A Novel Feature Extraction Technique for Facial Expression Recognition

***Mohammad Shahidul Islam[1], Surapong Auwatanamongkol[2]**

**[1] Department of Computer Science, School of Applied Statistics,
National Institute of Development Administration,
Bangkok, 10240, Thailand**

**[2] Department of Computer Science, School of Applied Statistics,
National Institute of Development Administration,
Bangkok, 10240, Thailand**

Abstract

This paper presents a new technique to extract the light invariant local feature for facial expression recognition. It is not only robust to monotonic gray-scale changes caused by light variations but also very simple to perform which makes it possible for analyzing images in challenging real-time settings. The local feature for a pixel is computed by finding the direction of the neighboring of the pixel with the particular rank in term of its gray scale value among all the neighboring pixels. When eight neighboring pixels are considered, the direction of the neighboring pixel with the second minima of the gray scale intensity can yield the best performance for the facial expression recognition in our experiment. The facial expression classification in the experiment was performed using a support vector machine on CK+ dataset The average recognition rate achieved is 90.1 ± 3.8%, which is better than other previous local feature based methods for facial expression analysis. The experimental results do show that the proposed feature extraction technique is fast, accurate and efficient for facial expression recognition.

Keywords: *Emotion Recognition, Facial Expression Recognition, Image Processing, Local Descriptor, Pattern Recognition.*

1. Introduction

Facial Expression plays an important role in human-to-human interaction, allowing people to express themselves beyond the verbal world and understand each other from various modes. Some expressions incite human actions, and others fertilize the meaning of human communication. Human-centered interfaces must have the ability to detect shades of and changes in the user's behavior and to start interactions based on this information rather than simply responding to the user's commands. Facial expression recognition is a challenging problem in computer vision. Due to its potential important applications, it attracts much attention of the researchers in the past few years (Z. Zeng *et al*, 2009). Appearance-based methods have been heavily employed in this domain with great success. Popular methods are Gabor filters, local binary patterns (LBP) descriptors, Haar wavelets and subspace learning methods. Facial expression recognition process is a part of facial image analysis. A. Mehrabian (1968) mentioned in his paper that the verbal part of a message contributes only 7% of its meaning as a whole, the vocal part contributes 38% while facial movement and the expression gives 55% to the effect of that message, see Fig. 1. This means that the facial part does the major contribution in human communication. There are seven basic types of facial expressions. They are contempt, fear, sadness, disgust, anger, surprise and happiness. From the review of papers on facial expression, it is clear that most of the facial expression recognition systems (FERS) were based on the Facial Action Coding System (FACS), Y.L. Tian *et al.* (2001), Y. Tong *et al.* (2007), M. Pantic *et al.* (2000). In this system, the changes in the facial expression are described with FACS in terms of 44 different

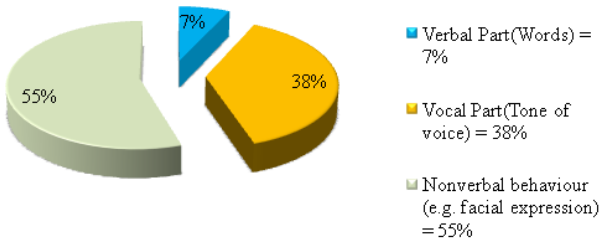

Fig. 1: 7%-38%-55% rule by A. Mehrabian (1968)

action units (AUs), each of which is related to the facial muscle movements. 44 AUs can give up-to 7000 different combinations, with wide variations due to age, size and ethnicity. M. Pantic *et al.* (2000) gave detail survey on facial expression recognition in their paper. Most of the research works on facial expression recognition (FER) are grounded on still images. The psychological experiments

