Feature and Decision Fusion based Facial Recognition in Challenging Environment

Md. Rabiul Islam
Assistant Professor
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh.

Md. Fayzur Rahman
Professor
Department of Electrical & Electronic Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh.

ABSTRACT
This paper introduces a face recognition system that contributes the feature and decision fusion in challenging environment. In this work, we investigate the proposed facial recognition system in typical office environments conditions. Though the traditional HMM based facial recognition system is very sensitive to the facial parameters variation, the proposed feature and decision fusion based face recognition is found to be stance and performs well for improving the robustness and naturalness of human-computer-interaction. At first appearance and shape based features are extracted using Active Appearance Model and Active Shape Model. The other task combines appearance and shape based features that have been used by the multiple Discrete Hidden Markov Model classifiers with likelihood ratio based score fusion and majority voting method. The performances of all these uni-modal and multi-modal system performance have been evaluated and compared with each other according to the VALID database.

General Terms
Face Recognition, Feature Fusion and Decision Fusion.

Keywords
Face Recognition, Feature and Decision Fusion, Facial Feature Extraction, Human Computer Interaction, Discrete Hidden Markov Model..

1. INTRODUCTION
In noisy environment, perfect solutions are often difficult to achieve for pattern recognition system [1]. A multiple classifier system is a powerful solution to these complex pattern recognition problems because it allows simultaneous use of arbitrary feature descriptors and classification procedures [2].

Several excellent survey papers on face recognition techniques are available with a wide variety of methods [3], [4], [5], [6] that covers early face recognition approaches. While humans quickly and easily recognize faces under variable situations or even after several years of separation, the problem of machine face recognition is still a highly challenging task in pattern recognition and computer vision [7], [8]. A face is inherently a 3D object illuminated by a variety of lighting sources from different directions and surrounded by arbitrary background objects. Therefore, the appearance of a face varies tremendously when projected onto a 2D image. Different pose angles cause significant changes in 2D appearance [9]. Rotation independent face recognition using optical neural network has also been developed [10], [11].

Various learning algorithm such as probabilistic decision-based neural network based face recognition/detection [12], neural network based face detection [13], Gabor Wavelets Transform and Extended Nearest Feature Space Classification [14], Karhunen-Loeve Procedure [15], ICA based face detection [16], support vector machine based approach [17], eigenfaces for recognition [18] etc. have been developed for the face recognition purpose. Learning systems in face recognition that employ hybrid strategies [19], [20], [21], [22] can potentially offer significant advantages over single-strategy systems.

In this work, appearance and shape based facial features fusion and likelihood ratio based and majority voting method based decision fusion has been proposed for face recognition. Facial image processing and the components of proposed facial recognition system are shown and design tradeoffs are focused on the following sections. The performances of each uni-modal and multi-modal systems performance are compared and shown in the results and performance analysis section.

2. FEATURE AND DECISION FUSION BASED FACE RECOGNITION SYSTEM
The paradigm of the proposed appearance and shape based feature fusion and decision fusion method are shown in figure 1. Principal Component Analysis (PCA) has been used to reduce the dimension of the facial feature vector. Log likelihood ratio based decision fusion is performed to combine the appearance and shape based HMM classifier output. Finally all of the classifiers output i.e. appearance based HMM classifier, shape based HMM classifier and combined appearance-shape based HMM classifier are simulated using majority voting method to get the overall face recognition result.

3. FEATURE EXTRACTION FROM THE FACIAL IMAGE
High quality digital camera has been used to capture the face image. After acquisition of face image, Stams [23] Active Appearance Model (ASM) has been used to detect the facial features. Then the binary image has been taken. The Region Of Interest (ROI) has been chosen according to the ROI selection algorithm [24, 25]. Lastly the background noise has been eliminated [26] and finally appearance based facial feature has been found. The procedure of the facial image pre-processing parts is shown in figure 2. To reduce the dimensionality of the facial feature vector, PCA has been used.

