{15}{4}$	5	$\frac{25}{4}$	$\frac{15}{2}$	$\frac{35}{4}$	10	0	$\frac{5}{32}$	$\frac{5}{8}$	$\frac{45}{32}$	$\frac{5}{2}$	$\frac{125}{32}$	$\frac{45}{8}$	$\frac{245}{32}$	10	10
$C3$	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1
$C4$	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1

SOLUTION TABLE

Sl.No	Objective	Solution		Objective Function Value
		X_1	X_2	
1.	$\frac{Z_1}{Z_2}+\frac{Z_3}{Z_4}$	0	$\sqrt{5}$	$\frac{930}{319}$
2.	$\frac{Z_2}{Z_1}+\frac{Z_3}{Z_4}$	0	$\sqrt{5}$	$\frac{1049}{432}$
3.	$\frac{Z_1}{Z_2}+\frac{Z_4}{Z_3}$	0	$\sqrt{5}$	$\frac{1049}{517}$
4.	$\frac{Z_2}{Z_1}+\frac{Z_4}{Z_3}$	0	$\sqrt{5}$	$\frac{949}{297}$
5.	$\frac{Z_1}{Z_3}+\frac{Z_2}{Z_4}$	2	0	$\frac{28}{15}$
6.	$\frac{Z_1}{Z_3}+\frac{Z_4}{Z_2}$	0	$\sqrt{5}$	$\frac{1445}{517}$
7.	$\frac{Z_3}{Z_1}+\frac{Z_2}{Z_4}$	0	$\sqrt{5}$	$\frac{1445}{432}$
8.	$\frac{Z_3}{Z_1}+\frac{Z_4}{Z_2}$	0	$\sqrt{5}$	$\frac{949}{176}$
9.	$\frac{Z_1}{Z_4}+\frac{Z_2}{Z_3}$	2	0	$\frac{1008}{540}$
10.	$\frac{Z_1}{Z_4}+\frac{Z_3}{Z_2}$	2	0	$\frac{109}{42}$
11.	$\frac{Z_4}{Z_1}+\frac{Z_2}{Z_3}$	2	0	$\frac{981}{378}$
12.	$\frac{Z_4}{Z_1}+\frac{Z_3}{Z_2}$	0	$\sqrt{5}$	$\frac{1049}{176}$

The Value of the ratio $\frac{Z_4}{Z_1}+\frac{Z_3}{Z_2}$ is 5.960227 which is maximum, and the best solution obtained.

TABLE I

	Sign of z_{nj}	Sign of z_{dj}	Sign of $z_{nj}-c_j$	Sign of $z_{dj}-d_j$	Sign of R_j	Condition for Improvement
1	Positive	Positive	+	+	+	Positive $R_j < Z_{nj} / Z_{dj}$
2	Positive	Positive	-	-	+	Positive $R_j > Z_{nj} / Z_{dj}$
3	Positive	Positive	+	-	-	No Improvement
4	Positive	Positive	-	+	-	**1.1 Negative $R_j < Z_{nj} / Z_{dj}$**
5	Negative	Positive	+	+	+	No Improvement (i) Provided there is no change of sign for Z_{dj} (ii) Check for condition Number 16 if there is change in sign of Z_{dj}
6	Negative	Positive	-	-	+	Positive $R_j > Z_{nj} / Z_{dj}$
7	Negative	Positive	+	-	-	Negative $R_j > Z_{nj} / Z_{dj}$
8	Negative	Positive	-	+	-	Negative $R_j < Z_{nj} / Z_{dj}$
9	Positive	Negative	-	-	+	No improvement (i) Provided there is no change of sign for Z_{dj} (ii) Check for condition number 16 if there is change in sign of Z_{dj}.
10	Positive	Negative	-	+	-	Negative $R_j > Z_{nj} / Z_{dj}$
11	Positive	Negative	+	-	-	Negative $R_j < Z_{nj} / Z_{dj}$
12	Positive	Negative	+	+	+	Positive $R_j > Z_{nj} / Z_{dj}$
13	Negative	Negative	-	-	+	Positive $R_j < Z_{nj} / Z_{dj}$
14	Negative	Negative	+	-	-	Negative $R_j < Z_{nj} / Z_{dj}$
15	Negative	Negative	-	+	-	**1.2 No Improvement**
16	Negative	Negative	+	+	+	Positive $R_j > Z_{nj} / Z_{dj}$

Conclusion:

A new algorithm for solving a multi objective non-linear optimization problems with 't' number of objectives is presented. Optimal solution is determined by exploring different orderings of objective functions playing the roles of numerators and denominators of objective fractions. In this way the multi objective optimization problem is approached in a new pattern.

Variables are allowed to enter only if they improve the value of Z (sum of ratios of objectives). Promisibility is checked by calculating the Z value every time a variable enters into basis. Variables are not merely allowed to enter the basis even after ordering them, because previously entered variable may change the value of Z and affects the next promising variable.

The algorithm finds the best optimal solution among the solutions found for each ordering of objectives. The solution is best if it gives the best value for the sum of ratios. Algorithm is implemented using C language program. Efficiency of the algorithm is computationally tested by solving large scale multi objective optimization problems. The results are tabulated.

The proposed algorithm is computationally tested and it is found that the execution time and number of iterations are saved by this algorithm.

References:

[1] Frenk, J.B.G., Scheible, S.: Fractional programming. In: Floudas, C.A., Pardalos, P.M. (eds.) Encyclopedia of Optimization, vol. II, pp. 162–172. Kluwer, Dordrecht (2001)

[2] Rao, M.R.: Cluster analysis and mathematical programming. J. Am. Stat. Assoc. **66**, 622–626 (1971)

[3] Almogy, Y., Levin, O.: Parametric analysis of a multi-stage stochastic shipping problem. In: Proceedings of the 5th IFORS Conference (IFORS), pp. 359–370 (1964)

[4] Colantoni, C.S., Manes, R.P., Whinston, A.: Programming, profit rates, and pricing decisions. Account. Rev. 44, 467–481 (1969)

[5] Konno, H., Inori, M.: Bond portfolio optimization by bilinear fractional programming. J. Oper. Res. Soc. Jpn. 32, 143–158 (1989)

[6] Schaible, S.: Fractional programming. In: Horst, R., Pardalos, P. (eds.) Handbook of Global Optimization, pp. 495–608. Kluwer, Dordrecht (1995)

[7] Schaible, S.: Fractional programming with sums of ratios. In: Castagnoli, E., Giorgi, G. (eds.) Scalar and Vector Optimization in Economic and Financial Problems, Proceedings of the Workshop in Milan, 1995, pp. 163–175. Elioprint, Modena (1995)

[8] Naganathan E R (1997). A Ratio Multiplex Algorithm to solve large scale linear programming problems, Ph.D. thesis, Department of Computer Science and Engineering, Alagappa University, Karaikudi

[9] Nagarajan S, "A Complex Algorithm to Solve Large Scale Linear Programming Problems", Ph.D. thesis, Department of Computer Science and Engineering, Alagappa University, Karaikudi

[10] Sakthivel,S., "A Multiplex algorithm to solve large scale linear programming problems", Ph.D. thesis, Anna University, India, 1984.

[11] Meyyappan. T "A new Computer Oriented Technique To Solve Large Scale Multi Objective Optimization Problems", ", Ph.D. thesis, Department of Computer Science and Engineering, Alagappa University, Karaikudi

[12] Palanisamy.V "Computer oriented algorithms to solve special cases of non-linear fractional programming problems", ", Ph.D. thesis, Department of Computer Science and Engineering, Alagappa University, Karaikudi

[13] Ramesh babu.R Divyaa devi.S Palanisamy.V "A New Computer Oriented Technique to Solve Sum Of Ratios Non-Linear Fractional Programming Problems" International Journal of Computing Algorithm Volume: 01 Issue: 01 January-June 2012.

Handwritten Arabic Digits Recognition Using Bézier Curves

Aissa Kerkour El Miad and Azzeddine Mazroui

University Mohammed First, Faculty of Sciences, Oujda, Morocco

Abstract

In this paper we propose a new recognition approach for Arabic numerals. Given that the performance of recognition systems for Arabic numerals are closely linked to the choice of features and classification system used in the recognition phase, we seek to exploit the possibilities of the theory of Bézier curves that allows representing parametric curves from a limited number of data (some characteristic dots with their derivatives). Indeed, the characteristic dots of the Arabic digit that we have adopted are those such that their associated Bézier curve is close to the shape of the digit. The used classifier in this work is the k-nearest neighbour. The obtained results testify to the interest and the strength of our approach.

Keywords: Handwritten digits recognition, Image processing, Spline, Bézier curves, Skeletonization, Feature extraction, Training.

1. Introduction

The recognition of handwriting text is a topic widely studied by the scientific community. It has been in recent decades the subject of several research works. Some studies are devoted to the digit recognition given its growing interest in many applications such as postal mail sorting and bank check processing. The accuracy and speed of execution of these applications are widely characteristics that are sought by users. However, the diversity of writing styles of writers makes this subject difficult and stimulates many researchers to develop high performance applications. The performances of recognition systems depend strongly on the choice of approaches used in the feature extraction approach and the classification techniques relating to training and testing phases.

Among the most used feature extraction approaches, we quote statistical methods based on histograms [1-3], and those using local variations in digit shapes [4-5]. Our approach fits in line with the approaches of the second category. In fact, the features that we have adopted are the dots where the digit shape shows a strong variation (change of direction, inflection dots and cusps) accompanied by their tangents. The theory of Bézier curves explains the choice of these features [6]. Indeed, an appropriate selection of a limited number of dots with their tangents allows building very close shape to that of digit.

A broad family of classifiers have been used by several research teams. The artificial neural networks (ANN), the k-nearest neighbours (k-NN), the hidden Markov models (HMM) and the support vector machine (SVM) are among the most frequently used classifiers [7]. To improve the performance of recognition systems, some authors have used hybrid classifiers which consist in using two classifiers [8]. In this work we have used the 1-nearest neighbour classifier.

The paper is organized as follows. In Section 2 we give a state of the art related to Arabic recognition digit systems. Section 3 is devoted to the pre-processing steps. After, we recall in Section 4 the main properties of Bézier curve theory. We describe in Section 5 the proposed feature extraction approach. Then, we explain in Sections 6 and 7 respectively the training phase and the recognition phase. Finally, Section 8 addresses and analyzes the experimental results, and we end this paper by a conclusion and a brief description of future works.

2. State of the art

S. Mahmoud [9] presented a recognition system for handwritten Indian numerals. He used a set of 120 features computed from angle span, distance span, horizontal span and vertical span. The HMM and 1-NN were used as classifiers. He tested these classifiers with different sets of these 120 features in order to select the features and the classifier giving the highest recognition rate. He concludes that the results obtained with HMM classifier are better than the 1-NN classifier.

D. Sharma et al. [10] proposed a Zone feature extracted method for the recognition of handwritten numerals. It consists to divide firstly the whole image in 4×4 zones. In order to gain more accuracy these zones are divided into 6×6 zones. The division of zones carried out up to 8×8 zones. The features are the densities of object pixels in each zone. Finally, 116 of such features are extracted for classification and recognition. 1-NN classifier was used for classification and recognition.

S. Impedove et al. [2] developed a novel prototype generation technique to recognize handwritten digit. The features are computed from binary histograms of oriented gradients, and the k-NN classifier was used in two stage processes to reduce the classification time. The first step used the adaptive resonance theory to determine the number of prototypes and select an effective initial solution, and an evolution strategy was used in the second step to generate the final solution.

In order to improve the performance of recognition systems, some approaches based on hybrid classifier have been developed in recent years [8, 11, 12]. In [8], X. Niu et al. used the convolutional neural network to extract the features and the SVM classifier to recognize the unknown pattern.

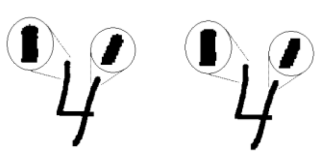
A state of the art of the main methods developed in the field of OCR for Arabic was made by L. M. Lorigo et al. [13], and more recently by A. M. AL-Shatnawi et al. [14] and by A. Mesleh et al. [15]. For digits, C. L. Liu et al. describes in [1] the different techniques used in digit recognition systems.

3. Pre-processing

Before extracting the features of each digit image, a pre-processing step is required. It consists in removing the unnecessary information in the image and keep only useful information.

3.1 Removing noise

The approach that we have adopted in the feature extraction phase is to use the skeleton of the digit instead of its original form. Following the phase of skeletonization (see paragraph 2 below), some branches appear in the skeleton of the digit in the form of noise. During the feature extraction phase, these branches can be detected as primitives. So, we conducted a filtering before skeletonization phase to prevent the appearance of these branches (see fig 1).



(a) (b)

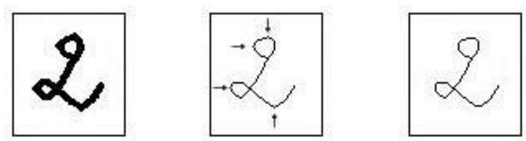
Fig. 1 - (a): before filtering ; (b): after filtering.

3.2 Skeletonization

In many cases, the treatment of skeleton of the digit instead of its raw form is less expensive in terms of time and more interesting in terms of accuracy. Thus, we chose to analyze the skeleton of the digit instead of its initial shape. The skeletonization algorithm that we used is that developed by Zhang-Wang [16]. It is known by its speed and its adaptation to Arabic numerals (see Fig. 2(b)).

3.3 Straightening of shapes

Sometimes we noticed, after the skeletonization phase, the onset of a quirky pixel in straight parts of the skeleton. To avoid being selected as feature in the next step, we proceed to the straightening of these pixels (see Fig. 2(c)).



(a) (b) (c)

Fig. 2 (a) The initial digit ; (b) The digit after skeletonization ; (c) The skeleton after adjustment of shape

3.4 Resizing

A standardization phase of image sizes is necessary in order to compare the features of the digit to recognize to those of learned digit. To do this, we first frame the digit (i.e. identify the smallest rectangle containing the digit (see Fig. 2 (d)), then we put the framed image at the center of a 128 × 128 window (see fig. 2 (e)).

4. Bézier Model

Bézier curves are parametric piecewise polynomial curves. They were introduced for the first time by Pierre Bézier [2]. They are easy to build and they have interesting properties for graphic design. Indeed, for each four distinct dots P_0, P_1, P_2 and P_3, there exists a unique cubic Bézier curve which starts at P_0 and arrives at P_3, and has the vectors P_0P_1 and P_2P_3 as tangent vectors respectively at the dots P_0 and P_3. The shape of this cubic curve is controlled by the envelope P_0P_1, P_1P_2 and P_2P_3 (see Fig. 3).

Furthermore, if we move only one dot P_i, we obtain variations of the initial curve. This is used by typographers to refine the plotted curves (see Fig. 3).

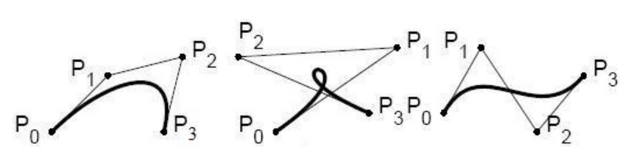


Fig.3 According to positions of four dots P_i, we obtain various forms of Bézier curves.

The continuous splines of degree 3 are obtained by connecting these curves. So, the skeleton of the digit can be seen as a continuous spline (see Fig. 3(d)). The features of the digit are the points and the derivatives which the associated spline is close to the shape of the skeleton. To

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

illustrate this feature extraction method, we treat the case of the digit 2. We partition the shape of this digit into two curves. The first curve starts at Q_1 and arrives at Q_2 and the second starts at Q_2 and arrives at Q_3. After building a spline from dots Q_i and adapted tangents at these dots, we obtain a shape close to that of the digit 2 (see Fig. 4).

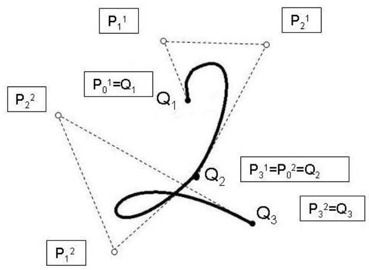


Fig. 4 Bézier data required to reconstruct the digit 2.

Thus, we can characterize each digit by a limited number of dots equipped with their tangents. The Features are the extremity dots of the digit, in addition to some dots where the shape of the digit presents a variation (changes of direction, inflexion dots and cusps).

5. Features extraction

After the pre-processing step, the resulting shape is a sufficiently smooth skeleton. A study on local variations in the skeleton shape will allow us to identify the features.

5.1 Digit classes

We distinguish two classes of the Arabic digits:

- the class LD of digits that have a loop,
- the class NLD of digits without a loop.

By analyzing the shapes of digits written by different writers, we noticed that some writers write the numeral 2 with a loop and 4 without loop, so:

- LD class is composed of digits 0, 4, 6, 8, 9 and the digit 2 written with a loop.
- NLD class is composed of remaining digits and the digit 4 written without loop.

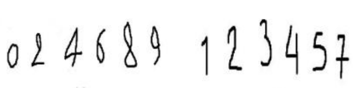


Fig. 5 (a): class LD ; (b): class NLD

A digit is an element of the LD class if browsing its skeleton in a given direction we meet a pixel twice.

5.2 Extremity dots

As the extremity pixels have only one neighbour, their identification will be through one of the following eight masks:

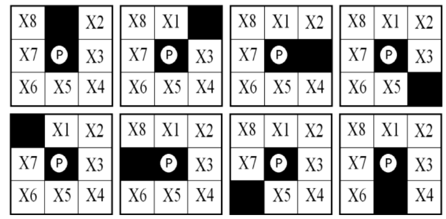


Fig. 6 detection masks of extremity pixels

5.3 Singular dots

For each pixel, we consider the eight possible derivative directions d_k:

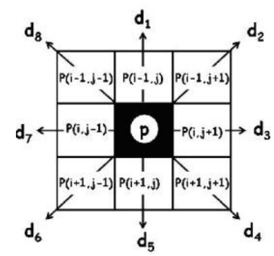


Fig. 7 Eight derivative directions of the pixel P

Definition: A pixel P is non-singular if it is aligned with its two neighbour pixels, i.e. the two associate directions d_i and d_k of its derivatives are collinear, which is equivalent to $|i-k|=4$.

The two first masks below are examples of used masks to detect singular dots and the last one is an example of used masks to detect non-singular dots.

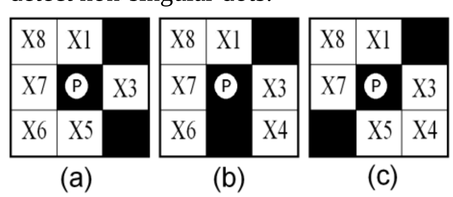


Fig. 8 (a) and (b): singular pixels ; (c): non-singular pixel

The first step of our algorithm consists in removing the non-singular pixels.

5.4 Characteristic dots

The change in directions of derivatives for some singular dots is not necessary due to a strong variation in the digit shape, but can be a result of the digital image processing or the skeletonization algorithm. In this case, these dots should not be considered as features. Hence, we use a second filter to remove them. It consists to eliminate any singular dot P that is almost aligned with its nearest neighbours P_b and P_a, where P_b (resp. P_a) is the singular dot detected just before (resp. just after) P. In the case where P is the first detected singular dot, the dot P_b will be the first extremity dot and where P is the last detected singular pixel, P_a will be the last extremity dot.

For a given digit D, we denote by P_1 and P_m the two extremity dots and $(P_i)_{2\leq i\leq m-1}$, the singular dots of D. Let (u_i,v_i) the two derivative directions of the dot P_i, with the extension $u_0 = v_m = 0$.