by J.N. Bassili (1979) have proposed that facial expressions are more precisely recognized from video than single static image. I. Kotsia *et al.* (2007) applied facial wire frame model and a Support Vector Machine (SVM) for classification. Y. Zhang *et al.* (2005) proposed IR (Infra Red) illuminated camera for facial feature detection, tracking and recognized the facial expressions using Dynamic Bayesian networks (DBNs). Y.L. Tian *et al.* (2001) proposed multi state face component model of AUs and neural network for classification. M. Yeasin *et al.* (2007) created discrete hidden Markov models (DHMMs) to recognize the facial expressions. K. Anderson *et al.* (2006) used the multichannel gradient model (MCGM) to determine facial optical flow in videos. The motion signatures achieved are then classified using Support Vector Machines. I. Cohen *et al.* (2003) employed Naive–Bayes classifiers and hidden Markov models (HMMs) together to recognize human facial expressions from video sequences. M. Pantic *et al.* (2006) applied face-profile-contour tracking and rule-based reasoning to recognize 20 AUs taking place alone or in a combination in nearly left-profile-view face image sequences and they achieved 84.9% accuracy rate. T. Ahonen *et al.* (2006) proposed a new facial representation strategy for still images based on Local Binary Pattern (LBP). The basic idea for developing the LBP operator was that two-dimensional surface textures can be identified by two complementary measures: 2D local spatial patterns and the gray scale difference. The original LBP operator by T. Ojala *et al.* (1996), labels for

a(1)	a(2)	a(3)
a(8)	e	a(4)
a(7)	a(6)	a(5)

Fig. 2: Local 3x3 pixels Image region

the image pixels by comparing the 3 x 3 neighborhood of each pixel with the center pixel value and transforming the result as a binary number.

$$LBP = \sum_{i=1}^{P} 2^{i-1} f(a(i) - e) \qquad (1)$$

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \qquad (2)$$

Where (e) denotes the gray value of the center pixel, a(i) is the gray value of its neighbors, P stands for the number of neighbors, see Fig. 2. Fig. 3 shows an example of

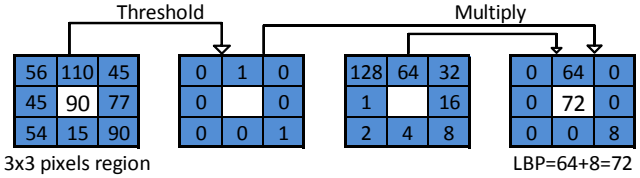

Fig. 3: Example of obtaining LBP from a 3x3 local region.

obtaining an LBP from a given 3×3 pattern. The histogram of these patterns for a local block of an image represents a local feature for the block. The histograms for all blocks can be concatenated to represent the feature vector for the image. G. Zhao *et al.* (2007) applied facial dynamic texture data in conjunction with Local Binary Pattern on the Three Orthogonal Planes (LBP-TOP) and Volume Local Binary patterns (VLBP) to combine motion and appearance. In her earlier work (G. Zhao *et al.*, 2004), she tested with the two-dimensional (2-D) discrete cosine transform (DCT) over the entire face image but got less accuracy on the facial expression recognition. The FACS approaches involve more complexity in facial feature detection and extraction procedures while the appearance-based approaches using local features such as LBP are less complex but still need to be improved to get higher recognition rates. Hence, this paper proposes an alternative local feature extraction technique that would be simple and more effective for facial expression recognition.

The rest of the paper is organized to explain the proposed methodology in section 2, results and analysis in section 3, and conclusion in section 4.

2. Proposed Methodology

2.1 Local Minima (LM)

The proposed method computes the local feature for a pixel from the gray scale value of its neighboring pixels. From a 3x3 local pattern, shown in Fig. 2, the center pixel of the pattern is surrounded by 8 neighboring pixels in 8 possible directions. The directions are denoted by 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°. The direction of the neighboring pixel with the minimum, the second minimum and so on for the gray scale values can be considered as the local feature for the given pixel. To identify the neighboring pixels with the first minima, second one and so on, the gray scale color values of all the eight