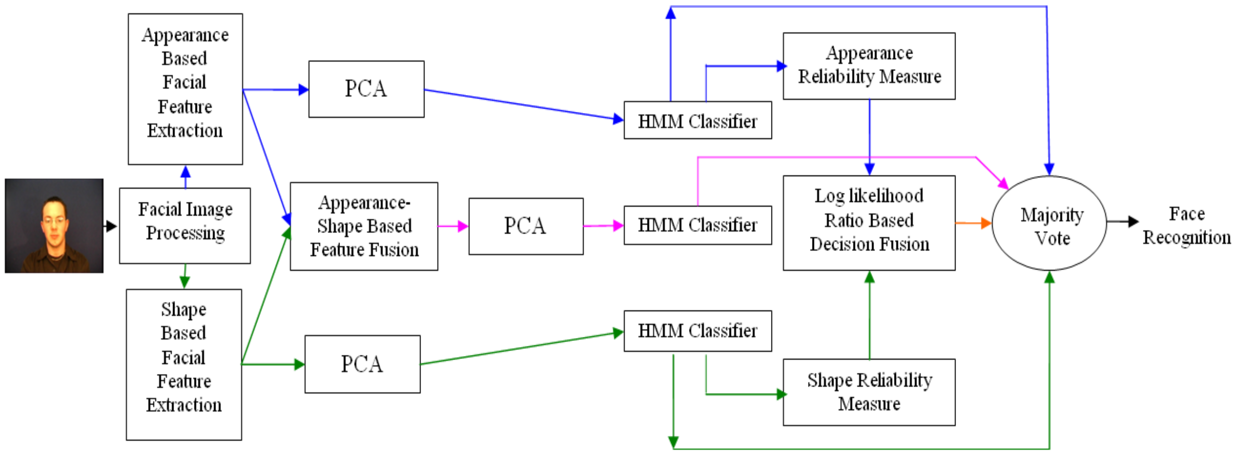

Fig 1: Block diagram of the proposed feature and decision fusion based face recognition system

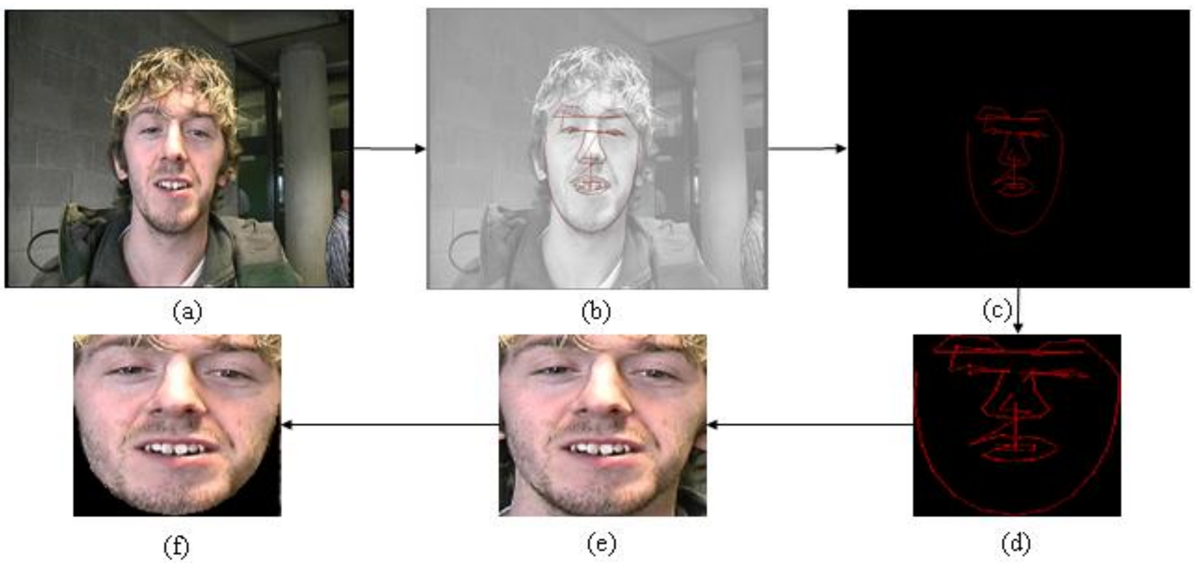

Fig 2: Facial image pre-processing for the proposed system (a) Original image (b) Output taken from Stams Active Appearance Model (c) Facial edges are extracted (d) Shape based features (e) Region Of Interest (ROI) selection with background noise (f) Appearance based facial features