Lemma: for each $2\leq i \leq m-1$*, the directions* v_{i-1} *and* u_i *are parallel and the directions* v_{i-1} *and* v_i *are not parallel.*

Proof: as P_i is the first singular dot detected after the singular dot P_{i-1}, it is clear that the directions v_{i-1} and u_i are parallel. Similarly, since the directions u_i and v_i are not parallel (because P_i is a singular dot), we conclude that the directions v_{i-1} and v_i are also not parallel. ■

Thus, for a given singular dot P_i, if $P_{\hat{\imath}}$ is the singular dot judged as characteristic dot just before P_i, then we are in front of one of the two following situations:

- First case: the angle between the direction v_i and one of the above directions v_j, $\hat{\imath} \leq j \leq i-1$, is a right angle. In this case, the dot P_i is a real cusp and will thereafter be considered as a characteristic dot of the digit D.
- Second case: the angle between the direction v_i and all previous directions v_j, $\hat{\imath} \leq j \leq i-1$, is a non-right angle (so it's equal to 135° or 180°). In this case, the dot P_i is a false cusp and should not be considered as a characteristic dot of the digit D.

The following algorithm shows how to detect the characteristic dots from the singular dots $(P_i)_{1\leq i\leq m}$:

Algorithm to detect the characteristic dots from the singular dots

```
The starting dot P1 is considered as a characteristic dot
i = 2
î =1
while i≤m-1
    for j=î:i-1
        if (the angle between the direction vi and the
              direction vj is a right angle)
              Pi is a characteristic dot
              î=î+1
              i=i+1
        else
              i=i+1
        endif
    endfor
endwhile
The last dot Pm is considered as a characteristic dot
```

So, the features of the digit D are:

1. the obtained characteristic dots $(Q_i)_{1\leq i\leq r}$ $(r \leq m)$,
2. the associated derivatives $(u_i, v_j)_{1\leq i\leq r}$, ($(u_i, v_j)$ are the two derivative directions of the dot Q_i),
3. the associated information $(e_i)_{1\leq i\leq r}$ identifying the characteristic points that belong to a loop (e_i=1 if the dot Q_i belongs to a loop and e_i=0 otherwise),
4. the class affiliation c, ($c \in LD, NLD$).

So, the features of the digit D are its class affiliation and the following characteristic matrix:

$$\begin{pmatrix} Q_1 & Q_2 & \dots & Q_r \\ u_1 & u_2 & \dots & u_r \\ v_1 & v_2 & \dots & v_r \\ e_1 & e_2 & \dots & e_r \end{pmatrix}$$

6. Training phase

The training phase consists in choosing some writers and characterizing the different digits written by these writers. Indeed, given n writers S_i, $1 \leq i \leq n$, and a digit D, we denote by M_i the characteristic matrix of the digit D written by the writer S_i and c_i its class affiliation. As the sizes of the matrices M_i aren't necessarily equal (because we don't have the same writer), a standardization of matrix sizes is necessary.

6.1 Standardization of matrix sizes

Put $n^*=max\ n_i$, where n_i, $1 \leq i \leq n$, is the number of columns of the matrix M_i (n_i = the number of characteristic dots of the digit D written by the writer S_i).
Let M^* be one of the matrices M_i such that its number of columns is equal to n^*. The standardization approach of matrix sizes that we have adopted is to complete each matrix M_i for which $n_i < n^*$, by additional $(n^*- n_i)$ characteristic dots.

Let N^* (resp. N_i) be the sub matrix of M^* (resp. M_i) formed by the first two lines of M^* (resp. M_i) (it is obtained by keeping in M^* (resp. in M_i) only the coordinates of characteristic dots). First, we match each dot of the matrix N_i the nearest dot of the matrix N^* in the sense of Euclidean norm. So, it will remain (n^*-n_i) dots in N^* for which we have no counterpart in N_i. For each of these (n^*-n_i) dots of N^*, we seek the nearest dot of the digit D written by the writer S_i (see Fig. 9). By respecting the position of corresponding dot in the matrix M^*, we add to the matrix M_i this dot with its associated derivatives and information relating to its belonging or not to a loop. So we get a new matrix M_i^* with the same size as M^*.

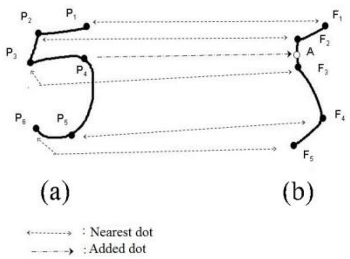


Fig. 9 Standardization of sizes applied to digit 5

6.2 Digit features

The characteristic matrices $(M_i^*)_i$ of the digit D according to the writers $(S_i)_i$ are all of the same size ($5\times n^*$). Thus, the features of the digit D are:

1. The characteristic matrix M_D equal to the mean of the matrices $(M_i^*)_i$:

$$M_D = \frac{1}{n}\sum_{i=1}^{n} M_i^*$$

2. The class affiliation c of the digit D equal to the value c_k which is the most common in the sequence $(c_i)_{1\leq i\leq n}$.

7. **Recognition**

By following the steps in the previous paragraph, we compute the features of each digit of classes LD and NLD. The recognition of an unknown digit D will occur in four steps.

7.1 Features of the digit D

We first compute the class affiliation and the characteristic matrix M_D of the digit D by following the steps developed in Paragraph 5. The next step consists to identify from characteristic matrices M_i of digits D_i belonging to the same class as D, the nearest matrix to $M_.$

7.2 Distance between M_D and the characteristic matrix M_i of the digit D_i

Since the matrices M_D and M_i do not necessarily have the same size, we first standardize their sizes. For this, we distinguish three cases on their respective numbers of columns n_D and n_i.

7.2.1 First case: $n_D = n_i$

We ordain the columns of the matrix M_i so that the j^{th} column of the obtained matrix is the closest to the j^{th} column of the matrix M_D.

7.2.2 Second case: $n_D < n_i$

To complete the matrix M_D by $(n_i - n_D)$ additional dots, we proceed as in sub-paragraph 6.1.

7.2.3 Third case: $n_i < n_D$

We first ordain the columns of the matrix M_i so that the j^{th} column, for $j \leq n_i$, of the obtained matrix is the closest to the j^{th} column of the matrix M_D. As M_i is the characteristic matrix of D_i according to several writers (this is the mean of characteristic matrices $(M_{ij})_{1\leq j\leq r}$ of the digit D_i according to the r writers $(w_j)_{1\leq j\leq r}$ used in training phase), we first complete each matrix M_{ij} by $(n_D - n_i)$ additional dots by following the steps of the sub-paragraph 6.1. After, we substitute the matrix M_i by the matrix M_i^* mean of these matrices which is the same size as M_D.
After the standardization phase of matrix sizes, we denote $M_i^*$and M_D^* the obtained matrices, and we put $d_i=\|M_i^*- M_D^*\|$ where $\| \;\|$ is the Frobenius norm.

7.3 Identification of the digit D

The digit D will be identified with the digit D_i^* satisfying the following minimization equation:

$$d_{i^*} = \min_i d_i$$

The minimum is taken for all digits belonging to the same class as the digit D.

8. **Recognition Results**

The database DB consists of 360 digits. Each digit between 0 and 9 has been written by 36 different writers (see Fig. 10).

Fig. 10 Sample of handwritten digits.

One part of DB, denoted Tr_DB was used in the training phase, and the rest, denoted Te_DB was reserved to evaluate the system.
We sought to identify the best choice of the set Tr_DB giving the highest recognition accuracy in the test phase.

For this, we denote by S_r the r[th] writer and *RR* the recognition rate.
Given an integer $k \geq 1$, and for any combination of k writers among 36 writers, we first use these k writers as *Tr_DB* in the training phase, and after compute the corresponding recognition rate *RR*. Finally, we identify the combination of *k* writers giving the highest *RR*.
The results obtained for $k < 3$ and $k > 8$ are not interesting. So, we give in Table 1 only the results for $3 \leq k \leq 8$.

Table 1: Set *Tr_DB* of k writers giving the best recognition rate (RR)

K	Set Tr_DB of k writers giving the best RR	RR (%)
3	S_8 ; S_3 ; S_{16}	91.94
4	S_8 ; S_3 ; S_{16} ; S_{22}	94.44
5	S_8 ; S_3 ; S_{16} ; S_{22} ; S_{26}	96.39
6	S_8 ; S_3 ; S_{16} ; S_{22} ; S_{26} ; S_{14}	96.94
7	S_8 ; S_3 ; S_{16} ; S_{22} ; S_{26} ; S_{14}; S_{25}	90.00
8	S_8 ; S_3 ; S_{16} ; S_{22} ; S_{26} ; S_{14}; S_{25} ; S_{18}	86.67

The best performance has been achieved when we use the six writers S_8, S_3, S_{16}, S_{22}, S_{26} and S_{14} in the training phase. The explanation that we can advance on high performance obtained with this list is that the writing styles of these writers cover the different writing styles of all writers. Recognition errors are mainly due to writing styles of some writers. Indeed, the digits 1, 4, 7 and 9 are in some cases very confused, and even humans have difficulties to identify them (see Fig. 10).
For more details, we give in Table 2 the confusion matrix along with the recognition rate of each digit. These results are related to the use of the optimal list (S_8,S_3,S_{16},S_{22},S_{26},S_{14}) in the training phase.

Table 2: Confusion Matrix and the recognition rate (RR) of each digit

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**RR (%)**
0	36	0	0	0	0	0	0	0	0	0	100
1	0	34	0	0	0	0	0	**2**	0	0	94,44
2	0	0	36	0	0	0	0	0	0	0	100
3	0	0	0	36	0	0	0	0	0	0	100
4	0	**4**	0	0	30	0	0	0	0	**2**	83,33
5	0	0	0	0	0	36	0	0	0	0	100
6	0	0	0	0	0	0	36	0	0	0	100
7	0	0	0	0	0	0	0	36	0	0	100
8	0	0	0	0	0	0	0	0	36	0	100
9	0	0	0	0	**3**	0	0	0	0	33	91,66

9. Conclusion

We presented in this work a new approach of digit recognition. It is based on the extraction the Hermite data from the digit shape (dots with their derivatives). The choice of this approach was dictated by the possibility of recovering a close shape to that of the digit using the Bézier curve theory on these data.
The obtained results are very interesting, and we plan to improve them using other classifiers (the artificial neural networks, the hidden Markov models and the support vector machine) during both training and testing phases. Similarly, we will enrich our database in order to perform tests on a more consistent data base.

References

[1] C. L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques", Pattern Recognition, Vol. 36, 2003, pp. 2271–2285.
[2] S. Impedovo, F.M. Mangini and D. Barbuzzi, "A novel prototype generation technique for handwriting digit recognition", Pattern Recognition, Available online 3 May 2013 http://dx.doi.org/10.1016/j.patcog.2013.04.016
[3] O. Rashnoodi, A. Rashnoodi and A. Rashnoodi", Off-line Recognition of Persian Handwritten Digits using Statistical Concepts", International Journal of Computer Applications, Vol. 53, No. 8, 2012, pp. 20–28.
[4] M. Shi, Y. Fujisawa, T. Wakabayashi and F. Kimura, "Handwritten numeral recognition using gradient and curvature of gray scale image", Pattern Recognition, Vol. 35, No. 10, 2002, pp. 2051–2059.
[5] A. Mazroui, Aissa Kerkour El Miad, "*Handwritten Arabic characters modeling by Bézier curves for recognition",* in proceeding of 4[th] International Conference on Approximation Methods and Numerical Modelling in Environment and Natural Resources, 2011, pp. 535-538.
[6] G. Farin, and P. Massart, Curves and Surfaces for Computer Aided Geometric Design, San Diego, CA, Academic Press, 1983.
[7] S. Arora, D. Bhattacharjee, M. Nasipuri, L. Malik, M. Kundu and D. K. Basu, "Performance Comparison of SVM and ANN for handwritten Devnagari Character Recognition", International Journal of Computer Science Issues, Vol. 7, No. 6, 2010, pp. 18-26.
[8] X. X. Niu, and C. Y. Suen "A novel hybrid CNN–SVM classifier for recognizing handwritten digits", Pattern Recognition, Vol. 45, No. 4, 2012, pp. 1318–1325.
[9] S. Mahmoud, "Recognition of writer-independent off-line handwritten Arabic (Indian) numerals using hidden Markov models", Signal Processing, Vol. 88, No. 4, 2008, pp. 844–857.
[10] D. Sharma, and D. Gupta, "Isolated Handwritten Digit Recognition using Adaptive Unsupervised Incremental Learning Technique", International Journal of Computer Applications, Vol. 7, No. 4, 2010, pp. 27-33.
[11] K. Mori, M. Matsugu, and T. Suzuki, "Face recognition using SVM fed with intermediate output of CNN for face detection", in Proc. of the IAPR Conf. on Machine Vision Applications (MVA), 2005, pp. 410–413.
[12] M. Szarvas et al., "Pedestrian detection with convolutional neural networks", in Proc. of the IEEE Symp. On Intell. Vehicles (IV), 2005, pp. 224–229.
[13] L.M. Lorigo, and V. Govindaraju, "Offline Arabic handwriting Recognition: A survey", IEEE Trans. Pattern Anal. Machine Intell., Vol. 28, 2006, pp.712-724.
[14] A.M. AL-Shatnawi, S. AL-Salaimeh, F. H. AL-Zawaideh, and O. Khairuddin, "Offline Arabic Text Recognition-An

Overview", World of Computer Science and Information Technology Journal (WCSIT), Vol. 1, No. 5, 2011, pp. 184-192.
[15] A. Mesleh, A. Sharadqh, A. Al-Azzeh, J. Abu-Zaher, M. Al-Zabin, N. Jaber,and T. Odeh, Hasn, "An optical character recognition", Contemporary Engineering Sciences, Vol. 5, No. 11, 2012, pp.521-529.
[16] Y. Y. Zhang, P. S. P. Wang, "A New Parallel Thinning Methodology", International Journal of Pattern Recognition and Artificial Intelligence, Vol. 8, 1994, pp 999-1011.

AUTHENTICATION SYSTEM OF BANKING TRANSACTION BY FINGERPRINTING

Neural Network Approach

Eugène MBUYI MUKENDI[1], Jean Didier BATUBENGA MWAMBA[2]

[1] **Professor University of Kinshasa, Computer Science Department, DR Congo**

[2] **University of Kinshasa, Computer Science Department, DR Congo**

Kinshasa Computer Science Laboratory.

Abstract

Shape recognition is a set of techniques and methods aiming at identifying patterns from raw data so as to make a decision depending on the category assigned to these patterns. Examples of content to which methods are applied are numerous. It can be visual content such as bar code, face, fingerprint, speech, images and much more. In fingerprint recognition, the application makes the registration of a person as well as his fingerprint in a database for future authentication, that is to say determine whether two fingerprints are identical to conclude that they come from the same person. In our case, the application authenticates the client of the Bank and then asks for his/her account number and password to ca rry out the withdrawal. The algorithm for fingerprint recognition which is used has been proposed by **D. Maio** and allows the location of minutiae in a more direct way by using neural networks. [2], [4]

Key words: authentication, fingerprinting, neural networks, identification.

I. INTRODUCTION [1], [2], [3], [4]

I.1. Pattern recognition

A. Kind of pattern recognition

In general, there are two types of pattern recognition:

- **Identification**: determines whether an observation or input data is a manifestation of the individual previously known.
- **Classification**: determines that the observation is a manifestation of a member of a class. In both cases, individuals or classes are characterized by a vector of properties. The recognition is summed up in a comparison of the properties.

B. Pattern recognition system

A pattern recognition system principally possesses an equipment of acquisition of the object to recognize. These equipments can be a video, a fingerprint reader, microphone etc. From them, the system acquires the input data that will undergo pretreatment to improve their qualities. Generally, this task is performed by hardware. After the preprocessing stage, the system goes to the extraction of features that will be used to identify the object to recognize. The next step will consist in the classification which aims, for a given input object, to specify which class it belongs to.

With the help of special equipment depending on the type of recognition to achieve, the acquisition of data to be processed can be manipulated by machine. The pattern to recognize is often a digitized image or a speech segment. It could undergo pretreatment so as to be ready for manipulation. The structure of a pattern recognition system contains the following items:

Pretreatments: have to determine hash information and retain the relevant one, useful for recognition. They can provide standardized data. They consist of noise reduction, the segmentation, normalization and skeletonization.

Parameterization: consists in the reduction of the complexity of data to be processed by some useful parameters stored in the database. The parameters will represent the actual data in an irrefutable way. They will be made thanks to characteristics extracted from data. In the parameterization, the system can calculate the type of representation, extract features and finally pass to learning.

The post-processing: is to validate the results of the classification using specific tools scope.

I.2. Phases of pattern recognition

Applications of pattern recognition are made through a pre-registration of the form to be recognized. It is only from this record that the application may proceed to the recognition of the so called form by comparing it with those previously recorded.

I.3. Algorithm for fingerprint recognition

Generally there are two categories of algorithms for fingerprint recognition: the first category concerns "conventional" algorithms that rely on the relative position between minutiae while the second includes algorithms to extract other features of the fingerprint, such as the local direction of the grooves, or

the local frequencial components of the texture in the heart of the image.
Note also the alternative proposed by **D. Maio** that allows the location of minutiae in a more direct way by using neural networks.

Figure 1. Fingerprints

A neural network is composed of a set of interconnected formal neurons giving rise to various network structures. For our application, we used a structure in successive layers (Multi-Layer Perceptron: MLP). Such a structure (See Figure 3) broadcasts the information of the input layer, composed by the artificial neural receiving primitive information, to the output layer, which contains the final neurons transmitting output information processed by the all network while traversing one or more intermediate layers, said hidden layers. So established, the network is a non-linear system that combines the input feature vectors to the outline of the image of the output layer.

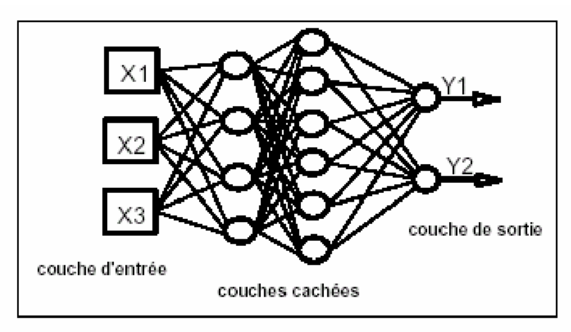


Figure 2. Multilayer neural network (also called Multi-Layer Perceptron MLP).

The neural network developed uses a threshold sigmoidal function and a gradient retropropagation to optimize its learning phase. Retro propagation consists in retropropagating error committed by a neuron's synapses and neurons that are connected. For neural networks, gradient error retropropagation are commonly used to correct errors in the importance of the elements that have just participated in the realization of these errors: the synaptic weights that contribute to generate an important error will be more significantly modified than weights which have led to a marginal error.

II. MODELLING AND IMPLEMENTATION OF APPLICATION [4], [5] [6]

II.1. Description of Application

The application that we are going to design will be responsible for authenticating a client prior to a transaction, especially in case of withdrawal. Being a biometric system, the application makes the customer's record as well as his/her fingerprint for future authentication. Thus customers register their fingerprints as well as useful information.

During withdrawal, the application authenticates the client then asks his/her account number and password to complete the withdrawal.

II.2. Modeling the Application

II.2.1. Use case diagrams

II.2.1.1. Determination of system actors

Actors represent different entities that will interact with the system.

The different actors are:

- Fingerprint readers: the role is to capture customers' fingerprints
- Customer: Who can make a withdrawal or deposit
- The user creates the account of the customer, records the customer in the enrollment phase and receives deposits of money from different client. She/he administers the program

II.2.1.2. Description of use cases

In the table below we describe the different use cases used in our application.