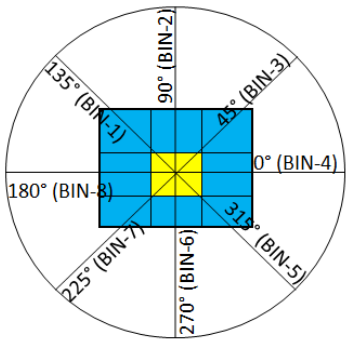

Fig. 4: 8 possible BINS denoted as 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°

neighboring pixels can be sorted in ascending order. If there are several of the pixels with the same gray color value, the positions of the pixels starting from the northwest one in clockwise direction can be considered to break the ties. The direction would represent the changing direction of the gray scale color values at the particular center pixel. Thus, eight possible bins are needed to build the histogram on the numbers of pixels in a block for the possible 8 directions as shown in Fig. 4**.** The histograms for all blocks for an image can then be concatenated to form the feature vector for the whole image. Notice that the direction is insensitive to light changes since the light changes would change the gray scale color values of all the pixels by nearly same amount but not the direction of the minima for each of the pixels.

2.2 Experimental Setup

The experiments for the facial expression recognition include three distinguished phases. i.e. facial feature extraction, SVM training and facial expression determination. The Extended Cohn-Kanade Dataset (CK+) (P.Lucey *et al.*, 2010) is used for both training and testing images. There are 326 peak facial expressions of 123

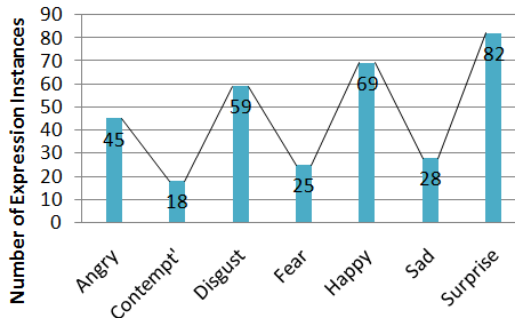

Fig. 5: CK+ Dataset, 7 expressions and numbers of instances of each expression

subjects. Seven emotion categories are in this dataset. They are 'Anger', 'Contempt', 'Disgust', 'Fear', 'Happy', 'Sadness' and 'Surprise'. No subject with the same emotion has been collected more than once. The data distribution of the dataset is shown in Fig. 5**.** This is the most common dataset used in FER (Facial Expression recognition). All the images are posed in this dataset. Facial feature extraction phase is illustrated in Fig. 6, which includes detecting face, masking the face, dividing the cropped face into equal blocks, calculating feature histogram for each block and concatenating all histograms to build feature vector.

Face detection is done using **fdlibmex**, free code is available for Matlab. The library consists of single mex file with a single function that takes an image as input and returns the frontal face. It is then resized to 180x180 resolutions and masked using a round shape, outside which all the pixels are removed from the consideration. In experiment, the 180x180-size face is equally divided into 9x9=81 blocks of 20x20 resolutions each. Feature is extracted from each block using the proposed method, concatenating histograms of all the blocks into a unique feature vector. Therefore, the length of the feature vector is 8x9x9=648. In the training phase, LIBSVM, by C.C. Chang *et al.* (2011) is used to train a multiclass Support Vector Machine to classify the facial expression for an image. The 10-fold cross validation was used to evaluate the performance of the classifier when the proposed local feature was used. Each expression instances are divided into equal size 10 folds. Ten rounds of evaluations were conducted. For each round, nine alternative folds (90% for each expression) are used for training and the rest one fold are used for testing. The kernel parameters are set to: **(-s 0 -t 1 -c 100 -g 0.00125 -b 1),** where s=0 for SVM type C-Svc, t=1 for polynomial kernel function, c=100 is the cost of SVM, g=0.00125 is the value of 1/ (length of feature vector), b=1 for probability estimation. The kernel

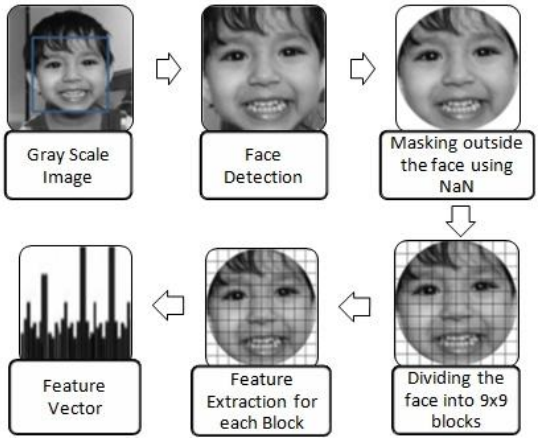

Fig. 6: Facial Feature Extraction

parameters are tuned so that it can produce optimal results during this phase.

