Evaluation of Activity Recognition Algorithms for Employee Performance Monitoring

Mehreen Mumtaz[1], Hafiz Adnan Habib[2]
[1] Software Engineering Department , UET Taxila
Taxila, 47080, Pakistan

[2] Computer Engineering Department, UET Taxila,
Taxila, 47080, Pakistan

Abstract

Successful Human Resource Management plays a key role in success of any organization. Traditionally, human resource managers rely on various information technology solutions such as Payroll and Work Time Systems incorporating RFID and biometric technologies. This research evaluates activity recognition algorithms for employee performance monitoring. An activity recognition algorithm has been implemented that categorized the activity of employee into following in to classes: job activities and non-job related activities. Finally, the algorithm will compute the time which employee spent in job related and non-job related activities. This paper presents a novel architecture based upon video analytics that can facilitate Human Resource Managers in real time.

Keywords: *Video Analytics; Human Activity Recognition; Human Resource Management.*

1. Introduction

Human resource management (HRM) plays a vital role in managing the people working in an organization and it helps in achieving business objective of the organization. A Human Resource Management System refers to systems that interconnect human resource management (HRM) and information technology. Such systems help in reducing the workload of managers and automate their administrative activities [23].

1.1 Human Resource Management Systems

Given below is a brief description of HRM systems [23].

Payroll systems: These systems can automate the pay process. They contain the data about employee attendance, allowance, charges, tax and other deductions. Based on this data, the pay of each employee is calculated.

Benefit administration systems: They manage and check the participation of employee in benefit programs such as insurance and profit sharing.

Online recruiting systems: They provide details of available vacancies and potential applicants. They allow applicants to register and submit their resume online.

Work time: Such systems gather information about time and work related efforts. They provide cost and efficiency analysis of the workforce.

Training systems: These systems manage and track efforts of employees in training and skill development.

Employee self-service systems: They allow employees to gather information about their attendance record and HR related data.

In [1], computer vision based real time system is proposed for modeling and recognizing human behavior in a visual surveillance task. This approach compared HMMs and CHMMs for behavior and interaction modeling. Their result shows that CHMM model is much more efficient and it works accurately. In [2] human actions are identified by tracking head movement of the actor in successive frames of videos. In this research work 12 different actions were recognized. Ayers and shah proposed system that recognizes activities of a person in a room using state machine model [3]. In [4] single person human activities are described. First step is image feature selection and extraction. Then properties of moving regions are computed and then scenarios are recognized from these properties using Bayesian networks. Leden [5] proposed a solution for human activity recognition that is based on Bayesian network and Fourier transform. Huang and Luo [6] proposed a method for video analysis and interpretation. First step of this method is to segment video clip. This is followed by semantic feature extraction. A mapping from low level features to high level semantic concept is required to acquire semantics of video clips. For this purpose dynamic Bayesian network is employed. Oliver and Horvitz [7] used hidden Markov model for sensing, learning, inferring and representing human activities. Tan and De Silva [8] recognized simple human activities by combining Neural Nets with HMM. Niu and Abdel-Mottaleb [9] proposed view invariant approach for activity recognition using a set of HMM where one model represents the activity from one view. [10] provides a solution for human activity recognition by using the statistics compiled on the basis of point trajectories. In [11], a framework for activity recognition is described in the form of multi-layered finite state machines. Low level layers involve spatio temporal detections and higher level layer recognizes activities and flag them as normal or unusual. Design of a classifier and data modeling for activity recognition are discussed in [12].

Thi V. Duong, Hung H. Bui, Dinh Q. Phung and Svetha Venkateshl in [13] described a system that recognizes human activities using Hidden Semi Markov model HSMM. In this case activities are recognized using two layers. Bottom layer recognize the atomic actions and the higher layer describes the sequence of high level activity where each activity is composed of atomic actions. Nascimento et al[14] proposed a system for human activity recognition using a bank of switched dynamical models. Space-dependent Markov chain governs the model switching. Neil Robertson and Ian Reid [15] developed a system for human behavior recognition by modeling it as stochastic sequence of actions. Higher level layers use Bayesian networks and belief propagation and low level uses sampling from learned database of actions.