II.2.1.3. Use case diagrams

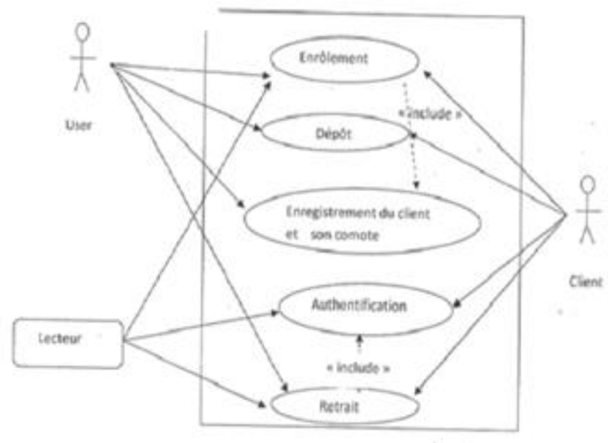


Figure 3. Use case diagrams

II.3. Data base design

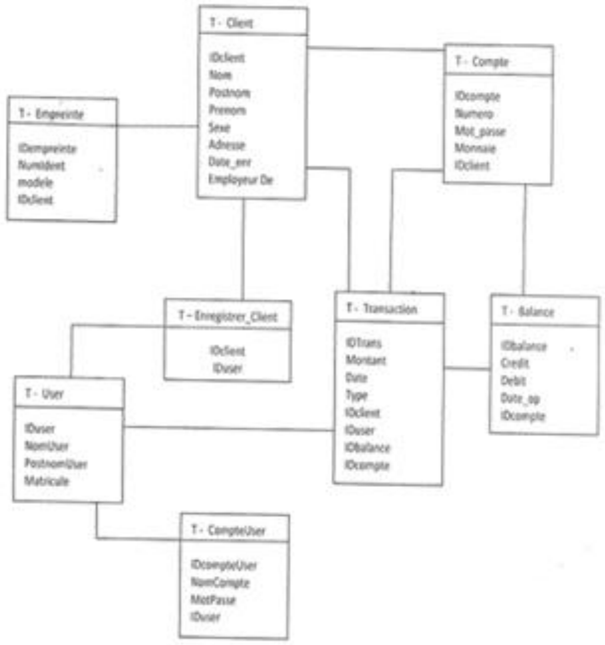


Figure 4. Passage from class diagram to MLDR

II.4. Interface presentation for the Application

II.5. Extract source code of the application

```
import java.awt.Dimension;
import java.awt.Toolkit;
import javax.swing.JOptionPane;
public class Appi extends javax.swing.JFrame {
  public Appi(String user) {
    initComponents();
     Dimension dim=Toolkit.getDefaultToolkit().getScreenSize();
    this.setLocation(dim.width/2-this.getWidth()/2,dim.height/2-this.getHeight()/2);
    txtCompte.setText(user);
  }
private void creerDossierClientActionPerformed(java.awt.event.ActionEvent evt) {
    Client monClient = new Client();
    panClient.setRightComponent(monClient);
  }
  private void jButton1ActionPerformed(java.awt.event.ActionEvent evt) {
    User user = new User();
    panConfig.setRightComponent(user);
  }
  private void bEnrollerActionPerformed(java.awt.event.ActionEvent evt) {
    FormMain classeEnroll= new FormMain();
     classeEnroll.setVisible(true);
  }
  private void bRetraitActionPerformed(java.awt.event.ActionEvent evt) {
   TransactionRetrait retrait = new TransactionRetrait();
    panTransact.setRightComponent(retrait);
     FormMain classeEnroll= new FormMain();
          classeEnroll.setVisible(true);

  }
  private void depotActionPerformed(java.awt.event.ActionEvent evt) {
    // depot
    TransactionDepot depot = new TransactionDepot();
    panTransact.setRightComponent(depot);
  }
  private void jLabel3MouseClicked(java.awt.event.MouseEvent evt) {
    Login login;
    int rep = JOptionPane.showOptionDialog(panClient,"Voulez vous réellement quitté", "Fermeture
du Programme",
JOptionPane.OK_CANCEL_OPTION,JOptionPane.INFORMATION_MESSAGE,null, null,null);
    if(rep == JOptionPane.OK_OPTION)
      login = new Login();
  }
```

```
public static void main(String args[]) {
  final String compte = null;
  java.awt.EventQueue.invokeLater(new Runnable() {
    public void run() {
      new Appi(compte).setVisible(true);
    }
  });
```

Code pour le dépôt

```
package contenu;
import java.io.FileInputStream;
import java.sql.ResultSet;
import java.text.DateFormat;
import java.util.Date;
import java.util.HashMap;
import javax.swing.JOptionPane;
import net.sf.jasperreports.engine.JasperFillManager;
import net.sf.jasperreports.engine.JasperPrint;
import net.sf.jasperreports.view.JasperViewer;
public class TransactionDepot extends javax.swing.JPanel {
   Connecion connexion;
   double soldeAncien;
   int id,id_client,id_bal;
  public TransactionDepot() {
    initComponents();
     connexion = new Connecion();
  }
private void jButton1ActionPerformed(java.awt.event.ActionEvent evt) {
     String nom = tnom.getText();
     String numero = txtNumeroCompte.getText();
      id = 0; double credit = 0,debit = 0;
      try{
       java.sql.Statement stmt=connexion.connect.createStatement();
       String requete="SELECT
`Nom`,`PostNom`,`Prenom`,`Adresse`,`Numero`,`Monnaie`,IDcompte,client.IDclient " +
            "FROM `client`,compte " +
            "WHERE `Numero`='"+numero+"' and client.Nom='"+nom+"'";
       ResultSet res=stmt.executeQuery(requete);
       while(res.next()){
         txtNom.setText(res.getString(1));
         txtPostnom.setText(res.getString(2));
         txtPrenom.setText(res.getString(3));
         tAdresse.setText(res.getString(4));
           jNumeroCompte.setText(numero);
           jMonnaie.setText(res.getString(6));
           id = res.getInt(7);
           id_client = res.getInt(8);
         }
         String reque="SELECT Credit,Debit,IDbalance FROM balance WHERE
balance.IDcompte='"+id+"'";
         ResultSet re=stmt.executeQuery(reque);
         while(re.next()){
          debit = re.getDouble(2);
          credit = re.getDouble(1);
          id_bal = re.getInt(3);
         }
       soldeAncien = credit - debit;
       jSoldeAncien.setText(""+soldeAncien);
      }catch(Exception ex){
          JOptionPane.showMessageDialog(null,"Erreur"+ex);
      }
    }
    private void jButton2ActionPerformed(java.awt.event.ActionEvent evt) {
      // Crediter :le compte
      Date datejr = new Date();
      String date = DateFormat.getDateInstance().format(datejr);
      date =date+" a "+DateFormat.getTimeInstance().format(datejr)  ;
       double soldeActuel;
       soldeActuel = soldeAncien + Double.parseDouble(txtMontant.getText());
       txtActuel.setText(""+soldeActuel);
       //processus de creditation du solde
       try{
    java.sql.Statement stm=connexion.connect.createStatement();
    String requete="insert into trasanction "
                  +
"values(NULL,'"+Double.parseDouble(txtMontant.getText())+"','"+date+"','Depot','"+id+"'," +
           "'herve','"+id_client+"','"+id_bal+"')";
     int re=stm.executeUpdate(requete);
     String reque= "UPDATE balance SET Credit = Credit +
"+Double.parseDouble(txtMontant.getText())+" WHERE(IDcompte='"+id+"')";
     int r=stm.executeUpdate(reque);
     String reqet="SELECT MAX(IDtrans) FROM trasanction";
        ResultSet reponse =stm.executeQuery(reqet);
        reponse.next();
         JOptionPane.showMessageDialog(null,"Compte Crédité ! ");
          //impression du document
        String cheminFichier="C:\\Rapport\\InfoTransaction.jasper";
        FileInputStream cheminReport=null;
        cheminReport=new FileInputStream(cheminFichier);
```

```
        HashMap param=new HashMap();
        param.put("id_client",id_client);
        param.put("numero",jNumeroCompte.getText());
        param.put("id_trans",reponse.getInt(1));

        JasperPrint jprint=JasperFillManager.fillReport(cheminFichier, param,connexion.connect);
        JasperViewer.viewReport(jprint,false);
      }catch(Exception ex){
    JOptionPane.showMessageDialog(null,"Erreur"+ex);
    }
  }
  private void txtMontantFocusLost(java.awt.event.FocusEvent evt) {
    // calcul de la solde actuel
  }
```

Code du le retrait

```
package contenu;
import java.io.FileInputStream;
import java.sql.ResultSet;
import java.text.DateFormat;
import java.util.Date;
import java.util.HashMap;
import javax.swing.JOptionPane;
import net.sf.jasperreports.engine.JasperFillManager;
import net.sf.jasperreports.engine.JasperPrint;
import net.sf.jasperreports.view.JasperViewer;
public class TransactionRetrait extends javax.swing.JPanel {
  Connecion connexion;
  double soldeAncien;
  int id,id_client,id_bal;
  public TransactionRetrait() {
    connexion = new Connecion();
    initComponents();
  }
private void jButton1ActionPerformed(java.awt.event.ActionEvent evt) {
    // TODO add your handling code here:
    String motPasse= tnom.getText();
    String numero = txtNumeroCompte.getText();
    String numCompte="";
    id = 0; double credit = 0,debit = 0;

    try{
      java.sql.Statement stmt=connexion.connect.createStatement();
      String requete="SELECT
`Nom`,`PostNom`,`Prenom`,`Adresse`,`Numero`,`Monnaie`,IDcompte,client.IDclient " +
          "FROM `client`,compte " +
          "WHERE `Numero`='"+numero+"' and compte.MotPasse='"+motPasse+"' and
client.Idclient = compte.IDclient ";
      ResultSet res=stmt.executeQuery(requete);
      while(res.next()){
        txtNom.setText(res.getString(1));
        txtPostnom.setText(res.getString(2));
        txtPrenom.setText(res.getString(3));
        tAdresse.setText(res.getString(4));
        jNumeroCompte.setText(res.getString(5));
        jMonnaie.setText(res.getString(6));
        id = res.getInt(7);
        id_client = res.getInt(8);
        numCompte = res.getString(5);
      }
      String reque="SELECT Credit,Debit,IDbalance FROM balance WHERE
balance.IDcompte='"+id+"'";
      ResultSet re=stmt.executeQuery(reque);
      while(re.next()){
        debit = re.getDouble(2);
        credit = re.getDouble(1);
        id_bal = re.getInt(3);
      }
      soldeAncien = credit - debit;
      jSoldeAncien.setText(""+soldeAncien);

    }catch(Exception ex){
      JOptionPane.showMessageDialog(null,"Numero compte n'est pas valide : "+ex);
    }
}
  private void jButton2ActionPerformed(java.awt.event.ActionEvent evt) {
    // Crediter :le compte
    Date datejr = new Date();
    String date = DateFormat.getDateInstance().format(datejr);
    date =date+" a "+DateFormat.getTimeInstance().format(datejr) ;
    double soldeActuel;
    soldeActuel = soldeAncien - Double.parseDouble(txtMontant.getText());
    txtActuel.setText(""+soldeActuel);
    //processus de creditation du solde
    try{
      java.sql.Statement stm=connexion.connect.createStatement();
      String requete="insert into trasanction "
        +
"values(NULL,'"+Double.parseDouble(txtMontant.getText())+"','"+date+"','Retrait','"+id+"'," +
```

```
            "'herve','"+id_client+"','"+id_bal+"')";
        int re=stm.executeUpdate(requete);

        String reque= "UPDATE balance SET Debit = Debit +
"+Double.parseDouble(txtMontant.getText())+", Credit = Credit -
"+Double.parseDouble(txtMontant.getText())+" WHERE(IDcompte='"+id+"')";
        int r=stm.executeUpdate(reque);
        String reqet="SELECT MAX(IDtrans) FROM trasanction";
          ResultSet reponse =stm.executeQuery(reqet);
          reponse.next();
        JOptionPane.showMessageDialog(null,"Compte Débité ! ");

        //impression du document
        //impression du document
          String cheminFichier="C:\\Rapport\\InfoTransaction.jasper";
          FileInputStream cheminReport=null;
          cheminReport=new FileInputStream(cheminFichier);
          HashMap param=new HashMap();
          param.put("id_client",id_client);
          param.put("numero",jNumeroCompte.getText());
          param.put("id_trans",reponse.getInt(1));

        JasperPrint jprint=JasperFillManager.fillReport(cheminFichier, param,connexion.connect);
        JasperViewer.viewReport(jprint,false);
      }catch(Exception ex){
        JOptionPane.showMessageDialog(null,"Erreur"+ex);
      }
}
```

Extrait code d'enrollement

```
/
  public void enroll(int IDclient) {
    try {
      //Inserts the template on the database
      enrollStmt.setBinaryStream(1,new ByteArrayInputStream(template.getData()),
template.getData().length);
      // enrollStmt.setInt(2, template.getQuality());
      enrollStmt.setInt(2, IDclient);
      enrollStmt.executeUpdate();

      //Picks the ID generated for it.
      ResultSet rs = insertedIdStmt.executeQuery();
      rs.next();
```

CONCLUSION

We presented a biometric authentication system that identifies customers by means of fingerprints before carrying out banking transactions while withdrawing money. A complete processing chain from the acquisition of fingerprints until the withdrawal has been developed with satisfaction. At the moment of enrollment, neural networks do learn the different fingerprints. During recognition, they combine a fingerprint stored in the database at the time of the learning process with the input in order to state the correspondence.

By adding authentication, notably by fingerprint during transactions we increase safety and minimize the risk of fraud and theft. The false acceptance rates and those of discharges of a potential client were minimized.

REFERENCES

1. BELGUECHI R: Contribution to the fingerprint recognition by a hybrid approach INI, 2006.
2. CHAARI Anis: New approach to identification in biometric databases based on an unsupervised classification. PhD thesis, University of Enry, 2009.
3. DREYFUS G and others: Networks of neuron methodology and applications. Eyrolle 2004
4. MBUYI MUKENDI Eugene and others: Datamining and Neuronal networks I IJCSI, Volume 9, issue 5, september 2012.
5. MBUYI MUKENDI Eugène and others : Biometric system by pointing finger IJCSI volume 10, issue 4, juillet 2013.
6. MUNTENGE Kalili Herve: Automatic authentication system, UNIKIN 2010.

Customer Churn Prediction in Telecommunication

A Decade Review and Classification

Nabgha Hashmi[1], Naveed Anwer Butt[2] and Dr.Muddesar Iqbal[3]

1 Department of Computer Science and Information Technology

University of Gujrat, Pakistan

2 Faculty of Computer Science and Information Technology

University of Gujrat, Pakistan

3 Faculty of Computer Science and Information Technology

University of Gujrat, Pakistan

Abstract

Acquisition and the retention of customers are the top most concerns in today's business world. The rapid increase of market in every business is leading to higher subscriber base. Consequently, companies have realized the importance of retaining the on hand customers. It has become mandatory for the service providers to reduce churn rate because the negligence could be resulted as profitability reduction in major perspective. Churn prediction helps in identifying those customers who are likely to leave a company. Telecommunication is coping with the issue of ever increasing churn rate. Data mining techniques enable these telecommunication companies to be equipped with effective methods for reducing churn rate. The paper reviews 61 journal articles to survey the pros and cons of renowned data mining techniques used to build predictive customer churn models in the field of telecommunication and thus providing a roadmap to researchers for knowledge accumulation about data mining techniques in telecom.

Keywords: *Retention, Higher Subscriber Base, Customer Churn, Telecommunication, Data mining.*

1. Introduction

Customer churn is referred to as the inclination of a customer to leave a service provider. (Chitra Phadke, Huseyin Uzunalioglu et al., 2013; Vivek Bhambri, 2013; Zhen-Yu Chen et al., 2012; Clement Kirui, Li Hong et al., 2013; Yaya Xie a, Xiu Li et al., 2009; Chandar, Laha and Krishna, 2006). Customer churn prediction is the process of identifying those customers who could leave or switch from the current service provider company due to certain reasons (Coussement and Van den Poel, 2008; Buckinx and Van den Poel, 2005). The major aim of churn prediction model is to identify such customers so that the retention strategies could be targeted upon them and the company may flourish by maximizing its overall revenue (Junxiang Lu, 2003).

Customer churn prediction has been raised as a notorious issue in many fields such as telecommunication (Umman, Tuğba, Şimşek, & Gürsoy,2010; Verbeke et al., 2011; Tarik Rashid ,2008; Adnan Idris et al.,2012; Bingquan Huang et al., 2012; Hung, S. Y.et al., 2006), Credit Card (Guangli Nie, Wei Rowe et al., 2011), Internet Service Providers (Afaq Alam Khan et al., 2010; B.Q. Huang et al., 2009; Li, S. T., Shue, L. Y., 2006), Electronic Commerce (Chang, S. E., Changchien et al., 2006; Changchien, S. W., Lee, C. F., & Hsu, Y. J., 2004; Etzion, O., Fisher et al., 2005; Kuo, R. J. et al., 2005, Kim et al., 2004), Retail Marketing (Chen, M. C., Chiu, 2005), Newspaper publishing companies (Dries F. Benoit b et al., 2010;Douglas, S., Agarwal, D., and Alonso, T., 2005), Banking (Yaya Xie a, Xiu Li, E.W.T. Ngai b, Weiyun Ying, 2009;Koh, H. C., & Chan, K. L. G., 2002; Au, W. H., & Chan, K. C. C. , 2003; Chiang, D. A., 2002) and financial services (B Larivière, D Van den Poel, 2004).

But among all other fields, telecommunication companies over the years are experiencing the highest annual churn rate from 20% to 40% (Jae-Hyeon Ahna,Sang-Pil Hana and Yung-Seop Leeb, 2006; Kim, Park, & Jeong, 2004; Berson, Smith, and Therling, 1999; Madden, Savage, & Coble-Neal, 1999). This has financial implications on a company, as it costs 5 to 10 times more to add a new customer than retaining an old customer with the company (Junxiang Lu, 2003).

Data mining, as a major field of computer sciences, is defined as the process of extracting hidden patterns from very large datasets by using statistical, mathematical, artificial intelligence and machine learning techniques (Turban, Aronson, Liang, and Sharda, 2007; Berson et al., 2000; Lejeune ,2001; Ahmed, 2004 and Berry and Linoff ,2004; Lau, Wong, Hui, & Pun, 2003).

The organizations may have an ocean of data but still they are starving for information or more specifically for valuable knowledge. Data mining tools could best help these organizations to extract hidden patterns of useful information (Vivek Bhambri, 2013; Berson et al., 2000).

This paper aims at reviewing the research intensity during the year of 2002 to June, 2013. It subsequently discusses the churn prediction problem specifically in telecom as section 2. Then the research methodology would be discussed along with flow chart view in section 3. Afterwards, the heart of paper, the comprehensive analysis in four different dimensions would be presented as section 4 along with the evaluation and interpretation of results. Also the challenges faced by the researchers in this sector would be highlighted in section 5. Limitations as well as findings of the paper will be described in section 6 and 7. Section 8 concludes the paper.

2. Customer Churn in Telecommunication

Telecommunication has gained one of the top positions in the list of fastest growing industries of the world by covering 90% of its population (Rajanish Dass and Rumit Jain; 2011). It is one of the sectors where customer base plays a very important role in maintaining the revenue (Adnan Idris , Muhammad Rizwana, Asifullah Khan, 2012). The telecommunication sector is facing a severe threat of customer churn (Jae-Hyeon Ahna,Sang-Pil Hana and Yung-Seop Leeb, 2006; Kim, Park, & Jeong, 2004; Berson, Smith, and Therling, 1999; Madden, Savage, & Coble-Neal, 1999).

According to Wie and Chiu, 2002, wireless telecom industry is facing with the threat of loosing 27% of its customers every year, which would definitely result in huge revenue loss. Its is also an adopted fact that adding or acquiring a new customer costs 5 to 10 times more to add a new customer than retaining an old customer with the company (Junxiang Lu, 2003). Therefore, Richeldi and Perrucci, 2002, suggest that the company should pay more attention to retain its current subscribers rather than adding new ones.

Now a days business firms pay more attention to make firm relationship with their customers (Zhen-Yu Chen et al., 2012; Coussement and Van den Poel, 2008; Ngai et al., 2009; Kim & Yoon, 2004).

Hence it has become a belief that the best marketing strategy is to retain the existing subscribers or more simply to avoid customer churn (Golshan Mohammadi et al., 2013; Chih-Fong Tsai et al., 2009; Kim, Park, and Jeong, 2004; Kim and Yoon, 2004).

To tackle with this problem, data mining techniques have been proved as the best tools to fight against ever increasing customer churn rate (Au, Chan, and Yao, 2003; Bin, Peiji, and Juan, 2007; Coussement and den Poe, 2008; Hung, Yen, and Wang, 2006; John, Ashutosh, Rajkumar, and Dymitr, 2007; Lazarov and Capota, 2007; Wei and Chiu, 2002).