3. Experimental Results and Analysis

Table 1 shows the achieved classification accuracy rates when the directions of the 1st to 8th local minima are individually considered as the local feature. The 2nd minima (LM-2) gives the peak accuracy of 90.1%±3.8. This is the average accuracy of 10-fold cross validation. The highest accuracy achieved by any one fold from the tenfold is 93.9% and the lowest is 87.1%. We also tried the

same experiment with different block sizes as shown in Table 2. However, the block size of 15x15 pixels gives the highest 90.33% accuracy rate but there is a penalty in feature vector length.

Table 1: Classification Accuracy of the 1st to the 8th Local Minima

Local Minima	Classification Accuracy
1st (LM-1)	88.8%
2nd (LM-2)	**90.1%**
3rd (LM-3)	89.8%
4th (LM-4)	88.9%
5th (LM-5)	88.9%
6th (LM-6)	89.2%
7th (LM-7)	89.5%
8th (LM-8)	86.4%

Table 2: Classification Accuracy Vs Block Dimension.

Face Dimension (Pixels)	Number of Blocks	Block Dimension (pixels)	Classification Accuracy (%)	Feature Vector Length
180x180	6x6	30x30	88.28	288
180x180	**9x9**	**20x20**	**90.1**	**648**
180x180	10x10	18x18	88.8	800
180x180	**12x12**	**15x15**	**90.33**	**1152**
180x180	15x15	12x12	88.25	1800
180x180	18x18	10x10	87.63	2592

Table 3: Confusion Matrix for LM-2

LM-2 (Local Second Minima) = 10-fold validation
Feature Extraction time for 326 Image = 96 Seconds
Average Classification Accuracy = 90.1 ± 3.8%
Kernel parameter: = (-s 0 -t 1 -c 100 -g 0.0015 -b 1)
Confusion Matrix:

C : Actual

prediction	Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise
Angry	38	1	2	0	0	4	0
Contempt	2	15	0	0	0	1	0
Disgust	2	0	54	2	1	0	0
Fear	1	1	0	21	1	0	1
Happy	0	0	0	0	69	0	0
Sad	4	1	1	1	0	20	1
Surprise	0	1	0	1	0	0	80

The confusion matrix using proposed feature extraction method of LM-2 is shown in Table 3. The feature extraction takes 96 seconds including 30 seconds for face detection and round masking (Preprocessing) for the 326 images, or 0.29 seconds per image. Table 4 shows comparisons of the individual expression accuracy and the average all 7 class expression accuracy achieved by the proposed method and the other recent methods using shape or combination of shape and texture information.

Table 4: Comparison of individual expression accuracy and the average accuracy (∑ (Accuracy of all 7 expressions/7)) of different methods. [S: shape based method, T: texture based method. S + T: both shape and texture based method. (CLM-Constrained Local Model, AAM-Active Appearance Model, An.= Anger, Co.= Contempt, Di.= Disgust, Fe.=Fear, Ha.=Happy, Sa.=Sad, Su.=Surprise, Avg.=Accuracy of all expressions/7)]

Authors	Method	T/S	An.	Co.	Di.	Fe.	Ha.	Sa.	Su	Avg.
P. Lucey *et al.* (2010)	AAM + SVM	S	35.0	25.0	68.4	21.7	98.4	4.0	100.0	50.3
	AAM + SVM	T	70.0	21.9	94.7	21.7	100.0	60.0	98.7	66.7
	AAM + SVM	T + S	75.0	84.4	94.7	65.2	100.0	68.0	96.0	83.3
S.W. Chew *et al.* (2011)	CLM + SVM	T	70.1	52.4	92.5	72.1	94.2	45.9	93.6	74.4
L.A. Jeni *et al.* (2012)	CLM + SVM (AU0 norm.)	S	73.3	72.2	89.8	68.0	95.7	50.0	94.0	77.6
	CLM + SVM (personal mean shape)	S	77.8	94.4	91.5	80.0	98.6	67.9	97.6	86.8
Proposed Method (LM-2)	**No Registration + SVM**	**T**	**84.4**	**83.3**	**91.5**	**84.0**	**100.0**	**71.4**	**97.6**	**87.0**