M. S. Ryoo and J. K. Aggarwal [16] proposed a method for representing and recognizing human actions and activities that is based on context free grammar. In this approach human activities are divided into three groups: atomic actions, composite actions and interaction. Pose information is collected for recognizing atomic actions which are further used to recognize interactions. In [17] human activities such as walking, running and other are recognized using optical flow. This approach uses Burt-Adelson Pyramid approach for feature selection. [18] provides an overview of various methods of human activity recognition. It analyzes these methods and categorizes them from the best in business. Pavan Turaga and Rama Chellapa provided a comprehensive survey on approaches for activity and action recognition [19]. They showed that approaches for activity recognition fall into three basic categories-non parametric, volumetric and parametric. Zhou et al [20] proposed a system for automatic video analysis. Their approach includes silhouette extraction, human detection and tracking. After that, features are extracted on the basis of location and motion. Finally, actions are recognized using decisions tree. Meghna Singh, Anup Basu, and Mrinal Kr. Mandal [21] bring forth an algorithm for human activity recognition that is based on silhouette extraction. After background subtraction they performed silhouette extraction and then from silhouette contours, directional vectors are extracted for activity recognition. Joshua Candamo et al [22] also surveyed human behavior recognition algorithms. They presented comprehensive overview on human behavior-recognition methods for transit surveillance.

1.2 Application of Activity Recognition

Application of activity recognition algorithms is in following areas :-

Security and Surveillance: Security and surveillance systems use a network of video cameras monitored by a human operator. Manual analysis of video is labor intensive and prone to errors. Hence, security agencies are seeking vision-based solutions to these tasks, which can replace or assist a human operator. Examples are Loitering detection, crowd counting, intrusion or trespassing detection, moving in wrong direction and tailgating detection [19].

Behavioral Biometrics: Biometrics involves study of approaches and algorithms for uniquely recognizing humans based on physical or behavioral cues. Examples are crowd behavior analysis, person vehicle interaction and meeting behavior analysis [19].

Content Based Video Analysis: This involves analyzing the content of videos to understand consumer behavior [19].

Animation and Synthesis: The gaming and animation industry rely on synthesizing realistic humans and human motion [19].

2. Proposed Algorithm

The objective of this research work is to recognize/monitor the activities of employee by analyzing the video streams through video analytics algorithms. Our activity recognition algorithm consists of steps that are shown in Fig. 1.

2.1 Preprocessing

Static objects in an image are called background. Various methods have been proposed for extracting the foreground from a relative stationary background. Background subtraction and optical flow are most popular techniques used for foreground extraction.

2.2 Human Pose Estimation and Action Recognition

Human poses are described in terms of angles of joints of

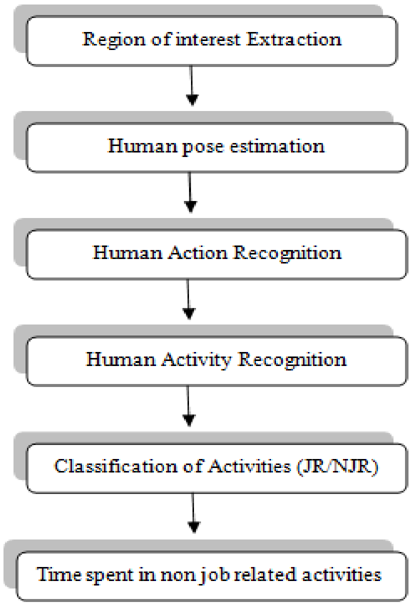


Fig. 1 Proposed Algorithm

2.2 Human Pose Estimation and Action Recognition

Human poses are described in terms of angles of joints of human body (arm, elbow, shoulder, neck).If both hands are on front and arm is making an "L" shape then this is an indication of typing. After that various statistics of a certain poses are computed.