4. LEARNING AND CLASSIFICATION MODEL OF THE FACE RECOGNITION SYSTEM

In training phase, for each face *k*, an erogodic DHMM (Discrete HMM), θ_k has been built [27, 28, 29, 30]. The model parameters (A, B, θ) have been estimated to optimize the likelihood of the training set observation vector for the k^{th} face by using Baum-Welch algorithm. The Baum-Welch re-estimation formula has been considered as follows [31]:

$$\bar{\Pi}_i = \gamma_1(i) \tag{1}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{2}$$

$$\bar{b}_j(\vec{k}) = \frac{\sum_{t=1(s,t,\vec{o}_t=\vec{v}_k)}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \tag{3}$$

where, $\xi_t(i,j) = \dfrac{\alpha_t(i)a_{ij}b_j(\bar{o}_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(\bar{o}_{t+1})\beta_{t+1}(j)}$ and

$$\gamma_t(i) = \sum_{j=1}^{N}\xi_t(i,j)$$

In the testing phase, for each unknown face to be recognized, the processing shown in figure 3 has been carried out. This procedure includes:

Measurement of the observation sequence, $O = \{o_1, o_2, ... o_n\}$, via a feature analysis of the speech corresponding to a face.
Transformation the continuous values of *O* into integer values.
Calculation of model likelihoods for all possible models, $P(O \mid \theta_k), 1 \le k \le K$.
Declaration of the face as *k** person whose model likelihood is highest – that is,

$$k* = \arg\max_{1\le k\le K}[P(O \mid \theta_k)] \quad (4)$$

In this proposed work the probability computation step has been performed using the Baum's Forward-Backward algorithm [31, 32].

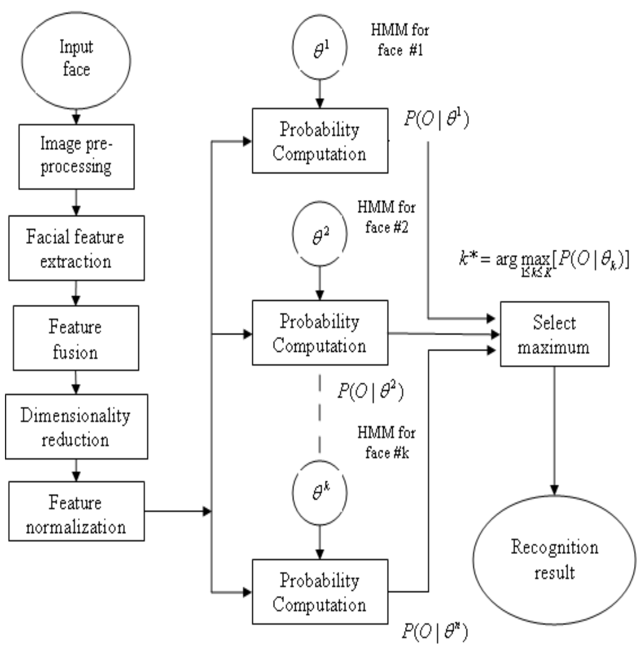

Fig 3: Block diagram of DHMM recognizer for face recognition

5. APPEARANCE AND SHAPE BASED LIKELIHOOD RATIO BASED SCORE FUSION

After appearance and shape recognition part separately, their outputs are combined by a weighted sum rule to produce the final decision. For a given appearance-shape facial test datum of O_A and O_S , the recognition utterance C^* is given by [33],

$$C^* = \arg\max_i\{\gamma \log P(O_A / \lambda_A^i) + (1-\gamma)\log P(O_S / \lambda_S^i)\} \quad (5)$$

Where λ_A^i and λ_S^i are the acoustic and the visual HMMs for the i^{th} utterance class respectively and $\log P(O_A / \lambda_A^i)$ and $\log P(O_S / \lambda_S^i)$ are there log likelihood against the i^{th} class.