It should be noted that the results are not directly comparable due to different experimental setups, version differences of the CK (T. Kanade *et al.*, 2000) dataset with different emotion labels, preprocessing methods, the number of sequences used, and so on, but they still point out the discriminative power of each approach. It is clearly

mentioned by L.A. Jeni *et al.* (2012) that aligned faces can give an extra 5-10% increase in the facial expression recognition accuracy and leave-one -subject-out validation can increase the accuracy by 1-2% , (M.S. Bartlett *et al.*, 2003) and incorporation of adaboost algorithm can also

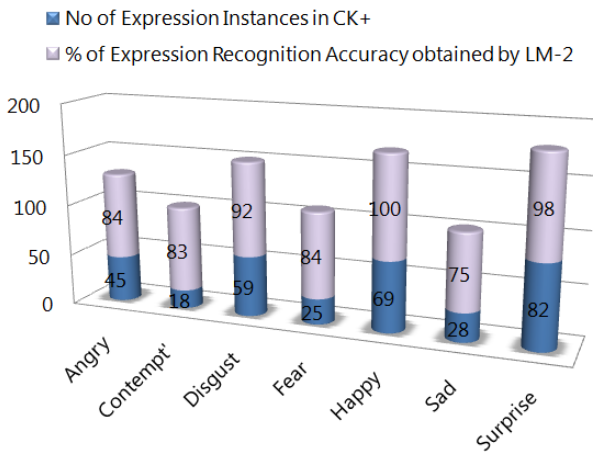

Fig. 7: Number of Instances Vs percentage of individual expression recognition Accuracy.

increase the accuracy by 1-2% on CK+ dataset. So overall an extra 7-12% accuracy can be obtained using proper alignment, increasing training data size and adding boosting algorithm along with the classifier.

A facial expression can be spontaneous or caused externally. In general, cases boundaries for spontaneous expressions are tough to determine. The dataset has only one peak expression for a particular subject. Some subjects do not contain all seven expressions. Training with multiple instances of the same expression and subject can increase accuracy. Fig. 7 clearly shows that in CK+ dataset number of 'Sad', 'Contempt' or 'Fear' instances are less in compare with the other expressions. Increasing these instances can increase accuracy like others.

4. Conclusion

A novel technique for facial feature extraction is proposed for facial expression recognition. It extracts from a gray scale image the direction of the neighboring pixel with local minima on the gray scale color value among those of the eight neighboring pixels. Eight possible Minima neighboring pixels can be considered as a local feature for a given pixel; however, the direction of the second minima yields the highest facial expression recognition rate in the experiment. Further techniques such as AdaBoost or SimpleBoost algorithms can be incorporated with the SVM classifier to increase the accuracy rate substantially.

References

[1] A. Mehrabian. "Communication without words." Psychology Today, 2, 4 (1968), 53-56.

[2] C.C. Chang and C.J. Lin." LIBSVM: a library for support vector machines". ACM Transactions on Intelligent Systems and Technology (2011).

[3] G. Zhao and M. Pietikainen." Dynamic texture recognition using local binary patterns with an application to facial expressions." IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)., 29, 6 (2007), 915–928.

[4] G. Zhao and M. Pietikainen. "Facial Expression Recognition Using Constructive Feed forward Neural Networks." IEEE Transactions on Systems, Man, and Cybernetics., 34, 3 (2004), 1588–1595.

[5] I. Cohen, N. Sebe, S. Garg, L. S. Chen and T. S. Huanga. Facial expression recognition from video sequences: temporal and static modelling. Computer Vision and Image Understanding, 91 (2003), 160-187.

[6] I. Kotsia & I. Pitas. Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines. IEEE Transaction on Image Processing, 16, 1 (Jan 2007).

[7] J.N. Bassili. Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Area of the Face. J.Personality and Social Psychology, 37 (1979), 2049-2059.