2.3 High Level Activity Recognition

High level activity recognition is performed using logic programming. If we have a list of actions detected then we can find relations between them to recognize the activity. For example, consider the following actions: ‘picking the book’, ‘holding the book’, ‘unlock the cupboard’, ‘place the book inside’, ‘close the door of cupboard’. All of these actions combine together in a specific sequence to generate the activity of locking the book.

2.4 Classification of Activities

Once our algorithm detects a particular activity, it then classifies it into two groups: job related activity and non-job related activity. The decision about putting any job in one of the above two groups depends on the management.

3. Experimentation and Results

3.1 Results for activity recognition

Preprocessing

Video is obtained from camera and then it is converted into avi format and is shown in Fig. 2(a). After this, the whole video is segmented into frames and all images constituting video are converted into gray scale.

Background Subtraction

Background subtraction is performed after converting the complete video sequence into frames. This gives those regions of image that change in successive frames. In later stages, all the operations are performed on these background subtracted blobs for recognition purpose. Result of background subtraction as shown in Fig. 2(b).

Thresh-hold Image

A gray scale image is obtained after background subtraction. This gray level detail is not needed for recognition purpose. Therefore, the result of background subtraction is converted into binary image as shown in Fig. 2(c).

Edge Detection

There are some very small blobs produced as a result of background subtraction. In this step, edges are drawn around each blob and in the next step all edges are dilated. This step is required for obtaining all solid blobs that contribute in obtaining the features that uniquely identify a certain pose. A boundary is drawn around all the blobs in image as a result of edge detection step as shown in Fig. 2(d).

Image Dilation

In the previous stage of our approach, edges are obtained around blobs. These edges are dilated in order to retain the useful information from loss. Result of this step as shown in Fig. 2(e).

Filling Image

All the regions that are obtained after edge detection and dilation are filled to get solid blobs as shown in Fig. 2(f).

Removing Un-Necessary Blobs

After dilation there are still some very small portions that do not aid in recognition and it is required to remove this unnecessary detail as shown in Fig. 2(g).

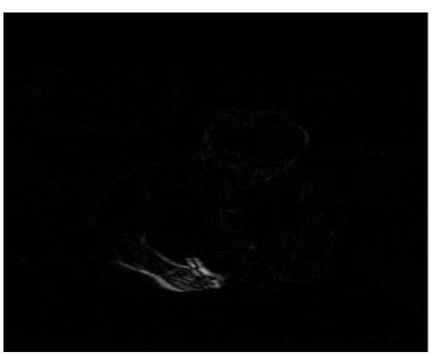
Recognizing Action and Activity

By applying all these operations, several templates are obtained for training the system for each action. When above refined images are obtained then different statistics are calculated for each template are obtained for example area perimeter, height to width ratio. Incoming frames are recognized on the basis of these values. This is good for action recognition but for activity recognition system needs to check the previous state for decision making. Result of activity recognition is shown in Fig. 2(h).

Report Generation

There is a unique timer set for each of the activity in activity recognition process. Timer for each particular activity increments on its recognition in video sequence. If the system cannot recognize a certain activity, then timer for unrecognized activity increments. After that total time elapsed in each of the activity is written to a text file as shown in Fig. 3. Then the percentage of the time consumed in job-related activity and non-job related activity is displayed in the form of pie chart as shown in Fig. 4.

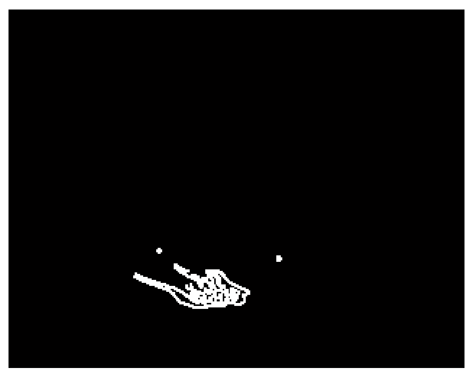
(a)



(b)

(c)

(d)



(e)

(f)

(g)