3. Research Methodology

Because journals are considered as the most reliable source of research (Nord & Nord, 1995), so firstly, some renowned online journal databases were explored to get a comprehensive academic literature on the topic. Here is a list:

- Elsevier
- IEEE Xplore
- SpringerLink
- ScienceDirect
- ACM Digital Library
- Microsoft Academic Search

The search criteria based on the search strings by following the syntax of each of the database. It originally produced 834 articles. As a next step, the articles actually related to the topic were filtered out. The criteria for the inclusion and exclusion of articles are as follow:

- The articles containing the keywords written in search string.
- Only those articles that have been published in significant journals. Conference papers, newsletter, lecture notes, books, doctoral dissertations, un-published work and conference proceedings were excluded.
- The search criteria were restricted to from the year 2002 to 2013.
- Only those articles were selected that were related to the telecommunication world.

- Then these selected articles were distributed into four different classes:

  — *Classification by Techniques being used for predicting customer churn.*
  — *Classification by Journals in which the papers were published.*
  — *Classification by the type of data set being used in research; wireless or land line.*
  — *Classification by Year in which the articles were published.*

Fig.1 depicts the overall scenario of research process described earlier in more comprehensive way.

4. Distribution of Articles

A detailed distribution of 61 journal articles has been shown in Table 1 along with their title, publication year and author names.

Total 61 articles were finally selected for analysis in four different dimensions:

- Distribution of articles by Dataset Type
- Distribution of articles by Techniques
- Distribution of articles by Journals
- Distribution of articles by Publication Year

4.1 Distribution of articles by Dataset Type

No one can deny the importance of data in an organization. It is truly called as an "asset". But it becomes to its intense significance level when we talk about predictive modeling.

The fact is that the quality of our predictive model is strictly dependent upon the quality of data being used for it. Un-reliable data will definitely lead to incorrect results (Jiayin Qi ,Yingying Zhang, Huaying Shu , Yuanquan Li, Lei Ge, 2006). In predictive modeling, data set plays the vital role.

Telecommunication sector could be divided into two sub categories in accordance with service providing nature. These two are Fixed line/ Land line Telecommunication and Cellular /Mobile /Wireless Telecommunication Service providers.

The classification of 61 articles in the above mentioned perspective has been shown in Fig.2. The figure is depicting the fact that among three categories, wireless telecom has grabbed the major part of research.

Table 1: Classification by Dataset Type

Data Type	Frequency	Percentage
Land Line	3	4.918032787
Wireless	35	57.37704918
Dataset not specified	23	37.70491803
Total	**61**	**100**

Table 1. is showing the statistics that 4.92% (3 among 61 articles) research has been done on taking fixed line telecommunication data type. And 57.38% (35 among 61) papers worked upon wireless or mobile telephony churn prediction.

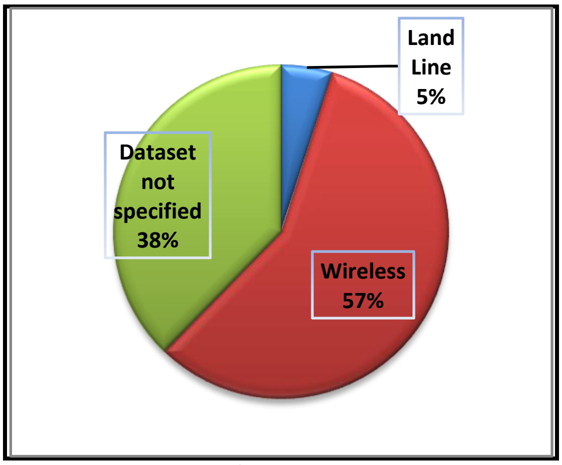


Fig. 1 : Classification by Dataset Type

The major reason behind the very little research work in fixed line area is the less amount of qualified information (Jiayin Qi ,Yingying Zhang, Huaying Shu , Yuanquan Li, Lei Ge, 2006). Customer call details, demographic, complaints, billing information and contractual data is usually used for predictive model building in telecommunication (Chih-Ping Wei, I-Tang Chiu, 2002). However, in fixed line service providers, researcher have access to only customers' call details and billing information (Jiayin Qi ,Yingying Zhang, Huaying Shu , Yuanquan Li, Lei Ge, 2006). This is the reason that invokes a challenge for researchers to build a predictive model with limited information for fixed line churners (Jiayin Qi ,Yingying Zhang, Huaying Shu , Yuanquan Li, Lei Ge, 2006).

Fig.3 depicts another split view of frequency of journal articles in terms of wireless and fixed line telephony which ultimately strengthens above mentioned facts.

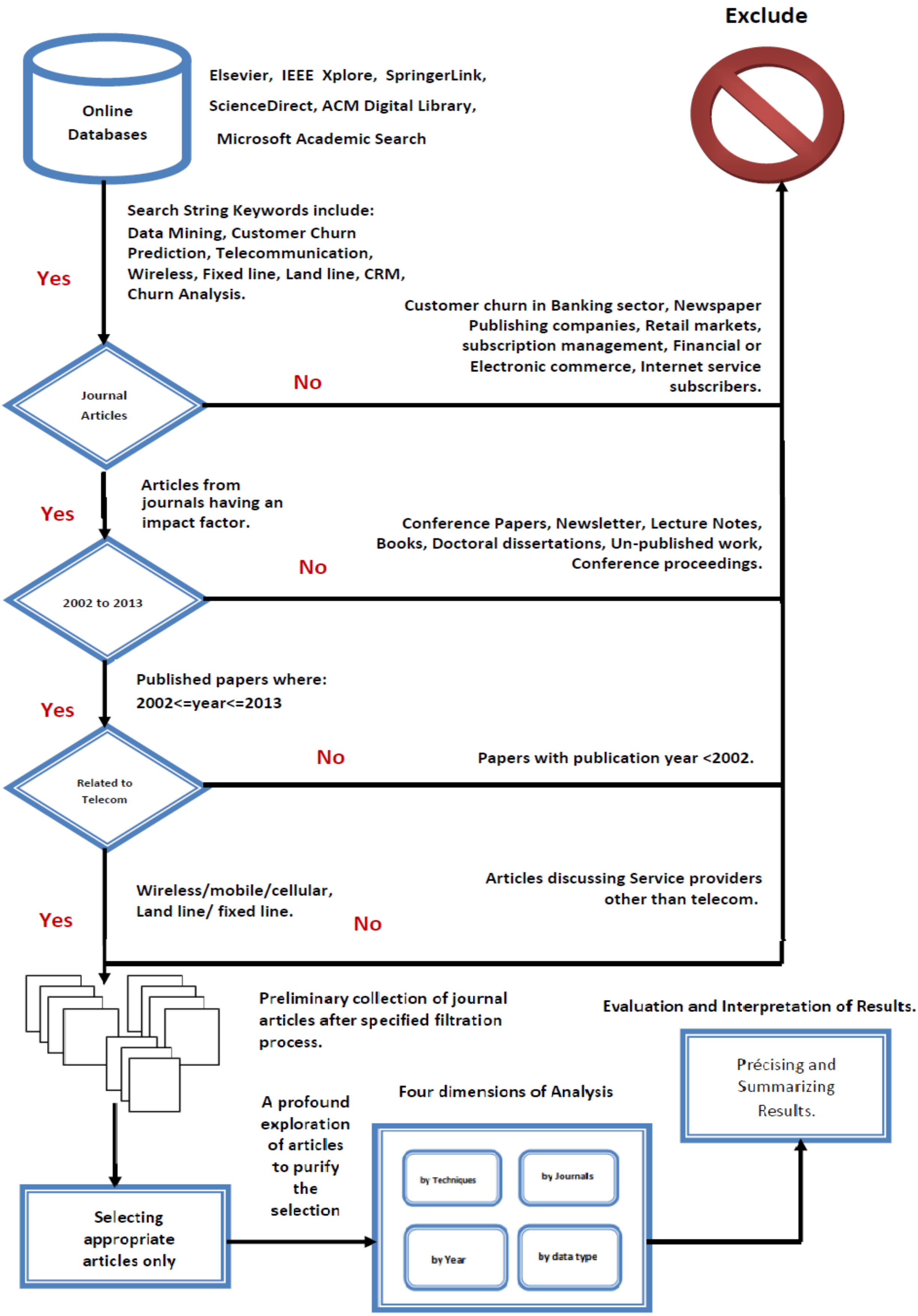


Fig. 2 : Research Methodology

Table 2: Classification by Techniques

Technique	**Frequency**
Decision Tree	26
Neural Network	16
Logistic Regression	15
Cluster Analysis	10
Genetic Algorithm	6
Markov Model	4
Naïve Bayes	4
k-nearest-neighbor	3
Bayesian Belief Network	3
Association Rule	2
Support Vector Machine	2
Bagging	2
CART	2
CHAID	2
K-Means	1
Fuzzy C means	1
influence diffusion model	1
Chr-PmRF	1
partial least squares (PLS)	1
C5.0	1
Structural Equation Model	1
Total	**104**

4.2 Distribution of articles by Techniques

Table 3 shows the distribution of journal articles by the technique being used for model building. Total 104 types of techniques have been used in 61 articles. Here it is a point to be noted that one article may have used more than single technique for model building.

Among these techniques, Decision Tree has been used most commonly. It has been used in 25% of articles. Neural network and Logistic Regression have been used with 15.4 and 14.4 percentage. Following are the top three popular techniques have been reviewed in the light of academic literature:

Decision Trees: Decision tree is used to predict future trends and to extract models based upon the interrelated decisions (Chu , Tsai and Ho, 2007; Berry & Linoff, 2004; Chen et al., 2003; Kim, Song, Kim, & Kim, 2005). It works upon the principal of classifying data into specific classes in accordance with their properties. Internal nodes follow the root node by covering all occurrence possibilities (Buckinx, Moons, Poel, & Wets, 2004; Chen et al., 2003 ;). Thus a tree is formed with its unique arc describing particular responses.

Neural Network: Currently, neural networks are used by researchers in the field of classification, clustering and prediction (Berry & Linoff, 2004; Turban et al., 2007). When applying neural network technique, it converts the data into dimensional array of neurons in an orderly manner which ultimately forms a prediction hierarch (Tsai and Lu, 2009; Song, Kim, Cho, & Kim, 2004).

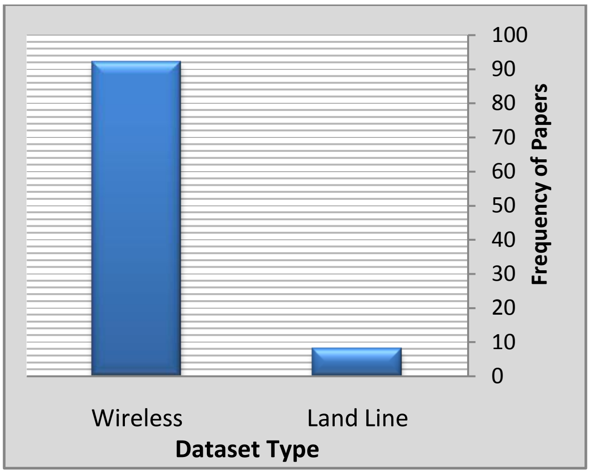


Fig. 3 : Classification by Two Dataset Type

Regression: Regression is also very popular technique for predicting behavior. It determines the impact of many independent variables to predict the possible reliance of one or more than one dependent variables (Bingquan Huang et al., 2012; Michel Ballings, Dirk Van den Poel, 2012; Marcin Owczarczuk, 2010; Jiayin Qi et al., 2006).

Cluster Analysis: Mainly, cluster analysis is done under the field of machine learning where objects with similar nature are kept in one cluster. It is also used for reliable statistical analysis of data (J. Hadden, A. Tiwari, R. Roy, D. Ruta, 2007; Larivi _ere and VandenPoel, 2005).

4.3 Distribution of articles by year of publication

The distribution of articles by the year of publication is shown in Fig.5. the decade along with the year of 2013 till June has been shown. It could be observed that 2012 and 2010 are the years where telecommunication sector got most research material on churn analysis.

Table 3: Classification by Publication Year

Year of Publication	Frequency of Papers
2002	1
2003	3
2004	2
2005	3
2006	7
2007	6
2008	4
2009	3
2010	10
2011	7
2012	11
2013 (June)	4
Total	**61**

4.4. Distribution of articles by Journals

Table 4 shows the distribution of articles by journal. Total 32 journals were explored to find articles related to churn prediction in telecommunication sector. "Expert System with Applications" leads the race with 16 significant articles. "Telecommunication Policy" is following the race with 6 papers published in context of customer churn prediction.

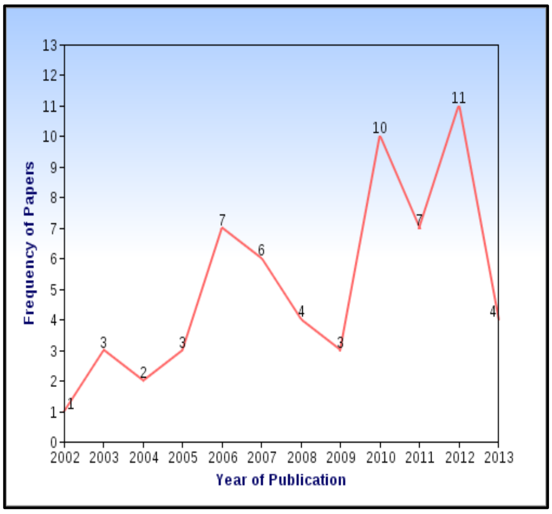


Fig.5: Classification by the Year

Table4: Classification by Journals

Journals	Frequency
Expert System with Applications	16
Telecommunications Policy	6
Decision Support Systems	6
Data Mining and Knowledge Discovery	2
Advanced Data Mining and Applications	2
International Journal of Computer Applications	2
International Journal of Intelligent Technology	2
IEEE Transactions on Knowledge and Data Engineering	1
Annals of Operations Research	1
Bell Labs Technical Journal	1
Computers and Electrical Engineering	1
Computers and Operations Research	1
Emerging Research in Artificial Intelligence and Computational Intelligence Communications in Computer and Information Science	1
European Journal of Operational Research	1
GE-International Journal of Management Research	1
IEEE Transactions on Evolutionary Computation	1
International Journal of Computer Science Issues	1
International Journal of Advanced Computer Science and Applications,	1
International Journal of Biometrics and Bioinformatics (IJBB)	1
International Journal of Production Research	1
Journal of Intelligent Manufacturing	1
Journal of Interactive Marketing	1
Journal of Organizational Computing and Electronic Commerce	1
Journal of Strategic Marketing	1
Journal of the Operational Research Society	1
Journal of the School of Business Administration	1
Knowledge-Based Systems,	1
Machine Learning and Data Mining	1
Marketing Intelligence and Planning	1
The Service Industries Journal	1
Applied Economics Letters	1
Systems Engineering — Theory & Practice	1
Total 32 Journals	**61 Papers**

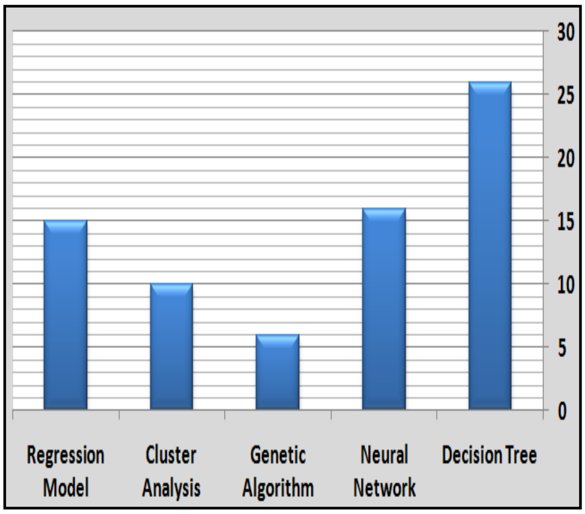


Fig. 4: Classification by mostly used techniques

5. Challenges

Researchers may face many problems in the process of modeling a churn prediction model. The major one is the missing or incomplete datasets for both wireless and fixed line telecommunication (B.Q.Huang a, T.-M. Kechadi , B. Buckley , G. Kiernan , E. Keogh , T. Rashid , 2010). On the other side, some telecom companies provide a very large enough data set which is sometimes very hard to handle (Adnan Idris, Muhammad Rizwan , Asifullah Khan, 2012) regarding with the problems of noisy data.

In the opinion of Yaya Xie , Xiu Li , E.W.T. Ngai , Weiyun Ying , 2009 and B.Q.Huang a, T.-M. Kechadi , B. Buckley , G. Kiernan , E. Keogh , T. Rashid , 2010, one of the biggest challenges for the researchers of telecom churn prediction is the imbalanced nature of data (Zhao et al., 2005).

The imbalanced term refers to un-equal ratio of regular customers with churners. Zhao et al., 2005 declare the fact that mostly, the number of churners' data is only of 2% of the total data which would certainly create problem in context with the reliability of the predictive model. Because of the confidential nature of telecom dataset, they are not publically available (Vivek Bhambri, 2013; B.Q. Huang et al., 2010). High dimensionality of these datasets is also another big issue (Adnan Idris et al., 2012).

While talking about the techniques mainly used for model building, XIA Guo-en and JINWei-dong, 2008, divide them into two major classes named as classification and artificial intelligence. The first class includes decision tree, logistic regression, naive bayesian classifiers and clustering are good for analyzing qualitative and continuous data and afterwards interpreting results but these techniques do not guarantee the appropriable accuracy of prediction model for large enough, highly dimensional, non linear or time series datasets (XIA Guo-en et al.,2008).

The second class including artificial neural network, self organization maps and evolutionary methods solve the problems of first class techniques with better prediction precision(XIA Guo-en et al.,2008). It means that exploring new methods or techniques for prediction model is very important.

6. Limitations of Research

As it has been mentioned earlier that only strong references were utilized to make the research more reliable in every perspective. But as nothing is perfect so this study has also some certain limitations:

- Firstly, only 61 articles have been review published from 2002 to 2013. More articles could be extracted by expanding the research duration.

- Secondly, the research was done under the search string including the main terms related to "customer churn", "churn prediction", "telecom""data mining" and "CRM" but the articles belonging to the churn prediction in telecom but not having any of these keywords could not be included in the review.

- Lastly, the research was restricted to only 6 online journals. Other academic journals may provide some more results.

7. Findings of Research

Although the research has certain limitations but yet it provided some implications:

- The majority of articles were related to wireless telephony. A very less amount of work (only 3 out of 61 articles) has been done in the field of fixed line telephony.
- The major reason behind the very little research work in fixed line area is the less or limited amount of qualified information.
- Decision tree has been emerged as the most commonly used technique for predicting churn rate.
- Un-availability or too large datasets with noisy nature are two major challenges in context of data quality constraints.
- High dimensionality and imbalanced nature of data are the barriers towards a précised prediction model.

- Although the classification techniques are good for analyzing qualitative and continuous data and afterwards interpreting results but these techniques do not guarantee the appropriable accuracy of prediction model for large enough, highly dimensional, non linear or time series datasets.
- Artificial intelligence and machine learning techniques work with high dimensional, bulky, non linear datasets with better prediction accuracy but complicated in terms of real world applications.

8. Conclusion

Now a days, telecommunication industry is struggling with a notorious issue of customer churn. The issue has serious allegation on customer loyalty along with revenue loss in major perspective. The only way is to retain the on hand customers by using customer churn prediction model. Data mining techniques help telecom industry in this perception by providing techniques to identify such customers so that retention actions could be targeted upon them.

The paper, in the beginning, highlighted the immense threat of customer churn for telecom companies by giving statistical reasons. Customer churn problem was discussed and reviewed both in general and subsequently in specific mode. Afterwards, a comprehensive comparison of selected articles was carried out in four intentional dimensions. 61 articles were reviewed in detail by year of publication, used technique, relative journal and by dataset type.

The findings of the research make a contribution in the field of customer churn predictive modeling in telecommunication. Thus the paper draws a sketch line for the researchers for reviewing and accumulation of the trends about data mining applications in the field of telecommunication.

Acknowledgments

A special thank goes to Dr. Javed Anjum Shiekh, Associate Director at University of Gujrat, Pakistan, for his valuable comments and sharing his worthy knowledge.