Among various types of score fusion techniques, baseline reliability ratio-based integratio has been used to combine the appearance and shape based facial recognition result. The reliability of each modality can be measured from the outputs of the corresponding HMMs. When the appearance based parameters are not corrupted by any noise, there are large differences between the appearance based HMMs output otherwise the differences become small. The reliability of each modality can be calculated by the most appropriate and best in performance [34],

$$S_m = \frac{1}{N-1}\sum_{i=1}^{N}(\max_j \log P(O/\lambda^j) - \log P(O/\lambda^i)) \quad (6)$$

Which means the average difference between the maximum log-likelihood and the other ones and *N* is the number of classes being considered to measure the reliability of each modality, $m \in \{A, V\}$.

Then the integration weight of audio reliability measure γ_A can be calculated by [35],

$$\gamma_A = \frac{S_A}{S_A + S_S} \quad (7)$$

Where S_A and S_S are the reliability measure of the outputs of the appearance and shape based HMMs respectively.
The integratio weight of visual modality measure can be found as,

$$\gamma_S = (1-\gamma_A) \quad (8)$$

6. MULTIPLE CLASSIFIER FUSION

An effective way to combining multiple classifiers is required when a set of classifiers outputs has been created. Various architectures and schemes have been proposed for combining multiple classifiers [36]. The majority vote [37, 38, 39, 40] is the most popular approach. Majority vote approach has been used to combine three classifiers i.e. appearance based, shape based and appearance-shape based outputs in this work. The general voting routine can be defined as [41],

$$E(d) = \begin{cases} c_i \underset{t \in \{1,..m\}}{\forall} \sum_{j=1}^{n} B_j(c_t) \le \sum_{j=1}^{n} B_j(c_i) \ge \alpha.m + k(d) \\ r \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \quad (9)$$

Where α is a parameter, k(d) is a function that provides additional voting constraints and the binary characteristics function can be defined as,

$$B_j(c_i) = \begin{cases} 1 \text{ if } d_j = c_i \\ 0 \text{ if } d_j \ne c_i \end{cases} \quad (10)$$

Where the output of the classifiers from the decision vector, $d = [d_1, d_2, ... d_n]^T$ and $d_i \in \{c_1, c_2, ..., c_m, r\}$, c_i denotes the label of the i[th] class and r the rejection of assigning the input sample to any class.

7. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

The critical parameter of HMM i.e. number of hidden states can affect the performance of the proposed system. A trade off is made to explore the optimal values of the above parameter and experiment is performed which is shown in sub-section 7.1.

Optimal values of the number of hidden states are chosen and finally find out the results of appearance based only, shape based only, combind appearance-shape based feature vector and overall performance of the appearance based, shape based, combined appearance-shape based feature which are elaborated in sub-section 7.2.

7.1 Optimum Value Selection of the Number of Hidden States of DHMM, N_H

In the learning phase of DHMM, We have chosen the hidden states in the range from 5 to 25. The highest performance of 87[%] have been achieved at N_H=10 which is shown in figure 4.

7.2 Performance Measurements of the Proposed System

VALID face database [42] has been used to measure the performance of the proposed system where four different office environment facial images are exists for each person. 150 persons facial images are chosen to evaluate the performance where three (i.e. 1, 2 and 4) faces are used for learning and other face (i.e. 3) is used for testing purpose. Figure 5 focuses the performance comparison among various uni-modal and multi-modal facial recognition system.