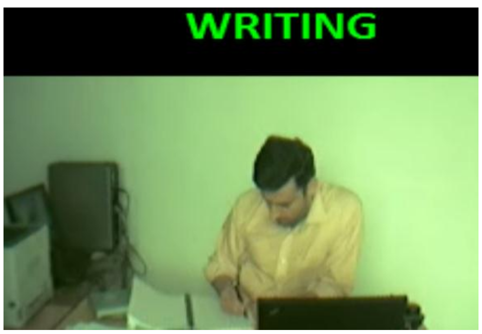
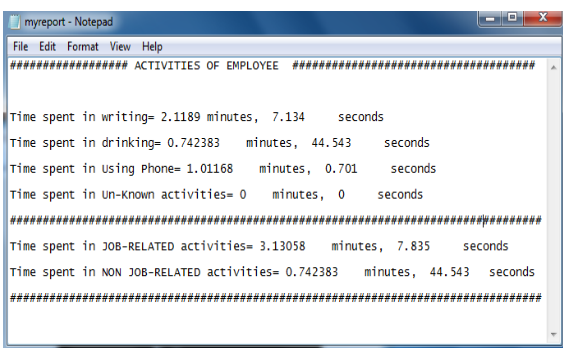


Fig. 2 Illustration of Pose estimation: (a) Original Image from video, (b) Background subtraction, (d) Thresh-holding, (c) Edge Extraction, (d) Dilation, (e) Filling the dilated version, (f) Removing un-necessary blobs (f) Pose estimation from image statistics and activity recognition.


myreport - Notepad
File Edit Format View Help

```
################# ACTIVITIES OF EMPLOYEE ###################################

Time spent in writing= 2.1189 minutes, 7.134 seconds
Time spent in drinking= 0.742383 minutes, 44.543 seconds
Time spent in Using Phone= 1.01168 minutes, 0.701 seconds
Time spent in Un-Known activities= 0 minutes, 0 seconds
###############################################################################
Time spent in JOB-RELATED activities= 3.13058 minutes, 7.835 seconds
Time spent in NON JOB-RELATED activities= 0.742383 minutes, 44.543 seconds
###############################################################################
```

Fig. 3 Activity recognition and classification report

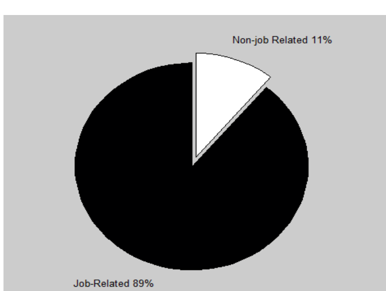


Fig. 4 Pie Chart showing the performance of employee

3.2 Results for Evaluation of Activity Recognition Algorithms

Recognition Using K-Nearest Neighbor (KNNR)

In k-NNR a given instance is compared with all other instances in defined neighbors and is assigned to one of neighbor classes [23]. This assignment is performed on the basis of probability of each class in defined neighborhood. This algorithm is evaluated in [8] with different values for k and results shows that this algorithm gives the best result when k is set to 5 as shown in Fig. 5.

Recognition using the HMM

Hidden Markov model is a probabilistic model used for activity recognition. In this model output states that are x1, x2, x3,….,xn are known. These are called observable states. And the goal is to find hidden states y1, y2, y3,…..,yn [24]. Simple activity recognition can be done using HMM. In HMM current observable state xt is only variable at yt depends on the previous hidden variable yt-1. Several experiments show that training can enhance the performance of HMM. HMM gives the best result when number of states is set to 3 [8]. This is demonstrated in Fig. 6.

Recognition using Context free Grammar

In [16], a system is proposed that use Bayesian networks for recognition of human pose at three levels for head, upper and lower portions of the body. After that predicate logic is used for finding the activity performed. Results are shown in Table 1.