References

1. N.Kamalraj, and A.Malathi, "A Survey on Churn Prediction Techniques in Communication Sector", International Journal of Computer Applications, Volume 64, Isssue 5, 2013, pp. 39-42.
2. Yen-Hsien Lee, Chih-Ping Wei, Tsang-Hsiang Cheng, Ching-Ting Yang, "Nearest-neighbor-based approach to time-series classification", Decision Support Systems, Volume 53, Issue 1, 2013, pp. 207–217.
3. Chitra Phadke, Huseyin Uzunalioglu, Veena B. Mendiratta , Dan Kushnir , and Derek Doran, "Prediction of Subscriber Churn Using Social Network Analysis", Bell Labs Technical Journal, Volume 17, Issue 4, 2013, pages 63–75.
4. Vivek Bhambri, "Data Mining as a Tool to Predict Churn Behaviour of Customers", GE-International Journal of Management Research (IJMR), April 2013, pages 59-69.
5. N.Kamalraj and A.Malathi, "A Survey on Churn Prediction Techniques in Communication Sector", International Journal of Computer Applications, Volume 64– No.5, 2013, pp. 0975 – 8887.
6. Clement Kirui, Li Hong, Wilson Cheruiyot and Hillary Kirui, "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining", International Journal of Computer Science Issues (IJCSI), Vol. 10, Issue 2, 2013, No 1, 165-172.
7. Golshan Mohammadi, Reza Tavakkoli-Moghaddam, and Mehrdad Mohammadi, "Hierarchical Neural Regression Models for Customer Churn Prediction", Journal of Engineering, Volume 2013 , 2013, Article ID 543940, 9 pages.
8. Hossein Abbasimehr, Mostafa Setak and Javad Soroor, "A framework for identification of high-value customers by including social network based variables for churn prediction using neuro-fuzzy techniques", International Journal of Production Research, 51:4, 2013,1279-1294.
9. Zhen-Yu Chen , Zhi-Ping Fan , Minghe Sun, "A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data", European Journal of Operational Research, 223, 2012, 461–472.
10. Adnan Idris, Muhammad Rizwan , Asifullah Khan,"Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies", Journal of Computers and Electrical Engineering, 38 , 2012,1808–1819.
11. Tomasz S. Ząbkowski, Wiesław Szczesny, "Insolvency modeling in the cellular telecommunication industry", Expert Systems with Applications, 39, 2012, 6879–6886.
12. Saravanan M.1 and Vijay Raajaa G.S., "A Graph-Based Churn Prediction Model for Mobile Telecom Networks", ADMA, LNAI , 7713, 2012, pp. 367–382.
13. Shui Hua Han a, Shui Xiu Lu a, Stephen C.H. Leung., "Segmentation of telecom customers based on customer value by decision tree model", Expert Systems with Applications, 39, 2012, 3964–3973.
14. Guangqing Li, Xiuqin Deng, "Customer Churn Prediction of China Telecom Based on Cluster
15. Analysis and Decision Tree Algorithm", Emerging Research in Artificial Intelligence and Computational Intelligence, Communications in Computer and Information Science, 2012, pp 319-327.
16. P Srinuan, E Bohlin, G Madden, "The determinants of mobile subscriber retention in Sweden", Applied

Economics Letters, Volume 19, Issue 5, 2012, 453-457.
17. Wouter Verbeke, Karel Dejaeger, David Martens, Joon Hur, and Bart Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach", Expert Systems with Applications, Volume 218, Issue 1, 2012, Pages 211–229.
18. Koen W. De Bock, Dirk Van den Poel, "Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models", Expert Systems with Applications 39, 2012, 6816–6826.
19. Michel Ballings, Dirk Van den Poel, "Customer event history for churn prediction: How long is long enough?", Expert Systems with Applications, Volume 39, Issue 18, 2012, Pages 13517–13522.
20. Bingquan Huan, Mohand Tahar Kechadi, Brian Buckley, "Customer churn prediction in telecommunications", Expert Systems with Applications, 39, 2012, 1414–1425.
21. Wouter Verbeke, David Martens, Christophe Mues, and Bart Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques", Expert Systems with Applications, Volume 38, Issue 3, 2011, Pages 2354–2364
22. Guangli Nie, Wei Rowe, Lingling Zhang, Yingjie Tian, Yong Shi, "Credit card churn forecasting by logistic regression and decision tree", Expert Systems with Applications, 38, 2011, 15273–15285.
23. Pınar Kisioglu, Y. Ilker Topcu, "Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey", Expert Systems with Applications, 38, 2011, 7151–7157.
24. Adem Karahoca, Dilek Karahoca, "GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system",Expert Systems with Applications, 38 , 2011, 1814–1822.
25. Torsten Dierkes,Martin Bichler,Ramayya Krishnan, "Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks", Decision Support Systems, Volume 51, Issue 3, 2011, Pages 361–371.
26. H Lee, Y Lee, H Cho, K Im, YS Kim, "Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model", Decision Support Systems, Volume 52, Issue 1, 2011, Pages 207–216.
27. Anuj Sharma, Dr.Prabin Kumar Panigrahi, "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services", International Journal of Computer Applications, Volume 27– No.11, 2011, pp. 0975 – 8887.
28. Abbas Keramati, Seyed M.S.Ardabili, "Churn analysis for an Iranian mobile operator", Telecommunications Policy, 35 , 2011, pp. 344–356.
29. Dass, Rajanish and Jain, Rumit, "An Analysis on the factors causing telecom churn: First Findings", *AMCIS 2011 Proceedings - All Submissions.* 2011, Paper 2.
30. Kristof Coussement, Dries F. Benoit, Dirk Van den Poel, "Improved marketing decision making in a customer churn prediction context using generalized additive models", Expert Systems with Applications, Volume 37, Issue 3, 2010, Pages 2132–214.
31. Dirk Thorleuchter, Dirk Van den Poel, Anita Prinzie, "Mining ideas from textual information", Expert Systems with Applications, Volume 37, Issue 10, 2010, Pages 7182–7188.
32. Rahul J. Jadhav, Usharani T. Pawar, "Churn Prediction in Telecommunication Using Data Mining Technology", International Journal of Advanced Computer Science and Applications, 2010, Vol. 2, No.2.
33. Ali Tamaddoni Jahromi, Mohammad Mehdi Sepehri, Babak Teimourpour and Sarvenaz Choobdar, "Modeling customer churn in a non-contractual setting: the case of telecommunications service providers", Journal of Strategic Marketing , Volume 18, Issue 7, 2010, 587-598.
34. Dejan Radosavljevik, Peter van der Putten, Kim Kyllesbech Larsen, "The Impact of Experimental Setup in Prepaid Churn Prediction for Mobile Telecommunications: What to Predict, for Whom and Does the Customer Experience Matter?", Transactions on Machine Learning and Data Mining , Volume 3 - Number 2, 2010, Pages 80-99.
35. KKK Wong, "Fighting churn with rate plan right-sizing: a customer retention strategy for the wireless telecommunications industry", The Service Industries Journal ,Volume 30, Issue 13, 2010, 2261-2271.
36. Umman, Tuğba, Şimşek, and Gürsoy, "Customer Churn analysis in telecommunication sector", Journal of the School of Business Administration, Vol: 39, No: 1, 2010, 35-49.
37. B.Q. Huang , T.-M. Kechadi , B. Buckley , G. Kiernan , E. Keogh , T. Rashid, "A new feature set with new window techniques for customer churn prediction in land-line telecommunications", Expert Systems with Applications, 37, 2010, pp. 3657–3665.
38. Bingquan Huang, B. Buckley , T.-M. Kechadi, "Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications", Expert Systems with Applications, 37, 2010, pp.3638–3646.
39. Marcin Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts", Expert Systems with Applications, 37, 2010,pp. 4710–4712.
40. Y. Huang, B. Q. Huang, M. T. Kechadi, "A New Filter Feature Selection Approach for Customer Churn Prediction in Telecommunications", Proceedings of the IEEM, 2010, pp. 338-342.
41. Chih-Fong Tsai , Mao-Yuan Chen, "Variable selection by association rules for customer churn prediction of multimedia on demand", Expert System with Applications 37, 2010, pp. 2006–2015.
42. Afaq Alam Khan, Sanjay Jamwal, M.M.Sepehri, "Applying Data Mining to Customer Churn Prediction in an Internet Service Provider", International Journal of Computer Applications, vol. 9, issue 7, 2010, pp. 8-14
43. Yaya Xie, Xiu Li, E. W. T. Ngai, and Weiyun Ying, "Customer churn prediction using improved balanced random forests", Expert Systems with Applications, Volume 36 Issue 3, 2009, Pages 5445-5449.

44. B.Q. Huang, M-T. Kechadi, and B. Buckley, "Customer Churn Prediction for Broadband Internet Services", Data Warehousing and Knowledge Discovery, Lecture Notes in Computer Science Volume 5691, 2009, pp 229-243.
45. Jiayin Qi, Li Zhang, Yanping Liu, Ling Li, Yongpin Zhou, Yao Shen, Liang Liang, Huaizu Li, "AD Trees Logit model for customer churn prediction", Ann Oper Res, 168: 2009, pp. 247–265.
46. I Bose, X Chen, "Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn", Journal of Organizational Computing and Electronic Commerce, Volume 19, Issue 2, 2009, 133-151.
47. E Lima, C Mues and B Baesens, "Domain knowledge integration in data mining using decision tables: case studies in churn prediction", Journal of the Operational Research Society , 60, 2009, pp.1096–1106.
48. Chih-Fong Tsai, Yu-Hsin Lu, "Customer churn prediction by hybrid neural networks", Expert System with Applications 36, 2009, pp. 12547–12553.
49. J. Burez, D. Van den Poel, "Handling class imbalance in customer churn prediction", Expert Systems with Applications, Volume 36, Issue 3, Part 1, 2009, Pages 4626–4636.
50. XIA Guo-en1, JINWei-dong, "Model of Customer Churn Prediction on Support Vector Machine", Systems Engineering — Theory and Practice Volume 28, Issue 1, 2008.
51. Tarik Rashid, "Classification of Churn and non-Churn Customers for Telecommunication Companies" , International Journal of Biometrics and Bioinformatics (IJBB), Volume 3, Issue 5, 2008.
52. Kristof Coussement , Dirk Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques", Expert Systems with Applications, Volume 34, Issue 1 , 2008, Pages 313–327.
53. Kiss, C. and Bichler, M., "Identification of influencers – Measuring influence in customer networks", Decision Support Systems, 46 (1), 2008, pp. 233–253.
54. Tarik Rashid, "Classification of Churn and non-Churn Customers for Telecommunication Companies", International Journal of Biometrics and Bioinformatics (IJBB), Volume 3, Issue 5, 2008.
55. DB Seo, C Ranganathan, Y Babad, "Two-level model of customer retention in the US mobile telecommunications service market", Telecommunications Policy, Volume 32, Issues 3–4, 2008, Pages 182–196.
56. Kristof Coussement , Dirk Van den Poel, "Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. Expert Systems with Applications, Volume 36, Issue 3, Part 2, 2008, Pages 6127–6134.
57. Efraim Turban, Ramesh Sharda, Dursun Delen, Turban Efraim, Decision Support and Business Intelligence Systems.published by Pearson Education, 2007.
58. Bin, Peiji, and Juan, "Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service", Service Systems and Service Management, 2007, International Conference, 1-5.
59. John, Ashutosh, Rajkumar, and Dymitr, "Computer assisted customer churn management: State-of-the-art and future trends", Computers and Operations Research, Volume 34, Issue 10, 2007, Pages 2902–2917.
60. Lazarov and Capota. "Churn Prediction", 2007, TUM computer science.
61. Chu, B. H., Tsai, M. S., and Ho, C. S., "Towards a hybrid data mining model for customer retention", Knowledge-Based Systems, 20, 2007, pp. 703–718.
62. Yun Chen, Guozheng Zhang, Dengfeng Hu,Chuan Fu., "Customer segmentation based on survival character", Journal of Intelligent Manufacturing, Volume 18, Issue 4, 2007 pp 513-517
63. A Eshghi, D Haughton, H Topi., "Determinants of customer loyalty in the wireless telecommunications industry", Telecommunications Policy, Volume 31, Issue 2, 2007, Pages 93–106.
64. Jiayin Qi, Feng Wu, Ling Li, Huaying Shu., "Artificial intelligence applications in the telecommunications industry", Expert System with Applications, Volume 24, Issue 4, 2007, pages 271–291.
65. J. Hadden, A. Tiwari, R. Roy, D. Ruta., "Computer assisted customer churn management: State-of-the-art and future trends", Computers and Operations Research, vol. 34, no. 10, 2007, pp. 2902-2917.
66. Jae-Hyeon Ahna, Sang-Pil Hana, Yung-Seop Lee., "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", Telecommunications Policy, Volume 30, Issues 10–11, 2006, Pages 552–568.
67. Shin-Yuan Hung , David C. Yen, Hsiu-Yu Wang., "Applying data mining to telecom churn management", Expert System with Applications, 31, 2006, 515–524.
68. Kim, Y. S., "Toward a successful CRM: Variable selection, sampling, and ensemble", Decision Support Systems, Volume 41, Issue 2, 2006, Pages 542–553.
69. J. Hadden, A. Tiwari, R. Roy, and D. Ruta., "Churn Prediction: “Does Technology Matter?”. International Journal of Intelligent Technology, 1(1), 2006, 104-110.
70. J. Hadden, A. Tiwari, R. Roy, and D. Ruta., "Churn Prediction using Complaints Data”. International Journal of Intelligent Technology, 13, 2006, 158-163.
71. Shin-Yuan Hung , David C. Yen, Hsiu-Yu Wang., "Applying data mining to telecom churn management”, Expert System with Applications 31, 2006, 515–524.
72. Zhiguang Qian, Wei Jiang and Kwok-Leung Tsui., "Churn detection via customer profile modeling", International Journal of Production Research , Volume 44, Issue 14, 2006, pages 2913-2933.
73. Ahn, H., Kim, K. J., and Han, I., "Hybrid genetic algorithms and case-based reasoning systems for customer classification”. Expert System with Applications, 23, 2006, 127–144.

74. M Chandar, Arijit Laha, P Krishna., "Modeling churn behavior of bank customers using predictive data mining techniques", National conference on soft computing techniques for engineering applications, 2006, Pages 24-26.
75. Kim, J. K., Song, H. S., Kim, T. S., and Kim, H. K., "Detecting the change of customer behavior based on decision tree analysis", Expert System with Applications, 22, 2005, 193–205.
76. Kim, Y. A., Song, H. S., and Kim, S. H., "Strategies for preventing defection based on the mean time to defection and their implementations on a self-organizing map", Expert System with Applications, 22, 2005, 265–278.
77. Wang, K., Zhou, S., Yang, Q., and Yeung, J. M. S., "Mining customer value: From association rules to direct marketing", Data Mining and Knowledge Discovery, 11, 2005, 57–79.
78. Kyoungnam Ha, Sungzoon Cho, Douglas MacLachlan., "Response models based on bagging neural networks", Journal of Interactive Marketing, Volume 19, Issue 1, 2005, Pages 17–30.
79. Ngai, E. W. T., "Customer relationship management research (1992–2002): An academic literature review and classification", Marketing Intelligence, Planning, 23, 2005, 582–605.
80. Poel, D. V. D., and Buckinx, W., "Predicting online-purchasing behavior", European Journal of Operational Research, 166, 2005, 557–575.
81. Etzion, O., Fisher, A., and Wasserkrug, S., "E-CLV: A modeling approach for customer lifetime evaluation in e-commerce domains, with an application and case study for online auction", Information Systems Frontiers, 7, 2005, 421–434.
82. Kuo, R. J., Liao, J. L., and Tu, C., "Integration of art neural network and genetic kmeans algorithm for analyzing web browsing paths in electronic commerce", Decision Support Systems, 40, 2005, 355–374.
83. Chen, M. C., Chiu, A. L., and Chang, H. H., "Mining changes in customer behavior in retail marketing", Expert Systems with Applications, Volume 28, Issue 4, May 2005, Pages 773–78.
84. Douglas, S., Agarwal, D., and Alonso, T., "Mining customer care dialogs for ‘‘daily news’’. IEEE Transactions on Speech and Audio Processing, Volume:13 , Issue: 5, 2005, 652-660.
85. Larivie‘Re, B., and Poel, D. V. D., "Predicting customer retention and profitability by using random forests and regression forests techniques", Expert Systems with Applications, 29, 2005, 472–484.
86. LarivieRe, B., and Poel, D. V. D. , "Investigating the post-complaint period by means of survival analysis", Expert Systems with Applications, 29, 2005, 667–677.
87. Yu Zhao, Bing Li, Xiu Li, Wenhuang Liu, and Shouju Ren., "Customer Churn Prediction Using Improved One-Class Support Vector Machine", ADMA , LNAI, 3584, 2005, pp. 300–306.
88. Changchien, S. W., Lee, C. F., and Hsu, Y. J., "On-line personalized sales promotion in electronic commerce", Expert Systems with Applications, 27, 2004, 35–52.
89. Hwang, H., Jung, T., and Suh, E., "An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry", Expert System with Applications, 26, 2004, 181–188.
90. Cho, Y. H., and Kim, J. K., "Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce", Expert System with Applications, 26, 2004, 233–246.
91. Kim, Park, and Jeong., "The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services", Telecommunications Policy. Volume 28, Issue 2, 2004, Pages 145–159.
92. Ahmed, S. R., "Applications of data mining in retail business", Information Technology: Coding and Computing, 2, 2004, 455–459.
93. Berry, M. J. A., and Linoff, G. S., "Data mining techniques second edition – for marketing, sales, and customer relationship management", 2004.
94. Wiley. Berson, A., Smith, S., and Thearling, K., "Building data mining applications for CRM", McGraw-Hill, 2000.
95. Kim H S, Yoon C H., "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market", Telecommunications Policy, Volume 28, Issues 9–10, 2004, Pages 751–765.
96. Buckinx, W., Moons, E., Poel, D. V. D., and Wets, G., "Customer-adapted coupon targeting using feature selection", Expert Systems with Applications, 26, 2004, 509–518.
97. Song, H. S., Kim, J. K., Cho, Y. B., and Kim, S. H., "A personalized defection detection and prevention procedure based on the self-organizing map and association rule mining: Applied to online game site", Artificial Intelligence Review, 21, 2004, 161–184.
98. B Larivière, D Van den Poel., "Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services", Expert Systems with Applications, Volume 27, Issue 2, 2004, Pages 277–285.
99. Au, W. H., and Chan, K. C. C., "Mining fuzzy association rules in a bank-account database", IEEE Transactions on Fuzzy Systems, 11, 2003, 238–248.
100. Au, W. H., Chan, K. C. C., and Yao, X., "A novel evolutionary data mining algorithm with applications to churn prediction", IEEE Transactions on Evolutionary Computation, 7, 2003, 532–545.
101. Lau, H. C. W., Wong, C. W. Y., Hui, I. K., and Pun, K. F., "Design and implementation of an integrated knowledge system", Knowledge-Based Systems, 16, 2003, 69–76.
102. Au, W. H., Chan, K. C. C., and Yao, X., "A novel evolutionary data mining algorithm with applications to churn prediction", IEEE Transactions on Evolutionary Computation, 7, 2003, 532–545.
103. Chen, Y. L., Hsu, C. L., and Chou, S. C., "Constructing a multi-valued and multilabeled decision tree", Expert Systems with Applications, 25, 2003, 199–209.
104. Chiang, D. A., Wang, Y. F., Lee, S. L., and Lin, C. J., "Goal-oriented sequential pattern for network banking churn analysis", Expert Systems with Applications, 25, 2003, 293–302.