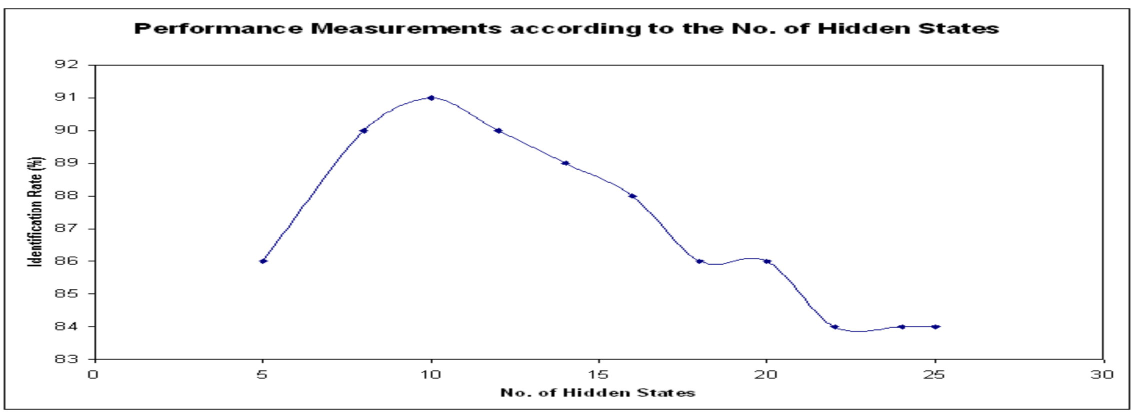

Fig 4: Results after setting up the hidden states of DHMM

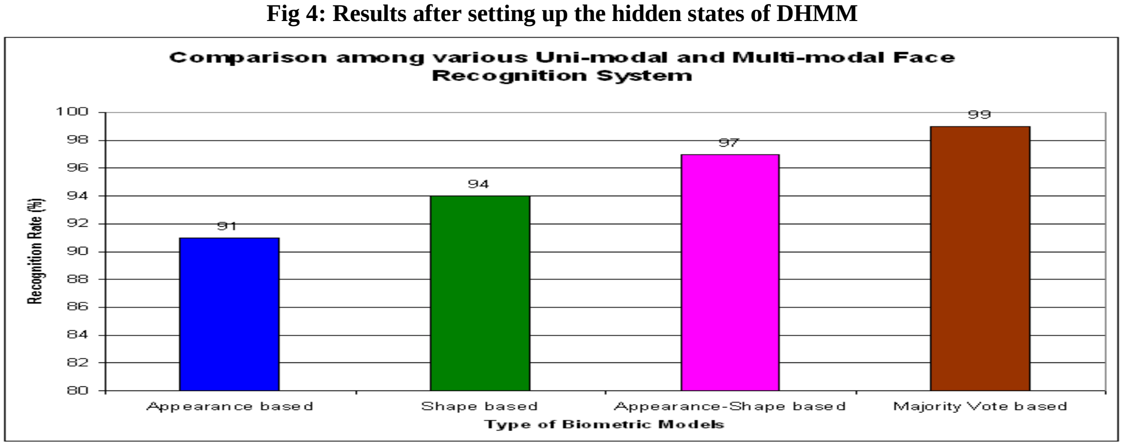

Fig 5: Performance comparison among appearance only, shape only, appearance-shape feature fusion based and combined classifiers based i.e. majority vote output of the proposed system

8. CONCLUSIONS AND OBSERVATIONS

The highest face recognition rate of 99% has been achieved for the proposed feature and decision fusion based technique. Experiments show that the proposed model of face recognition gives a promising result for official environmental conditions which can satisfy any practical demand. Future works include increasing the types of noise and improving the model so that it can handle new noises that are previously unknown for the system. Finally, hybrid classification methods introducing neural networks, genetic algorithm, fuzzy logic and so on can be used to optimally constructed given a large data set of facial images.

9. REFERENCES

[1] Bin Zhang and Sargur N. Srihari, "CLASS-WISE MULTI-CLASSIFIER COMBINATION BASED ON DEMPSTER-SHAFER THEORY", *Seventh International Conference on Control, Automation, Robotics and Vision (ICARV 2002),* Singapur, Dec. 2002.

[2] T. K. Ho. A theory of multiple classifier systems and itsapplication to visual word recognition. PhD Dissertation, State University of New York at Buffalo, 1992.

[3] A. Samal and P.A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Patt. Recog,*. 25 (1), pp. 65–77, 1992.

[4] D. Valentin, H. Abdi, A.J. O_Toole, "G.W. Cottrell, Connectionist models of face processing: a survey*," Patt. Recog,*. 27 (9), pp. 1209–1230, 1994.

[5] R. Chellappa, C.L. Wilson and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc. IEEE,* 83 (5), pp. 705–740, 1995.

[6] J. Zhang, Y. Yan and M. Lades, "Face recognition: eigenface, elastic matching, and neural nets," *Proc. IEEE*, 85 (9), pp. 1423–1435, 1997.