Recognition using Coupled Hidden Markov Model (CHMM)

In [1], a system is described for recognizing human interaction between people. In this method Bayesian approach is used for both bottom and top layers. HMM and CHMM are also compared on the basis of performance and the results are shown in Table 2. In this table, 'No inter' shows the accuracy of HMM and CHMM when there is no interaction. 'Inter 1'-'Inter 5' shows accuracy at classifying different interactions.

Recognition using Emerging Patterns

Emerging pattern is a feature vector that shows the difference among various classes [24]. Suppose there are certain number of activity classes and instances. Each instance has a set of attributes. These attributes are basis for putting different activity instances into certain class. For computing these attributes two things are required.

These are support and growth rate as shown in Equation (1) and Equation (2) respectively [25].

$$Supp(X) = \frac{\text{the number of instances containing } X \text{ in } D}{\text{the number of instances in } D} \quad (1)$$

where, X is an attribute and D is Class or dataset
The value of Supp(X) shows whether X can be a candidate for unique identification or not.
Growth rate for attribute from D1 to D is given by:

$$GrowthRate(X) \begin{cases} 0 & \text{if } \mathrm{Supp}_1(X) = 0 \text{ and } \mathrm{Supp}_2(X) = 0 \\ \infty & \text{if } \mathrm{Supp}_1(X) = 0 \text{ and } \mathrm{Supp}_2(X) > 0 \\ \dfrac{\mathrm{Supp}_2(X)}{\mathrm{Supp}_1(X)} & \text{otherwise} \end{cases} \quad (2)$$

Supp1(X) shows support for attribute X in class D1
Supp2(X) shows support for attribute X in class D2

Recognition Using Conditional Random Fields

In many cases, activities are not performed in a specific order. Recognition using CRF is quite suitable in such cases. This approach checks only the conditional probability and is more flexible then HMM. A probability factor is attached to each transition.
The probability factor p(Y|X) is computed using Equation (3) [26].

$$p(Y \mid X) = \frac{1}{Z(X)} \exp\left(\sum_i \lambda_i \sum_{t=1}^{n} f_i(y_{t-1}, y_t, X, t) \right) \quad (3)$$

where, Z(X) is normalizing factor, $f_i(y_{t-1}, y_t, X, t) = r(y_{t-1}, y_t)$, λ_i is probability factor. Table 3 shows the comparison of HMM, CRF and EP.

4. Conclusion and Future Works

Detecting multiple activities in real-time video is difficult task that is currently performed by assigning multiple analysts to simultaneously watch the video stream. If any one of them found some abnormal behavior then the immediate authorities are informed for further action. Such manual analysis of video is difficult because it requires much labor and still susceptible to errors.

This employee performance evaluation system facilitates the job of human resource managers. It recognizes the activities of employee and categorizes them into two groups i.e. job related activities and non-job related activities. Finally, a report is generated along with a graph showing time spent in job related or non-job related activities. This relieves manager of time-consuming task of keeping an eye over video from surveillance camera all day long. We have also evaluated the performance of various activity recognition algorithms. It is found that HMMs are suitable for simple and recursive activities. CRFs are suitable when activities are not performed in a specific order. CHMMs are suitable for recognizing activities that involve more than one actor. Emerging patterns perform much better than both CRF and HMMs. CFG also provides good level of accuracy but it needs strong and consistent logic for language construction.
The proposed architecture has number of possible extensions. The use of multiple cameras can make the system even more robust. Similarly, the system can be enhanced to operate over larger workspaces such as point of sales, workshop, industry etc. by incorporating human tracking across multiple cameras. In the next step, activities other than writing, using phone and drinking can be added according to the activities performed at specific work place. Recognizing interleaved and concurrent activities is a very important step towards automatic performance monitoring. Intention reasoning can be an important step in the future research for inferring what is going to happen next. Most important use of this step would be in security applications.
Alternate modalities can be added to improve the performance of the system such as audio-video based tracking. This will help to draw a system parallel with human intelligence by making use of artificial intelligence.