105. J Lu., "Modeling Customer Lifetime Value Using Survival Analysis: An Application in the Telecommunications Industry", Data Mining Techniques, 2003, SUGI 28.
106. Kim, Y. H., and Moon, B. R., "Multicampaign assignment problem. Knowledge and Data Engineering", IEEE Transactions, Volume:18 , Issue: 3, 2003, 405-414.
107. Rosset, S., Neumann, E., Eick, U., and Vatnik, N., "Customer lifetime value models for decision support", Data Mining and Knowledge Discovery ,Volume 7, Issue 3, 2003, pp 321-339
108. Koh, H. C., and Chan, K. L. G., "Data mining and customer relationship marketing in the banking industry", Singapore Management Review, 24, 2002, 1–28.
109. Chih Ping Wei, I-Tang Chiu, "Expert Systems with Applications", Volume 23, Issue 2, August 2002, Pages 103–112.
110. Lejeune, M. A. P. M., "Measuring the impact of data mining on churn management", Internet Research: Electronic Networking Applications and Policy, 11, 2001, 375–387.
111. Berson, A., Smith, S., and Thearling, K. (2000). Building data mining applications for CRM. McGraw-Hill.
112. Madden, G., Savage, S. J., and Coble-Neal, G., "Subscriber churn in the Australian ISP market", Information Economics and Policy,11 (2), 1999, 195–207
113. Nord, J. H., and Nord, G. D., "MIS research: Journal status and analysis", Information and Management, 1995, 29, 29–42.
114. Berson, A., Smith, S., and Therling, K. (1999). Building data mining applications for CRM . New York: McGraw-Hill.

Nabgha Hashmi is recently doing her MS from University of Gujrat, Pakistan. She has got her BS (Hons) degree in 2011 from the same university. Her research interests include Data Mining, Machine Learning, and Knowledge Discovery.

Naveed Anwer Butt received his B.Sc. from Punjab University, M.Sc. in Computer Science from Agriculture University Faisalabad and MS in Computer Science from International Islamic University, Islamabad Pakistan in 1997, 2000 and 2010, respectively. He is currently with University of Gujrat, Faculty of Computing & Information Technology as an Assistant Professor and having 12 Years of Industrial\Teaching\Research experience in various Organization and Universities. His research interests include Knowledge Discovery (data/text/web mining) for Electronic Commerce by using Sequential Patterns, Association Rules, Clustering, and Machine learning techniques i.e. Boosting, CBR, DTs, Natural Language Processing and Multi-agent Systems. His articles have been published in academic journal such as Robotics and Computer Integrated Manufacturing and, IEEE transactions.

Dr. Muddesar Iqbal is working as an Associate Professor with University of Gujrat Pakistan. He is also Director to the Faculty of Computing & Information Technology at the university. He has done PhD from Kingston University UK in the area of Wireless Mesh Networks entitled "Design, Development and Implementation of a High Performance Wireless Mesh Network for Application in Emergency and Disaster Recovery". He has undertaken research project both in the area of healthcare and disaster recovery. In 2007, he was granted a doctoral training award of £ 54000 + (Rs. 7,383,294+) from Engineering and Physical Sciences Research Council (EPSRC), UK to undertake a research entitled "4G-based Reconfigurable Mobile Healthcare System", In 2009, He was granted a funding of £24,900 (Rs. 3,404,902) by Technology Strategy Board, UK under grant number BV192W as principal investigator to conduct research feasibility Study for Digital Britain. The study entitled " A load balanced automated wireless network for community" was carried at Swansea University. Upon successful completion of the first round of feasibility study, in September 2010 technology strategy board UK awarded further funding of £72,075 (Rs. 9,855,441) in the second round against the proposal entitled "A Secure and Trusted Community Wireless Mesh Network (AST-Net) "under grant number: 3424-27170. Dr. Iqbal is worked as Co-investigator with Dr. Henry Wang from Swansea University (Principle Investigator) on this project .Dr. Iqbal has several years of teaching experience with British Institutes and taught Greenwich University UK BSC final year, British Computer Society (BCS), Association of Business Executive (ABE) courses at several Institutes in UK. In 2006, he won an Award of Appreciation from the Association of Business Executive (ABE) UK for tutoring the prize winner in Computer Fundamentals module. He also received Foreign Expert Certificate from State Administration of Foreign Experts Affairs, People's Republic of China in 2008 against his research collaborations in china. He won another Award of Appreciation from ABE UK for tutoring the prize winner in Information System Project management module in 2010. He has published 15 papers, all in International Journals and proceedings in the area of Wireless Networks targeting its application in Healthcare and Emergency and Disaster Recovery.

A proposed visualization tool for multilayer conceptual representation

Ahmed Sharaf Eldin[1] Helwan University

Adel Elsayed[2] University of Bolton

Mona M. Nasr[3] Helwan University

Mostafa Thabet[4] Fayoum University

Abstract

Visualization is the graphical or semi-graphical representation of information that aids human comprehension of and reasoning about that information. Visualization tools are critically important for creating, querying, visualizing and validating Conceptual Data. Conceptual Representation is an efficient infrastructure that enhances visibility of knowledge on the Content, the lack of robust and efficient tools to visualize, analyze and represent this Type of representation within Content framework constraint remains a big challenge. In this study, Conceptual delivery and their visualization tools were reviewed. ADS (Conceptual ,Delivery, Student Tool), a tool to evaluate and represent formal description of concepts, relationships of data models within a given framework for publishing contents of the Materials more efficiently is developed and proposed. Performance of ADS was investigated on samples of Delivery sessions. The proposed tool showed remarkable improvement over the existing tools as it aids a better comprehension of the syntax and semantics of Conceptual delivery investigated in this study.

Keywords: *Conceptual Representation, Concept Mapping, Conceptual Delivery, Visualization tool, Authoring tool, Delivery tool, Concepts maps Technologies.*

I. INTRODUCTION

People assumed that Students are proficient learners, can understand and apply knowledge gained from texts. However they may face difficulties in extracting important information, relating ideas, and organizing concepts in useful structures. These difficulties become worse by a number of factors.

One factor is that the knowledge domain may be new to them. Also the texts itself may cause comprehension problems. The failure to use appropriate learning strategies makes the problem worse. Learning strategies are flexible plans applied to put ideas together in a meaningful structure and apply them to different contexts, helping students meet academic requirements. Instructing students by using appropriate learning strategies can overcome these difficulties.

Concept mapping, as a learning strategy, has received considerable attention over the past twenty years. It is similar to other approaches such as networking, multiple-relationship mapping, and knowledge mapping. Concept mapping is a cognitive strategy which is suited to clarify relationships between ideas or concepts. [10][11]

Historically, from the early 1970s, the concept map appeared in the work of Ausubel, Novak, and Hanesian. Concept mapping was introduced originally to explore meaningful learning acquired through audio-tutorial instruction in science. Meaningful learning occurs when learners can connect new knowledge with something they already know. Ideas and knowledge need to be linked to each other when many ideas are presented at once; they are integrated into a conceptual whole. [8]

Concept maps can be used to improve learning and teaching in four ways:

- It can be used as a learning strategy.
- It can be used as a teaching strategy.
- As a tool in the instructional design process
- Used to assess the student's understanding. [9]

Concept mapping is one of visual-spatial strategies that facilitate thinking and learning about the information by extracting, reorganizing, and representing it externally.

Concept maps differ from other learning strategies:

- Concept maps are graphic illustrations of written statements.
- It enables learners to represent not only the ideas itself but also its interrelationships.
- Concept maps help teachers design better lesson by increasing students' cognitive learning, enhancing their understanding and help them thinking, analysis, and apply their gained knowledge. [10]

In the process of mapping concepts, a student's concepts and ideas are declared and, thus, misunderstanding may be corrected. Implementations of the concept-mapping strategy can take various forms. In all of them learner is required to take the initiative for his own learning by participating in the process of abstracting

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

ideas, clarify their relationships, and put them in a meaningful structure. [8][10]

A presentation is a structured and organised form of data delivery, operating through a suitable delivery medium. The goal of a presentation is illustrate some appropriate cognitive processes that lead to acquisition of information or knowledge.

Many students find electronic concept mapping very useful, as it minimizes the exhausting and time-consuming activity of erasing, revising, and beginning anew. It allows them greater freedom to adjust their conceptual thinking and mapped representations. [10]

Concept maps are visual models structured as directed or undirected graphs. Concept maps and their variations, such as mind maps, proved to be useful as graphical knowledge representation tools for human users who are able to derive semantics from context. In particular, concept maps are extensively used in education to present knowledge to students, and to enable students to "construct" their own Knowledge in a structured way. [1]

Conceptual mapping can be defined as a cognitive externalisation of a conceptualisation represented in a graphical form and the standard form used in this type of mapping is a node-arc graph in which the nodes represent concepts and the arcs connecting them represent the relationships between the respective concepts.

The above description of conceptual mapping is more akin to a ‘formal ontology’ than to a ‘cognitive activity’. This is an important distinction since formal ontologies are meant to support computer processing of knowledge representation while cognitive activities are more suited to human use. There are systems that sit somewhere between formal ontology representation and cognitive mapping activities e.g. Topic Maps, which has an ANSI standard. For this framework, Content description framework scheme is proposed [7]. The main contribution of the scheme is to integrate classical content hierarchies with conceptual models, thus closing the gap between the two entry points of classical documentation: the content at the front end and the index at the back end. Thus, the Content description framework scheme is composed of three parts: a discipline based hierarchy at the top, followed by a topic description layer in the middle and ends up with conceptual clusters representing domain models at the bottom. The Joint Academic Coding Scheme (JACS) used for subject classification provides a good framework for the top part, the discipline based hierarchy. Figure 1 illustrates the nine layers that make up the Content description framework, including three layers representing the JACS description.

<table>
<tr><td rowspan="3">Repository management level</td><td>Subject group</td><td rowspan="3">Subject classification</td></tr>
<tr><td>Principal subjects</td></tr>
<tr><td>Subject classification</td></tr>
<tr><td rowspan="3">User interface level

(document usability)</td><td>Topic Hierarchy</td><td rowspan="3">Topic description (semi-formal)</td></tr>
<tr><td>Rhetorical model</td></tr>
<tr><td>Domain model</td></tr>
<tr><td rowspan="3">Machine processing level

(mapping on domain ontology)</td><td>Activity relations</td><td rowspan="3">Ontology representation (formal)</td></tr>
<tr><td>Property relations</td></tr>
<tr><td>Class relations</td></tr>
</table>

Figure 1: Multilayer content model

JACS: Joint Academic Coding Scheme, available online at website of the Higher Education Statistics Agency [2]

II. PROBLEM OF THE RESEARCH

Students and learners are facing many difficulties during their learning process. Mainly they could not connect their old knowledge with the new knowledge which they learn. There are some factors may cause problems as the knowledge domain may be new to them and the texts itself may cause cognitive problems.

The concept mapping helps learners to learn, as it enables students to extract the new concepts from their learning materials and connecting it with their old knowledge, clearing the relations between all this concepts.

Students find that using concept mapping is very useful which facilitate understanding and retain but when use paper-and-pencil to apply the concept mapping technique it be time consuming and exhaustive tasks. They spend a significant amount of time and effort revising and maintaining concept maps. With the advancement of information and communication technologies, these problems will disappear.

Electronic concept map minimizes the time and effort that the students spend in constructing their concept map and its activities as erasing, revising, and beginning a new map also it enable them to share their concept maps. Here we will implement a tool that will support concept mapping methodology using Content description framework shown in Figure 1 and conceptual representations by seeding and harvesting of content shown in figure 2. And develop a delivery strategy based on conceptual representation to be used by tutor

A. Objectives of the research

Our main objective is to improve concept mapping process through the following:

- Help students get rid of the dull existed in the static presentations and in the traditional teaching ways.
- Help Learners to not spend a significant amount of time and effort revising and maintaining their new knowledge connecting it with the old knowledge.
- Help presenters to do interactive presentation without huge efforts.
- Visualize the Conceptual Mapping at each layer shown in figure 1 and show interaction and continuity from layer to layer and conceptual representations by seeding and harvesting of content like shown in figure 2.
- Development of a delivery strategy based on conceptual representation

B. Motivation of the research

From my studying and teaching experience I found that the students must connect their new knowledge with their old knowledge that makes them deeply understand their subjects. The concept map enables students to achieve that. But traditional concept mapping is time consuming process also it need more effort to be done and reviews.

Students found electronic concept mapping to be a way to encourage a focused discussion with others. They also found it is easily share. It makes the result from concept mapping understandable to others. There are many attempts to Develop Concept map tool. Here we will implement a tool that will support Populating conceptual representations by seeding and harvesting of content shown in figure 2 and apply field study on students which will enable us to evaluate this tool.

Search engine	**Data flow**	**Conceptual reps**	**Contributor**	**remarks**
Seeding concepts for search	<<<	Syllabus outline	Tutor	The process of populating CR by students should reflect level of engagement with content during learning
		Expanded bullet-points	Tutor	
		Populated from presentations	Student	
Harvesting docs for reading	>>>	Populated from reading	student	

Figure 2 - Populating conceptual representations by seeding and harvesting of content

C. Solution Approach

Since other conceptual mapping tools like (CMap, ICTool) don't care of different layers shown in figure 1 and representations by seeding and harvesting of content shown in figure 2 in concept map representation so we will implement a tool which will support concept mapping methodology using layers shown in Figure 1 and visualize this layers and interaction between them which will enhance concept mapping process.

D. Scope of the research

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Development of a delivery strategy based on conceptual representation which will enable tutor to represent lecture using concept maps and increase interactivity between student and learner and make learner get rid of usual delivery methods e.g. Bullet points

III. CONCEPT MAPS

Concept maps are graphical tools for organizing and representing knowledge. They include concepts, usually enclosed in circles or boxes of some type, and relationships between concepts indicated by a connecting line linking two concepts. Words on the line referred to as linking words or linking phrases, specify the relationship between the two concepts. We define concept as a perceived regularity in events or objects, or records of events or objects, designated by a label. The label for most concepts is a word, although sometimes we use symbols such as + or %, and sometimes more than one word is used. Propositions are statements about some object or event in the universe, either naturally occurring or constructed. Propositions contain two or more concepts connected using linking words or phrases to form a meaningful statement. Sometimes these are called semantic units, or units of meaning. Figure 3 shows an example of a concept map that describes the structure of concept maps and illustrates the above characteristics.[6]

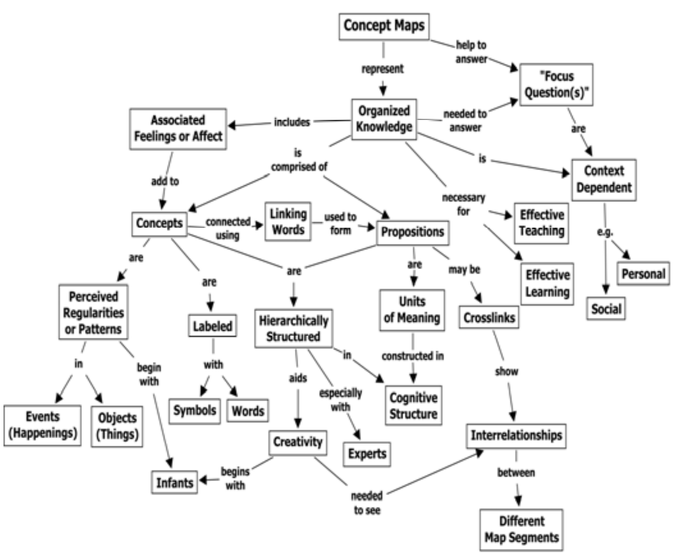


Figure 3 A concept map showing the key features of concept maps. Concept maps tend to be read progressing from the top downward.[5]

Concept Maps and Curriculum Planning

In curriculum planning, concept maps can be enormously useful. They present in a highly concise manner the key concepts and principles to be taught. The hierarchical organization of concept maps suggests more optimal sequencing of instructional material. Since the fundamental characteristic of meaningful learning is integration of new knowledge with the learners' previous concept and propositional frameworks, proceeding from the more general, more inclusive concepts to the more specific information usually serves to encourage and enhance meaningful learning. Thus, in curriculum planning, we need to construct a global "macro map" showing the major ideas we plan to present in the whole course, or in a whole curriculum, and also more specific "micro maps" to show the knowledge structure for a very specific segment of the instructional program. Faculty working independently or collaboratively can redesign course syllabi or an entire curriculum. For example, faculty working together to plan instruction in veterinary medicine at Cornell University constructed the concept map shown in Figure 4.

Using concept maps in planning a curriculum or instruction on a specific topic helps to make the instruction "conceptually transparent" to students. Many students have difficulty identifying the important concepts in a text, lecture or other form of presentation. Part of the problem stems from a pattern of learning that simply requires memorization of information, and no evaluation of the information is required. Such students fail to construct powerful concept and propositional frameworks, leading them to see learning as a blur of myriad facts, dates, names, equations, or procedural rules to be memorized.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 2, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

For these students, the subject matter of most disciplines, and especially science, mathematics, and history, is a cacophony of information to memorize, and they usually find this boring. Many feel they cannot master knowledge in the field. If concept maps are used in planning instruction and students are required to construct concept maps as they are learning, previously unsuccessful students can become successful in making sense out of science and any other discipline, acquiring a feeling of control over the subject matter [13]

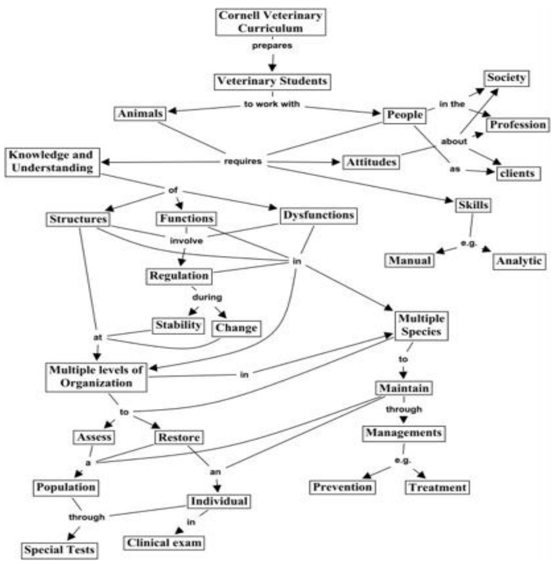


Figure 4. A concept map prepared cooperatively by the faculty of the College of Veterinary Medicine at Cornell University to show the overall structure for a revised curriculum[13].

Interaction with content

Interaction with content relates to the generation and description of content. Generation is based on seeding and harvesting through search engines. Description of content is based on a framework that maps content at three levels: subject level, topic level and ontological level that is shown in figure 1. The schematics shown in figure 2 provide a concise depiction of the above principles.

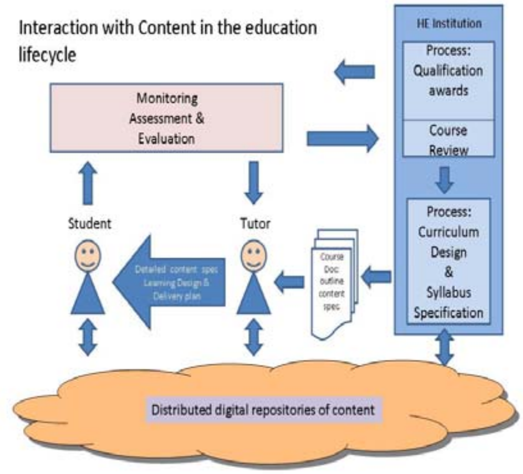


Figure 5.The curriculum lifecycle

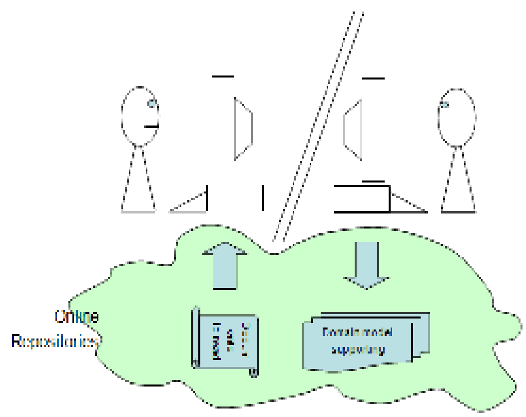


Figure 6 –Tutor-learner interactions through content

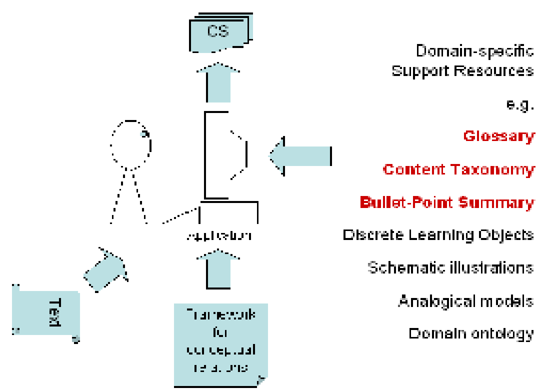


Figure 7: Learning support resource

IV. DIFFERENT HYPOTHESIS FOR DELIVERY METHODS

Since concept maps are approved to be better way for learning as its graphical representation for concepts and relationships and it represent concepts as long term memory of human save it So We introduce three strategies for delivery of content using concept maps:

1- Based totally in conceptual representations “map” and add only some rhetoric’s and some Illustration objects and when we applied it into first chapter of knowledge engineering material we found that it causes confusion to the learners but when it was applied to students in a real session they preferred this way than normal sessions “using bullet points”

2- Usual delivery method in which we use bullets ,narratives , definitions and diagrams to explain concept before adding it to the map and when we applied it into first chapter of knowledge engineering material we found that concept maps useful to :

- Determine semantic field
- Determine sequential order
- And previous points used to provide focus as a follow up to the delivery
- Reviewing

3- Interactive delivery based on conceptual representation in which we mention a concept or set of concepts to fire learners’ brains and then go though explaining them with an interactive way to save learners info. About concepts and add what I need to his info or modify his info about this concept check [appendix 1] for flow chart

We expect this strategy is the superior strategy in learning as it increase interactivity between student and learning program and learners “tutor” so we apply it to students after fishing the website that will help learner and student to apply this strategy

CONCLUSION A Concept Visualization tool has been developed for Representation of conceptual mapping and it’s usage in Lectures delivery.