[8] K. Anderson and Peter W. McOwan. A Real-Time Automated System for the Recognition of Human Facial Expressions. IEEE Transactions on Systems, Man, and Cybernetics, 36, 1 (2006), 96-105.

[9] L. A. Jeni, András Lőrincz, Tamás Nagy, Zsolt Palotai, Judit Sebők, Zoltán Szabó & Dániel Takács. 3D shape estimation in video sequences provides high precision evaluation of facial expressions. Image and Vision Computing, 30, 10 (October 2012), 785-795.

[10] M. Pantic and Ioannis Patras. Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments From Face Profile Image Sequences. IEEE Transactions on Systems, Man, and Cybernetics, 36, 2 (2006), 433-449.

[11] M. Pantic and L. J. M. Rothkrantz. Automatic analysis of facial expressions: the state of the art. IEEE Trans. Pattern Analysis and Machine Intelligence., 22, 12 (2000), 1424–1445.

[12] M. Yeasin, B. Bullot and R. Sharma. Recognition of Facial Expressions and Measurement of Levels of Interest From Video. IEEE Trans. Multimedia, 8, 3 (2006), 500-508.

[13] M.S. Bartlett, G. Littlewort, I. Fasel & R. Movellan. Real Time Face Detection and Facial Expression Recognition: Development and Application to Human Computer Interaction. In Proc. CVPR Workshop Computer Vision and Pattern Recognition for HumanComputer Interaction(2003).

[14] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar & I. Matthews. The Extended Cohn-Kande Dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression. Paper presented at the Third IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010) (2010).

[15] S.W.Chew, P.Lucey, S. Lucey, J. Saragih, J.F. Cohn & S. Sridharan. Person-independent facial expression detection using Constrained Local Models. In 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), (2011), 915-920.
[16] T. Ahonen, A. Hadid and M. Pietikainen. Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Trans. Pattern Analysis and Machine Intelligence, 28, 12 (2006), 2037-2041.
[17] T. Kanade, J. F. (2000). Comprehensive database for facial expression analysis. Fourth IEEE International Conference on Automatic Face and Gesture Recognition.
[18] T. Ojala, M. Pietikäinen & T. Mäenpää. Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. Pattern Analysis and Machine Intelligence, 24, 7 (2002), 971-987.
[19] Y. L. Tian, T. Kanade and J. F. Cohn. Recognizing action units for facial expression analysis. IEEE Trans. Pattern Anal. Mach. Intell, 23, 2 (2001), 97-115.
[20] Y. Tong, W. Liao, and Q. Ji. Facial Action Unit Recognition by Exploiting Their Dynamic and Semantic Relationships. IEEE Trans. Pattern Anal. Mach. Intell, 29, 10 (2007), 1-17.
[21] Y. Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans. Pattern Anal. Mach. Intel, 27,5(2005),699-714.
[22] Z. Zeng, M. Pantic, G. Roisman & T. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31, 1 (2009), 39–58.

Mohammad Shahidul Islam received his B.Tech. degree in Computer Science and Technology from Indian Institute of Technology-Roorkee (I.I.T-R), Uttar Pradesh, INDIA in 2002, M.Sc. degree in Computer Science from American World University, London Campus, U.K in 2005 and M.Sc. in Mobile Computing and Communication from University of Greenwich, London, U.K in 2008. He is currently pursuing the Ph.D. degree in Computer Science & Information Systems at National Institute of Development Administration (NIDA), Bangkok, THAILAND. His field of research interest includes Image Processing, Pattern Recognition, wireless and mobile communication, Satellite Commutation and Computer Networking.

Surapong Auwatanamongkol received a B.Eng. (Electrical Engineering) from Chulalongkorn University, THAILAND, in 1978 and M.S.(Computer Science) from Georgia Institute of Technology, U.S.A. in 1982 and Ph.D.(Computer Science) from Southern Methodist University, U.S.A. in 1991. Currently, he is an Associate Professor in Computer Science at the School of Applied Statistics, National Institute of Development Administration (NIDA), Thailand. His research interests include Evolutionary Computation, Pattern Recognition, Image processing and Data Mining.