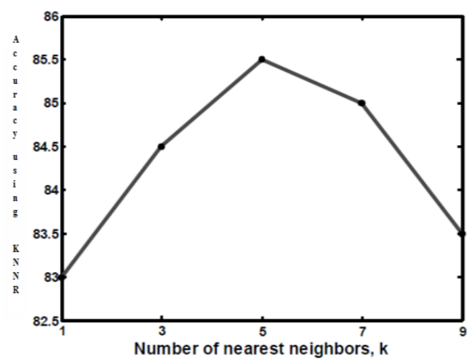


Fig. 5 Performance of KNNR [8]

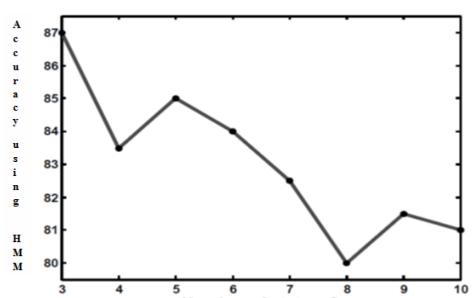


Fig. 6 Performance of HMM [8]

Interaction	Total	Correct	Accuracy
Approach	12	12	1.000
Depart	12	12	1.000
Point	12	11	0.917
Shake hands	12	11	0.917
Hug	12	10	0.833
Punch	12	11	0.917
Kick	12	10	0.833
Push	12	11	0.917
Total	96	88	0.917

Table 1: Performance of CFG [16]

Accuracy on Synthetic Data (%)		
	HMMs	CHMMs
No inter	68.7	90.0
Inter1	87.5	100
Inter2	85.4	100
Inter3	91.6	100
Inter4	77	100
Inter5	97.9	100

Table 2: Performance of HMM and CHMM [1]

	HMM	Linear Chain CRF	Emerging Pattern
Concurrent & Interleaved Activity	Not recognized	Not recognized	Recognized
Learning Method for Labeling	Supervised	Supervised	Partially unsupervised
Scalability	Change of HMM graph required	Change of CRF graph required	EP mining required

Table 3: Comparison of HMM, CRF and EP [24]

References

[1] Nuria M. Oliver, Barbara Rosario, and Alex P. Pentland, “A Bayesian Computer Vision System for Modeling Human Interactions”, IEEE transactions on pattern analysis and machine intelligence, vol. 22, no. 8, august 2000

[2] Anant Madabhushi and J. K Aggarwal, “Using Head Movement to Recognize Activity”, 15th International Conference on Pattern Recognition (ICPR'00) - Volume 4, 2000

[3] Douglas Ayers, Mubarak Shah, “Monitoring human behavior from video taken in an office environment”,2000

[4] Somboon Hongeng, Francois Br´emond and Ramakant Nevatia, “Representation and Optimal Recognition of Human Activities”,2000 IEEE Computer Vision and Pattern Recognition-Volume 1, 2000

[5] Rocio Diu de Ledn, Luis Enrique Sucar, “Continuous Activity Recognition with Missing Data”, ICPR '02 Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02) Volume 1 - Volume 1,2002

[6] Jenq-Neng Hwang, Ying Luo, “Automatic Object-Based Video Analysis And Interpretation: A Step Toward Systematic Video Understanding”, Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference, 2002

[7] Nuria Oliver, Eric Horvitz , Garg, A, “Layered Representations for Human Activity Recognition”, Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference, 2003

[8] Henry C. C. Tan and Liyanage C. De Silva, “Human Activity Recognition by Head Movement using Elman Network and Neuro-Markovian Hybrids”, Department of Electrical & Computer Engineering, National University of Singapore,Institute of Information Sciences and Technology, Massey University, New Zealand

[9] Feng Niu, Mohamed Abdel-Mottaleb, “View-Invariant Human Activity Recognition Based on Shape and Motion Features”, Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE’04), 2004

[10] Wei Niu, Jiao Long, Dan Han, and Yuan-Fang Wang, “Human Activity Detection and Recognition for Video Surveillance”, Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference, 2005