The implemented tool will help learners to deliver lectures in the help of conceptual delivery which will also help students to receive lecture in the way that his brain store it.

It was found that using conceptual Delivery using help of instructor “instructor led” or without help of instructor “instructor less” achieve more efficient results than using Normal Delivery Method.

Our Proposed tool will enhance the interaction between learner and students and give students hand on to start from what his instructor ended to build his concept map for each topic he is studying.

Also we found that the gap in exam degrees between students with different Levels has been decreased when using the proposed tool

V. REFERENCES

[1] Special Double Issue: Concept Mapping (1997). Journal of Interactive Learning Research, 8 (entire issue).

[2] http://www.hesa.ac.uk/index.php?option=com_content&task=view&id=158&Itemid=233

[3] Brian R. Gaines, Mildred L. G. Shaw," Web Map: Concept Mapping on the Web”, Knowledge Science Institute (http://ksi.cpsc.ucalgary.ca/KSI), University of Calgary, Alberta, Canada T2N 1N4.

[4] IHMC-A University Affiliated Research Institute, http://cmap.ihmc.us/.

[5] http://cmapskm.ihmc.us/servlet/SBReadResourceServlet?rid=1064009710027_1483270340_27090

[6] http://cmap.ihmc.us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps.htm

[7] Elsayed A ”Interaction with content through the curriculum lifecycle”

[8] Snead, Donald; Snead, Wanda L." Concept mapping and science achievement of middle grade students", Journal of Research in Childhood Education, 22-JUN-04.

[9] Wang, Charles Xiaoxue; Dwyer, Francis M." Instructional effects of three concept mapping strategies in facilitating student achievement", International Journal of Instructional Media, 22-MAR-06.

[10] De Simone, Christina "Applications of concept mapping", Heldref Publications,College Teaching, 01-JAN-07.

[11] aidman, eugenev v , egan, greald" academic assessment through computerized concept mapping: validating a method of implicit map reconstruction ", Westwood Press, Inc., International Journal of Instructional Media, 22-JUN-98.

[12] Attia S. , Sharaf Eldin A., Elsayed A., Nasr M, kamal A, “Effect of Using A Graphical Conceptual Representation Tool on Student Understanding”, 5th International Conference, ICSRA, Cairo University, Egypt, 22-24 Dec. 2009.

[13] Bascones & Novak, 1985; Novak, 1991, 1998

[14] Elsayed, A. “A framework for conceptual relations” M3CIW’09 @ ICALT, Riga, Latvia July 14-18 2009.

[15] Hartley R, Elsayed A, Pesheva M, “Presenting and representing knowledge to develop understanding”M3CIW’09 @ ICALT, Riga, Latvia July 14-18 2009.

[16] Qiu Y, Elsayed A. “Cognitive Support for Conceptual Acquisition through Reading“, M3CIW’08 @ ICALT,Santander, Spain July 2-7 2008.

[17] Novak J, Canas, A. “The Theory Underlying Concept Maps and How to Construct and Use Them” Technical Report IHMC CmapTools, Florida 2008.

[18] El-Khouly M, Elsayed A. “A Case Study on Using Conceptual Mapping“M3CIW’08 @ ICALT, Santander, Spain July 2-7 2008.

[19] Webster D, Xu J, Mundy D, Warren P. “A Practical Model For Conceptual Comparison Using A Wiki” M3CIW’09 @ ICALT, Riga, Latvia July 14-18 2009.

Preliminary Identification of Performance – Oriented Competences for Undergraduates' Entrepreneurial Education via Information Communication Technology (ICT) for Wealth Creation in Enugu State, Nigeria.

Ezegbe B.Nkiruka1, Eskay Michael 2, Anyanwu Joy3, Ikwumelu, SN 4

1 Social Science Department, University of Nigeria
Nsukka, Enugu, Nigeria

2. Educational Foundations, University of Nigeria
Nsukka, Enugu, Nigeria.

3. Educational Foundations, University of Nigeria
Nsukka, Enugu, Nigeria.

4 Educational Foundations, Ebonyi State University
Ebonyi, Nigeria

Abstract

This study sought to find out from the perceptions of the undergraduate students, the ICT and Entrepreneurial Competences needed for wealth creation. The study was carried out in Enugu State of Nigeria. The population of the study comprised of approximately 3000 final year students of University of Nigeria Nsukka (UNN)and Enugu State University of Technology Enugu (ESUT).Two hundred students was purposively sampled from the two schools, 100 from each school. Questionnaire was used to collect data. The instrument was validated by two lecturers in the Department of Social science Education, UNN. Data was analyzed using a modified four point rating scale of Strongly Agree (4points), Agree (3points), Disagree (2points) and Strongly Disagree (1point). Mean was used to answer the two research questions. The bench mark for the acceptable value is from 2.50 and above. The major findings of the study revealed that almost all the identified entrepreneurial competences were perceived by the respondents as the competences for wealth creation, while all the identified ICT competencies were accepted for entrepreneurial empowerment. Based on the findings of the study the following conclusion and recommendations among others were drawn. School Administrators should try to organize workshops intermittently, inviting resource persons who are experts in different field of study to acquaint the students with practical skills that will help them to be functional in a specific area of interest.

Keywords: ICT, Education, Entrepreneurial Competences, Special Education,

1. Introduction

The quest for a functional educational system that is responsive to the needs and aspirations of the society has been the heart beat of nations of the world. This is because education has been recognized as a powerful instrument that determines the level of individual and development of any nation. Various reforms in education sector all over the globe were geared towards producing effective and functional members of the society. Information and Communication Technology (ICT) and Entrepreneurial education are one of such reforms. This will help to ameliorate the problems of unemployment in the society.

The issue of unemployment among youths could be linked to the nature of school curriculum which failed to prepare the recipients to be job creators rather than job seekers. Lack of technological skills have rendered so many youths poor as they could not function effectively in the modern information technology. Yusuf, (2005) maintained that modern information technology can be of assistance in accelerating, enriching and deepening skills acquisition, help to relate school experience to work practices and equally, help to create economic viability for tomorrows working class. Equally, Nwagwu, in Oku and Onuoha (2010) pointed out that Nigerian educational system was not equipping the recipients with the needed skills for economic development. It was as a result of this anomaly that the Federal Government, according to Oku and Onuoha (2010) advocated for infusion of the National Economic Empowerment Development Strategy (NEEDS) into education in order to equip the youths with relevant skills for their present and future self-relevance.

Consequently, it is pertinent to note that the objectives of the NEEDS may hardly be attained, unless Nigeria is able to take advantage of the opportunities offered by modern technology, Modern technology, no doubt, holds promise for youths' education by helping to shape environment that will allow them to create, explore, and learn by doing. Modern technology has the capacity of helping students manipulate entrepreneurial skills to create wealth and participate effectively in the global knowledge economy.

Recognizing the importance of higher institution in the global knowledge economy, the Federal Government, according to Boroffice (2008), instructed that all higher institutions in Nigeria should introduce entrepreneurship as a compulsory course for all students across all disciplines with effect from the 2007/2008 academic session. Nevertheless, the Federal government's effort to establish entrepreneurship education was aimed at ensuring that our graduates acquire certain entrepreneurship sills that will make them self-reliant and independent, contributing their own quota to the development of the nation.

Entrepreneurship has been variously defined by different people in different ways and from different perspectives. For instance, Akabi (2002) and Nwafor (2007) defined it from economic perspective as the willingness and ability of an individual to seek out investment opportunities in an environment, and be able to establish and run an enterprise successfully based on the identified opportunities. Adibu and Dedekum (2006) saw it from sociological and psychological view point. Acccordind to them, entrepreneurship is the process of creating something new with value by devoting the necessary time and effort, assuming the accompanying financial, psychic and social risk. It has also been identified as the process of venturing, undertaking and assuming risks involving creative skills….Egboh (2009). Similarly, Kuratko and Hodgetts (2001) saw entrepreneur as an individual with the aim of establishing new resources or endow old ones with a new capacity, all for the purpose of creating wealth. Implicit in the above definition is that willingness, ability and capability are important factors for an entrepreneur. This is what educational system aimed at inculcating into the recipients in order to help them use the skills they have acquired to set up a profitable oriented enterprise. Through acquisition of entrepreneurial skills, a sustainable life style may be easily maintained. Muller (2004) succinctly stated that one of the benefits of entrepreneurship is that it helps students to change personal and career attitude including: ability to control one's own life, personality, creativity and interpersonal communications. Entrepreneurship Education (EE) has been defined as a process of inculcating knowledge skills, creative thinking etc. into the learner for the purposes of promoting a strong sense of self-worth and high sense of initiative (Wikipedia 2010, Etele 2007, and Osakwe 2011).

It is important to note that creativity and interpersonal communication in this present era could only be effective when one is ICT compliant The role ICT can play in enhancing the effectiveness of poverty reduction is not in doubt. According to Ezeugbor, (2008) ICT refers to a whole range of facilities or technologies involved in information processing and electronic communications, to be handled with skills and expertise for effective achievement and realization of its potentials in both education and socio-economic development. This means that for one to be a successful entrepreneur, one need to acquire certain ICT skills to function effectively in a specific area. Ezeugbor, (2008) enumerated such ICT competences as: awareness of overall development in ICT, desktop publishing, information retrieval, among others.

Despite its importance in the global economy, Nigerian ICT compliance seems to be very low compared with other countries of the world. Describing Nigeria's position in the global technology arena, Ezeugbor, (2008) reporting Osugi (2005) stated that there is a big gap between the average Nigeria computer skills and the skills of citizens of other countries. Lack of ICT skills by the majority of Nigerian youths has adversely affected entrepreneurial skills and potentials in both education and socio-economic development of individual and Nigerian society at large. There is still high rate of unemployment among Nigerian youths. Today's youths are very covetous and materialistic to the extent that they tend to lack critical thinking, logical reasoning, innovative and enduring spirit that are very vital for a good entrepreneur. This cannot help for a sustainable life style which EE promotes. Unsustainable life style according to Chike-Obi (2012) breads lawlessness, frustrates hard work and entrepreneurship and discourages genuine investment. The incessant increase of unemployed youths in Nigeria despite the on-going EE is an evidence of unsustainable life style. The problem behind the non-achievement of the objectives of EE may be attributed to students' lack of interest on the curriculum and its delivery designed for the EE and ICT.

In line with the fore-going the researchers want to find out from the youths the competences they may consider relevant for their EE and ICT. It is hoped that the findings of this study will provide additional information to the curriculum planners for EE and ICT. For youths, it will give room for wealth awareness creation as well as acquiring more knowledge on entrepreneurship and ICT skills.

The following research questions were raised to guide the study.

2. Research question:

1. What entrepreneurship competences do youths consider relevant for wealth awareness creation?
2. What ICT competences/skills do youths consider relevant for effective entrepreneurship empowerment?

3. Method:

The study was a survey of final year undergraduate youths in the two Universities in Enugu State of Nigeria, namely: University of Nigeria Nsukka (UNN), and Enugu State University of Technology (ESUT). The population of the study is approximately 3,000 final students. Purposive Simple random technique was used to select 100 students in each school. This means that 200 final year students were purposively sampled for the study.

The questionnaire consists of 3 sections-A, B and C. Section A, consists of personal data of the respondents, section B, consists of 10 items, while, section C, consists of 7 items. The questionnaire items was subjected for face validation of the expert opinion of two lecturers, in the Department of Social science Education, University of Nigeria, Nsukka. The validated instrument was further subjected to a test of reliability analysis using Cronbach Alpha statistics. Data from sections B and C were used which gave a reliability coefficient of 0.72 and 0.76 respectively.

Three research assistants were used to distribute and collect data. Data was analysed using four point rating scale of Strongly Agree (4points), Agree (3points), Disagree (2points) and Strongly Disagree (1point) to answer the two research questions. The bench mark for accepting the mean for item on the questionnaire is 2.50 and above. Data for answering research question 1 are presented in Tables 1 and 2.

4. Results:

Table1: Mean ratings of the undergraduate students on the needed entrepreneurial competence for wealth creation.

S/N	ITEMS	SA	A	D	SD	TOTAL	X	DEC
	Entrepreneurial Competence the youths considered relevant for wealth awareness creation:	(4)	(3)	(2)	(1)			
1.	Motor mechanics repairing	97	73	25	05	662	3.31	Accept

2.	Computer coupling/configuration and repairing	84	106	06	04	670	3.35	“
3.	Catering/baking	101	92	02	05	689	3.45	“
4.	Home dressing/interior decoration.	60	100	22	18	602	3.01	“
5.	Printing and Publishing	67	102	17	14	622	3.11	“
6.	Cloth making	126	54	04	16	690	3.45	“
7.	Hair dressing(barbing, plaiting, styling)	100	54	46	00	654	3.27	“
8.	Laundering (cloth washing/dry cleaning)	50	62	48	40	522	2.61	“
9.	Computerized artistic	103	64	21	12	658	3.29	“
10.	Iron Welding	10	11	111	68	363	1.82	Do not accept

Data from table 1, revealed that the mean ratings of the undergraduate youths on the needed entrepreneurial competencies for wealth creation for items 1-10 are 3.31, 3.35, 3.34, 3.01, 3.11, 3.45, 3.27, 2.61, 3.29 and 1.82 respectively.

Table 2: Mean ratings of the Undergraduate students on the needed ICT Competence skills for wealth creation.

S/N	Item	SA	A	D	SD	Total	X	Dec
	ICT competences the youths considered Relevant for their empowerment.							
1.	Ability to retrieve information using CD Roms And other commercial programmes.	102	84	08	06	682	3.41	Accepted
2.	Desktop Publishing.	180	16	03	01	775	3.86	Accepted
3.	Interactive presentation	101	94	04	01	695	3.48	Accepted
4.	Effective use of Internet.	122	78	00	00	682	3.41	Accepted
5.	Ability to prepare graphics and Artworks	120	70	04	06	704	3.52	Accepted
6.	Knowledge of the functions of the various							

Component of computer	55	72	30	43	539	2.70	Accepted

Table 2, revealed that the mean ratings of the undergraduate youths on the needed ICT competences for wealth creation for items 1-6 are: 3.41, 3.86, 3.48, 3.41 3.52 and2,70.

5. Discussion:

The findings of the study revealed that apart from item no. 13(iron welding), all other identified entrepreneurship competences were perceived as wealth awareness creation competences for the youths. The items which they agreed to be the needed entrepreneurship competences are as follows: catering/banking, cloth making, followed by computer coupling/configuration and repairing, computer artistic, hair dressing. Motor mechanics repairing, livestock rearing and maintenance, printing and publishing, home dressing/interior decoration, laundering, soap and pomade making, and fishing. However the youths disagreed that iron welding is entrepreneurship competence they need for wealth creation. Their reason might be probably because of the low status of the welding industry in Nigeria.

The above findings showed that undergraduate youths who are in different field of study still need skills in the identified areas so that on their graduation from the University, they would be able to establish small scale business that might not even need big capital to start. The implication of such is that the youth restiveness in our society will reduce as many youths would be self employed. This however, is in line with Olayinka, Adeyemi and Yusuf (2010) who succinctly stated that the relevance of technical and vocational education in the provision of skillful and technical manpower in the society will bring about gainful employment to ample numbers of the populace.

6. Conclusion and Recommendation:

Entrepreneurship Education has been accepted as one of the many reforms in education which aimed at solving the problem of graduate unemployment in the society. The acquisition of the entrepreneurship knowledge and skills will go a long way in making our graduates self-reliant and independent thereby contributing their own quota to National development. Wealth creation can be easily attained if the desired entrepreneurial skills are practically inculcated in them.

Based on the findings of this study, the researchers recommended as follows:

1. Widening of the scope of entrepreneurial curriculum to accommodate many special areas so as to allow students choice of choosing needed competencies according to need and interest.

2. School administrators should try as much as possible organize workshops, inviting resource persons who are experts in different fields of study to acquaint the students with practicals which will help them to be functional in a specific area of interest

3. Lecturers in our universities should try to adopt learner-centered approach which encourages active classroom through the use of innovative methods of curriculum delivery. This is against passive classroom where the learner is subjected to be dormant and inactive, thereby learning little or nothing.

4. Students should be allowed to select entrepreneurial skills they are interested in and should be helped to acquaint the skills through effective planning, implementation and internalization of such skills.

5. Examination on EE should be practical oriented. This will help to taste the recipient on the skills acquire.

References

Adibu, F. A. and Dedekum, S. E. (2006). Toward developing entrpreneurship consciousness among Nigerian youths. Journal of Business and Management Studies. 1,(1).

Akanbi, A.A. (2002). Entrepreneurship Education in the Tertiary school curriculum. Journal of the World Council for Curriculum and Instruction. 3(2). 102-I07.

Borffice, O. B. (2008). Building partnership for entrepreneurship development in Nigeria Universities. National Sensitization workshop on entrpreneurship development in Nigerian Universities organized by NBTI and NUC, 21st-22nd May.

Chike-Obi, M. (2012). Sustainability as a Key in life; Business and Government. A lecture presented at the University of Nigeria Nsukka during her 52 Founders Day Celebration held on oct; 4, 2012.

Egboh, S.H.O. (2009). Etrepreneurship development for employment and wealth generation, Benin City: Ethis Printer.

Etele, O. (2007). Entrepreneurship and emerging young entrepreneurs wealth creation knows no age. London: Withworth Books.

Ezeugbor, C.O.(2008). Information and Communication Technology.Competence Level of Nigerian Tertiary Institution Teachers as a Challenge to Harnessing the ICT gains in Education. In B. G. Nworgu (ed). Education in the Information Age: Global Challenges and Enhancement Strategies. Nsukka: Nigeria. University Trust Publishers. 137-143.

Kanno, T. N. in G.A. Badmus (1995). Enhancing Women Contribution to National Development Through Relevance Curriculum. Women Education and National Development and Crisis In Nigerian Education System. The Nigerian Academy of Education.

Kuratko, D, F. and Hodgets, R. M.(2001). Entrpreneurship, a contemporary approach, 5th ed. New York: Harcourt Inc.

Muller, S.L. (2004). Gender gaps in potential for entrepreneurship across countries and cultures. Journal of developmental entrepreneurship 9(3) 1999-220.

Nwafor, P.Z.(2007). Practical Approach to Entrepreneurship Small and Medium scale Enerprise.
Enugu Precision Publishers.

Okoli, J.N.(2010). Providing Entrepreneurship Education at the Secondary Education Level through Science Technology and Mathematics (STM) Education to meet the challenges of Globalization in Nigeria. International Journal of Educational Reshearch. 10 (1). 205-223.
Oku, O.O. and Onuoha, K.(2010). Entrepreneurship Education and Gender Equity: Exploring the Link. International Journal of Educational Research. 10,(1). 224-235.

Olayinka, O. Adeyemi, A.B. and Yusuf, A. O. (2010). Enhancing NCE Technical and Vocational Graduates to Develop Managerial Competencies in entrepreneurship. International Journal of Educational Research. 10,(1). 236-245.

Osakwe, R. N. (2011). A Survey on the perception of need for entrepreneurial skill acquisition among Udergraduate Students tertiary Institutions in Delta State. Delta State University (DELSU) Journal of Educational Research and Development. 10. 127-141.

Wikipedia, the Free Encyclopedia. (2010). Entrpreneurship education. Retrieved on Jan; 15, 2010, From mhtm/:file://:f: for 20%printing/ entrepreneurship20%education.

Yusuf, M. O. (2005). Information and Communication Technology: Analysis of the Nigerian National Policy for Information and Communication Technology. *International Education Journal*. Lagos: 6,(3). 316-321.

Problem of Universal Basic Education in Nigeria and the Role of Information Communication Technology (ICT) in Enhancing its Quality, Sociological and Counselling Perspective.

Anyanwu Joy1, Eskay Michael 2, Ezegbe B.Nkiruka 3, Ikwumelu, SN 4

1. Educational Foundations, University of Nigeria
Nsukka, Enugu, Nigeria.

2. Educational Foundations, University of Nigeria
Nsukka, Enugu, Nigeria.

3. Social Science Department, University of Nigeria
Nsukka, Enugu, Nigeria

4 Educational Foundations, Ebonyi State University
Ebonyi, Nigeria

Abstract

The study which was carried out in Ankpa local government of Kogi state to investigate the problems of the Universal Basic Education and the role of ICT I enhancing its quality, used survey research design. The sample of the study was 100 respondents- 80 teachers and 20 teacher counsellors randomly and purposively selected from four out of 400 (four hundred) registered UBE schools. Questionnaire was the instrument for data collection. Data collected were analysed using descriptive statistic. The decision rule was mean scores from 2.50 and above. The study revealed that the UBE program is still facing various implementation problems of personnel, funding, and infrastructure. To overcome these challenges, counsellors have a lot of roles to play to boost the quality of the man power needed for the nations, self-reliance, scientific and technological development and total empowerment. Teachers should be trained in ICT utilization in teaching science subjects and maths in *primary schools. It was recommended that adequate funds should be provided for the instructional and infrastructural facilities for effective implementation of the UBE programs in Nigeria. Qualified guidance counsellors and teachers should be provided for the UBE institutions to assist in overcoming learners academic and personal social problems.*

Key words, Universal Basic Education, Special education, Guidance Counsellors, ICT, Implementation.

1. Introduction

One of the central functions of government at all levels is the provision of qualitative and sound education for the citizens. Consequently, the Federal Government of Nigeria launched the Universal Primary Education (UPE), in 1976, aimed at providing free and qualitative education, establishing equal educational opportunities for all children of school age and gradually reduce illiteracy and ignorance in Nigeria. The UPE gave most citizens of school age and even those who had no means of pursuing education opportunity to receive free basic primary education. However, the Universal Primary Education program was faulted due to inaccurate data on children that would be affected by the scheme for planning and projection; implementation in terms of inadequate public enlightenment and social mobilization, poor quality and social welfare of teachers, inadequate funding which led to the dearth of facilities, poor management of resources available and poor supervision of personnel. These factors promoted its failure and subsequent abolition. Hence, the Universal Basic Education (UBE) was launched in September 1999 (Edho, 2009).

The introduction of UBE emanated from the world conference of Education for all (EFA) held in Jomitten, Thailand in 1990. It was also a response to international recognition of children education as stipulated by the Millennium Development Goals (MDGs) item two which proposed that by 2015, children of educable age should have free, affordable and accessible education and adopted by world nations in 2000.

The main thrust of the programme according to Anaekwe (2009) are to ensure that all 6 (six) years old children are enrolled in primary schools and complete their education; contribute to a massive reduction of illiteracy and dropout rates by making basic education free and compulsory for children; ensure that adults, handicapped children, migrant workers and out-of-school youths have access to basic education; creating skills at junior secondary school level through technical and vocational education and ensure equal access for both boys and girls, thus reducing gender disparities in school enrolment among others.

In order to achieve these, the Nigerian Educational Research and Development Council (NERDC), restructured the school curricular for the 9 years basic education. The contents of the new curriculum reflect both the emerging and national values. It has been noted however, that since the inception of the UBE scheme, qualified teachers to handle the challenges of quality of instruction has not been massively recruited. Modumogu (2006) opines that among the educational services promised by the government to provide in primary schools, educational resources centre and special teachers trained to teach particular subject are not inclusive. Besides, teachers are not provided in the right quantity and quality despite the stipulation in NPE that teachers be assigned to teach only subjects they are trained to teach.

NPE also states that practical, exploratory and experimental methods be used in teaching but this is not obtainable in the primary schools. Teachers rather use one type of method monotonously in teaching thereby producing passive learners who only memorize facts and cannot think critically and creatively. Enem (2004) equally observed that most products of this level of education lack the skill for literacy, clear expression and logical argument, integrity and self-discipline. This situation may be based on the fact that UBE accommodates children from different socio-cultural and socio-economic backgrounds and maladjustment cannot be ruled out.

Since teachers are vital in education which is the gateway to modernization and development according to kabiru (2001), they need proper training. Contributing, Nwosu (2002) opined that teachers should be trained in computer literacy to enable them educate students who would function effectively in the modern work environment and compete in the global job market. This is supported by the report of the United Kingdom's Department for International Development (DFID,2002) that Information Communication Technologies (ICT) have enormous potential as tools to increase information flow and empower poor people.

Information Communication Technologies (ICT) in simple language means all that is involved in gathering and processing information, using modern communication technologies such as computers and other related equipment so that the service generated can be assessed at reasonable cost and in good time by all that desire them to the overall benefit of mankind (Aneakwe 2008).

Recognizing the importance of ICT, the Federal Executive Council of Nigeria approved a national Information Technology (IT) policy in March 2001 with the vision to make Nigeria an IT capable country in Africa and a key player in the information society using IT as the engine for sustainable development and global competitiveness. Besides, IT is to be used for wealth creation, poverty eradication, education and global competitiveness. The establishment of the National Information Technology Development Agency (NITDA) charged with the implementation responsibility in April 2001, marked its implementation.

However, Robert-Okah (2012) noted that ICT is the least developed area in Nigerian primary and junior secondary schools as only few teachers and students can operate computer. Staff and students at this level are yet to adapt and internalize the new ICT ideas in order to appreciate and work with it thus impeding rapid diffusion of the new ICT technology.

In view of the apparent ignorance of many young people about career prospects and personal maladjustment among school children as stated by the Federal Government (FRN, 2004), the stipulation to appoint career officers and counselors ,to train interested teachers as teacher counsellors in schools becomes necessary. Experience has however shown that counselors are not posted to primary schools. The general outcome is maladjusted children with poor self-concept and lack of vocational direction which negates the objective of the UBE programme. Guidance and Counselling according to Okeke (2003) is specialized and developmental services aimed at assisting individuals by trained and qualified counsellors to achieve understanding, acceptance of self and live a productive life. Guidance and counselling furnishes clients with needed information for major decisions in vocational, educational and personal-social areas of life to achieve adjustment. While educational concerns relate to development of effective study habits and choice of subjects, vocational counselling relates to assisting individuals to understand self and the world of work, Socio-personal counselling deals with emotional, social and personal (relationship)problems encountered by individuals .

2. Statement of the Problem

In 1976, the Federal Government trained one hundred and sixty three thousand (163,000) teachers to raise quality staff for the UPE program. Yet, the UPE scheme did not record much success. The Federal Government then approved major changes in the nation's educational policies which culminated to the introduction and adoption of the UBE scheme to fulfil the aim of education for all with emphasis on free access, equity, literacy, numeracy and lifelong skills for all as endorsed at the world conference on education in 1990. In spite of the Federal Government's effort to achieve her anticipated objectives of UBE, the scheme has not registered meaningful impact in the society. Considering the importance of the UBE, it has become pertinent to investigate problems facing the implementation of the program, the role of ICT in enhancing its quality as well as its implications for counselling.

2.1 Purpose of the study:

This study investigated the problems of the UBE program, the role of ICT in enhancing the quality and its implications for counselling in terms of:

- the extent UBE has provided free, qualitative and universal basic education,
- the extent UBE has reduced the incidence of dropout from formal school system through the provision of quality basic education, and
- Assess the role of ICT in enhancing the quality of the UBE programme.

2.2. Scope of the study: This study was limited to the problems of the universal basic education, the role of ICT in its implementation and implication for counselling. The study was carried out in UBE primary schools in Ankpa Local Government Area of Kogi State.

2.3. Research Questions

- To what extent has UBE provided free and qualitative, universal basic education in the primary schools?
- To what extent has the UBE program reduced the incidence of drop out from formal school system through the provision of quality basic education?
- What are the roles of ICT in enhancing the quality of the UBE program?

3. METHOD

The study which was carried out in Ankpa local government of Kogi state used a descriptive survey research design. The sample of the study was 100 respondents consisting 80 teachers and 20 teacher counsellors randomly and purposively selected from four out of 400 (four hundred) registered schools where UBE program is operational in Ankpa local Government Area. Questionnaire was the instrument for data collection.

The questionnaire was face validated by experts in measurement and evaluation, and guidance and counselling whose corrections enhanced the final draft production. A reliability coefficient of .74 was obtained using Chronbach alpha method of reliability analysis. Data collected were analysed using descriptive statistic (mean scores). The decision rule was mean scores from 2.50 and above. Real limit of numbers was used to interpret the data.

4. RESULTS

4.1. Research Question One: To what extent has UBE provided free, qualitative and universal basic education?

Table 1:Mean and Standard Deviation Scores on the extent Ube Has Provided Free, Qualitative and Basic Education

s/n	Items	Mean	S.Dev	Remark
1	UBE has helped to provide free education	2.52	1.01	High extent
2	UBE has not done anything to improve quality education due to lack of qualified teachers.	2.35	1.07	Low extent
3	UBE has little quality to offer to education due to lack of basic teaching and learning materials like adequate classrooms and desks for both teachers and pupils.	2.70	1.08	High extent
4	UBE is not really free because of all the levies paid in the school	2.91	.93	“
5	There are schools for adults, challenged and migrants in the area.	1.75	1.02	Very low extent

N0= 100

From the Table, the respondents agreed that UBE has helped to provide free education to a high extent. It has little to offer to quality education because the basic materials are not on ground; it is not universal as the focus is mainly on normal children.

4.2. Research Question Two: To what extent has UBE reduced the incidence of dropout from formal school system through the provision of free and quality basic education?

Table 2: Mean Scores on the Extent UBE has Reduced the Incidence of Drop Out from schools.

S/No	Item	Mean	SD	Remark
6	UBE has reduced drop-out since books are provided freely.	3.04	.85	Very High extent
7	UBE practices free lunch for children thereby reducing drop-out	2.51	.88	High extent
8	UBE is free, so people no longer drop out due to none payment of school fees	2.36	.99	Low extent

9	UBE has helped more children to go back to school since schools are located close by.	2.83	1.07	High extent

No. = 100, SD= Standard deviation

Result from the table show that the respondents agreed that UBE has helped reduce drop out through free books, free schooling and proximity to their residences.

4.3. Research question three; what is the roles of ICT in enhancing the quality of UBE

Table 3; mean and standard deviation of responses on the role of ICT in enhancing the quality of UBE

s/n	Items	Mean	SD	Dec
10	Empowering citizens by enabling online teamwork for increased collaboration and information sharing through the use of email, phones among teachers	3.47	.64	Agree
11	Enabling the rapid creation and inexpensive distribution of educational information and knowledge among stake holders and teachers	3.19	..80	"
12	Encouraging professional development, in-service-training and mentoring for lifelong learning for teachers	3.30	.74	"
13	Increasing motivation for the learners through the use of multimedia such as sound, video, graphics, animation and text	3.14	..86	"
14	Allowing each student to learn at his/her level and speed thereby giving pupil greater control over their own learning.	3.26	.77	"
15	Promoting active rather than passive learning.	3.29	.87	"
16	Enhancing the development of the abilities of mentally and Physically challenged students.	3.30	.74	"
17	Engaging students in research and problem solving, thereby facilitating higher-order thinking processes such as synthesizing, interpreting and hypothesizing.	2.91	.93	"

From the table, ICT will enhance information sharing and help pupils become active among others.

5. DISCUSSION

The findings of the study show that the UBE program is not totally free because, levies charged by the UBE board or ministry of education have made the program more costly. This may be as a result of poor management of available resources to meet the needs of the programme according to Ikoya (2000). The program is also not universal because it only focuses only on the formal school learners from primary school to junior secondary school III (JSS III) neglecting the needs of the adults, and other special groups as embedded in the UBE policy.

The reduction in the rate of school dropout through UBE implementation was largely due to the free books, food and proximity to their homes. Lack of funds for the UBE program in the study area has led to non-provisions of adequate infrastructure and personnel towards achieving the educational goals. This confirms Eya and Nebos (2001) statement that lack of instructional and infrastructural materials and funding may affect implementation of UBE. These factors equally led to the fall of UPE and one would have thought that the government will take corrective measures to make UBE succeed.

On the role of ICT in enhancing the quality of UBE, the result shows that ICT will help the teachers share information with colleagues and stakeholders among others. It will equally help pupils become active learners and engage in problem solving among others. Since the Nigerian National Policy on Education lists computer education as one of the subjects at the primary school level (FRN, 2004) and considering the report of Department for International Development (DFID,2002,) that Information Communication Technologies (ICT) (which include computer) has enormous potential as tools to increase information flows and empower poor people, Nwosu's (2002) suggestion that teachers should be trained in computer literacy to enable them educate students who would function effectively in the modern work environment and compete in the global job market should be followed to enhance the UBE programme. Abdulfattah (2007) suggested that at the primary level, pupils should be taught logical reasoning under mathematics as from the fourth year. In the fifth and sixth years, computer studies should be introduced with emphasis on practical under computer education.

6. CONCLUSION

The Universal Basic Education (UBE) program is still facing various implementation problems especially, funding, personnel and infrastructure. To overcome these challenges, the use of ICT as a force to drive scientific and mathematical subjects in Nigerian schools should be a major undertaking, an investment in the future productivity of the Nigerian workforce because it can enhance the quality in educational acquisition in Nigeria, boost the man power needed for the nation's scientific and technological development and total empowerment for the future prosperity of the country. Moreover sufficient funds are needed to establish and maintain ICT in the schools. Also, continuing professional development of teachers, school heads and other educational personnel must be instituted.

It however requires the total support of and collaboration with all agencies in the education system especially the counselor whose major concern is the adjustment of individuals – both teachers and pupils in their academic, vocational and personal social aspects of life.

7. RECOMMENDATIONS:

I. Adequate funds should be provided for the instructional and infrastructural (including ICT) facilities for effective implementation of the national policy objectives of the UBE programs in Nigeria.

II. Adequate and qualified teachers and guidance counsellors should be provided for the UBE institutions to assist in overcoming learners academic and personal social problems.

REFERENCES

Abdulfattah A. Aboaba (2007) Information and Communication Technology (ICT) for Development. The Exigent and Foundational Approaches for Nigeria. *Continental J. Information Technology 1: 16 - 24,*

Aneakwe, J. U. (2008). Information and communication technology in M.O. Ofoefuna (ed) 198-209, Functional approach to educational technology. Onitsha ; Ofona Publishers.

Aneakwe, J. U. (2009). *Functional education through information and communication technology in Universal Basic Education.* In N. Ada, B. U Maduewesi, and C. I .Ihebereme (eds) Emergent issues in Universal Education. Onitsha; West and Solomon Pub. Ltd.

DFID (2002) *The significance of ICT for reducing poverty Strategy Paper*.

Edho, O.G (2009). The challenges affecting the implementation of the Universal Basic Education (UBE) I Delta State, Nigeria. *Journal of Social Sciences, 20(3), 183-187*

Enem, F. O. (2004). Vision and mission of teacher education in the 21st century Nigeria.*Journal of excellence 1(1), 124-131.*

Eya, P. And Nebo,O. (2001). Evaluation of the available instructional materials for the implementation of the UBE Programme in Enugu Education Zone. *The Nigrian Universal Basic Education Journal. 1(2). 27-35.*

Federal Republic of Nigeria, (2004).*National Policy on Education.*Lagos, Federal Ministry of Education, Printing Press.

Ikoya, P.O (2000).Efficient Management of the National Primary Education Fund for effective implementation of the U.B.E in Nigeria. *The Nigerian academy of education, 1-9*

Kabiru, Y. (2001). Teachers and sustainable national development. *Paper presented at the National Teachers Institute.*

Modumogu, C.A. (2006): Quality Nursery and Primary Education in Idoma Land in the 21st Century: Perspective, challenges and Strategies for Improvement. *Being a paper presented at the 21t catholic Education Summit of the catholic Diocese of Otukpo. St. Francis Cathedral Otukpo. 22nd – 24th Nov. 2006.*

Nwosu, S.E. (2002). *Fundamentals of computer education and educational technology*. Enugu: Cedartop.

Okeke, B.A (2003) Principles of Guidance ad counseling; An outline for Beginners. Enugu, EMA Publishing company.

Robert-Okah, I.(2012) Information and Communication Technology for UBE Implementation in Nigeria: An Imperative for Achieving Sustainable Millennium Development Goals, *Journal of Resourcefulness and Distinction, 1 (1),*

IJCSI CALL FOR PAPERS JANUARY 2014 ISSUE

Volume 11, Issue 1

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas. See authors guide for manuscript preparation and submission guidelines.

Accepted papers will be published online and indexed by Google Scholar, Cornell's University Library, DBLP, ScientificCommons, CiteSeerX, Bielefeld Academic Search Engine (BASE), SCIRUS, EBSCO, ProQuest and more.

Deadline: 15th January 2014
Online Publication: 31st January 2014

- Evolutionary computation
- Industrial systems
- Evolutionary computation
- Autonomic and autonomous systems
- Bio-technologies
- Knowledge data systems
- Mobile and distance education
- Intelligent techniques, logics, and systems
- Knowledge processing
- Information technologies
- Internet and web technologies
- Digital information processing
- Cognitive science and knowledge agent-based systems
- Mobility and multimedia systems
- Systems performance
- Networking and telecommunications
- Software development and deployment
- Knowledge virtualization
- Systems and networks on the chip
- Context-aware systems
- Networking technologies
- Security in network, systems, and applications
- Knowledge for global defense
- Information Systems [IS]
- IPv6 Today - Technology and deployment
- Modeling
- Optimization
- Complexity
- Natural Language Processing
- Speech Synthesis
- Data Mining

For more topics, please see http://www.ijcsi.org/call-for-papers.php

All submitted papers will be judged based on their quality by the technical committee and reviewers. Papers that describe on-going research and experimentation are encouraged.
All paper submissions will be handled electronically and detailed instructions on submission procedure are available on IJCSI website (www.IJCSI.org).

For more information, please visit the journal website (www.IJCSI.org)

IJCSI

The International Journal of Computer Science Issues (IJCSI) is a well-established and notable venue for publishing high quality research papers as recognized by various universities and international professional bodies. IJCSI is a refereed open access international journal for publishing scientific papers in all areas of computer science research. The purpose of establishing IJCSI is to provide assistance in the development of science, fast operative publication and storage of materials and results of scientific researches and representation of the scientific conception of the society.

It also provides a venue for researchers, students and professionals to submit ongoing research and developments in these areas. Authors are encouraged to contribute to the journal by submitting articles that illustrate new research results, projects, surveying works and industrial experiences that describe significant advances in field of computer science.

Indexing of IJCSI

1. Google Scholar
2. Bielefeld Academic Search Engine (BASE)
3. CiteSeerX
4. SCIRUS
5. Docstoc
6. Scribd
7. Cornell's University Library
8. SciRate
9. ScientificCommons
10. DBLP
11. EBSCO
12. ProQuest