[11] Dhruv Mahajan, Nipun Kwatra Sumit Jain Prem Kalra Subhashis Banerjee, “A Framework for Activity Recognition and Detection of Unusual Activities”, Department of Computer Science and Engineering Indian Institute of Technology New Delhi 110016

[12] Pedro Canotilho Ribeiro & Jos´e Santos-Victor, “Human Activity Recognition from Video: modeling, feature selection and classification architecture”, 2005

[13] Thi V. Duong, Hung H. Bui, Dinh Q. Phung, Svetha Venkatesh, “Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model”, Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), 2005

[14] Jacinto C. Nascimento , Mario A. T. Figueiredo,Jorge S. Marques, “Recognition of human activities using space dependent switched dynamical models”, Image Processing, 2005. ICIP 2005. IEEE International Conference, 2006

[15] Neil Robertson , Ian Reid , “A General Method for Human Activity Recognition in Video”, Computer Vision and Image Understanding - Computer Vision and Image Understanding - Special issue on modeling people: Vision-based understanding of a person's shape, appearance, movement, and behavior Volume 104 Issue 2, November 2006

[16] M. S. Ryoo and J. K. Aggarwal,”Recognition of Composite Human Activities through Context-Free Grammar based Representation”, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference, 2006

[17] Toru Nakata, “Recognizing Human Activities in Video by Multi-resolutional Optical Flows”, Digital Human Research Center”.

[18] Md. Atiqur Rahman Ahad, J.K. Tan, H.S. Kim and S. Ishikawa , "Human Activity Recognition: Various Paradigms", Control, Automation and Systems, 2008. ICCAS 2008. International Conference, 2008

[19] Pavan Turaga, Rama Chellappa, "Machine Recognition of Human Activities: A Survey" , IEEE Transactions On Circuits And Systems For Video Technology, Vol. 18, No. 11, November 2008 .

[20] Zhongna Zhou, Xi Chen, Yu-Chia Chung, Zhihai He, Tony X. Han, andJames M. Keller, "Activity Analysis, Summarization, and Visualization for Indoor Human Activity Monitoring", IEEE transactions on circuits and systems for video technology, vol. 18, no. 11, November 2008

[21] Meghna Singh, Anup Basu, and Mrinal Kr. Mandal, "Human Activity Recognition Based on Silhouette Directionality", IEEE transactions on circuits and systems for video technology, vol. 18, no. 9, september 2008

[22] Andreas Bulling,, Jamie A. Ward, Hans Gellersen, and Gerhard Troster, "Eye Movement Analysis for Activity Recognition Using Electrooculography", IEEE transactions on pattern analysis and machine intelligence, MARCH 2010

[23] wikipedia(website)

[24] M Eunju Kim, Sumi Helal, and Diane Coo, "Human Activity Recognition and Pattern Discovery", IEEE Pervasice Comput. 2010

[25] T. Gu, Z. Wu, X. Tao, H. K. Pung, and J. Lu, epSICAR: An Emerging Patterns based approach to sequential, interleaved and Concurrent Activity Recognition. Proceedings of the IEEE International Conference on Pervasive Computing and Communications, 2009.

[26] C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning," L. Getoor and B. Tas- kar, eds., Introduction to Statistical Relational Learning, MIT Press, 2006

Engr. Mehreen Mumtaz Received her MS degree in Software Engineering from University of Engineering and Technology Taxila, Pakistan in August, 2011. She did her BSc Software Engineering from the same university in September, 2008. She served as a Lab Engineer in Software Engineering Department at UET Taxila from December 2008 to September 2011. Her area of interest are Digital Image processing and Computer Vision.

Dr. Hafiz Adnan Habib
Received his PhD in Electrical Engineering from University of Engineering & Technology Taxila Pakistan 2007.His area of interest is Gesture Recognition Solutions for HCI/HMI in Smart Environments. He is also working on Gesture Recognition for Monitoring, Safety and Video Surveillance Algorithms for Activity Recognition. Currently he is working as Associate Professor in Computer Engineering Department University of engineering & Technology Taxila, Pakistan.