[7] I. Craw, N. Costen, T. Kato and S. Akamatsu, "How should we represent faces for automatic recognition?," *IEEE Trans. Patt. Anal. Mach. Intell.*, 21 (8), pp. 725–736, 1999.

[8] A.M. Burton, V. Bruce, P.J.B. Hancock, "From pixels to people: a model of familiar face recognition," *Cognitive Sci.* 23, pp. 1–31, 1999.

[9] S. Kong, J. Heo, B. Abidi, J. Paik and M. Abidi, "Recent Advances in Visual and Infrared Face Recognition - A Review," *The Journal of Computer Vision and Image Understanding*, Vol. 97, No. 1, pp. 103-135, 2005.

[10] K. Parimala Geetha, S. Sundaravadivelu and N. Albert Singh, "Rotation Invariant Face Recognition using Optical Neural Networks," *TENCON 2008 - 2008 IEEE Region 10 Conference,* Hyderabad, India, 2009.

[11] Kiyomi Nakamura, and Hironobu Takano, "Rotation and Size Independent Face Recognition by the Spreading Associative Neural Network," *International Joint Conference on Neural Networks*, Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada, 2006.

[12] S. H. Lin, S. Y. Kung, and L. J. Lin., "Face recognition/detection by probabilistic decision-based neural network," *IEEE Transactions on Neural Networks, Special Issue on Artificial Neural Networks and Pattern Recognition*, 8(1), 1997.

[13] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, "Neural network based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1), 1998.

[14] Jianke Zhu, Mang I Vai and Peng Un Mak, "Gabor Wavelets Transform and Extended Nearest Feature Space Classifier for Face Recognition," *Proceedings of the Third IEEE International Conference on Image and Graphics (ICIG'04),* 2004.

[15] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, 1990.

[16] J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Joutsensalo, " A class of neural networks for independent component analysis," *IEEE Trans. Neural Networks*, Vol. 8, pp. 486–504, 1997.

[17] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 130–136, 1997.

[18] M. Turk and A. Pentland, "Eigenfaces for recognition*," J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71–86, 1991.

[19] Y.-S. Ryu and S.-Y. Oh, "Simple hybrid classifier for face recognition with adaptively generated virtual data*," Pattern Recognit. Leters.*, vol. 23, pp. 833–841, 2002.

[20] J. Bala, P. Pachowicz, and K. De Jong, "Multistrategy Learning from Engineering Data by Integrating Inductive Generalization and Genetic Algorithms, in Machine Learning: A Multistrategy Approach," *Michalski and G. Tecuci (Eds.),* Vol. IV, R.S. Morgan Kaufmann, San Mateo, CA., pp. 121-138, 1994.

[21] F. Gruau and D. Whitley, "Adding Learning to the Cellular Development of Neural Networks: Evolution and the Baldwin Effect*," Evolutionary Computation,* Vol.1, No.3, pp. 213-234, 1993.

[22] H. Vafaie and K. De Jong, "Improving a Rule Induction System Using Genetic Algorithms, in Machine Learning: A Multistrategy Approach," Michalski and G. Tecuci (Eds.), Vol. IV, R.S., Morgan Kaufmann, San Mateo, CA., pp. 453-469, 1994.

[23] Stephen Milborrow and Fred Nicolls, "Locating Facial Features with an Extended Active Shape Model," available at http://www.milbo.org/stasm-files/locating-facial-features-with-an-extended-asm.pdf.

[24] R. Herpers, G. Verghese, K. Derpains and R. McCready, "Detection and tracking of face in real environments," *IEEE Int. Workshop on Recognition, Analysis and Tracking of Face and Gesture in Real- Time Systems,* Corfu, Greece, pp. 96-104, 1999.

[25] J. Daugman, "Face detection: a survey," *Comput. Vis. Imag. Underst,* 83, 3, pp. 236- 274, 2001.

[26] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing.* Addison-Wesley, 2002.

[27] Stephen Milborrow and Fred Nicolls, "Locating Facial Features with an Extended Active Shape Model," available at http://www.milbo.org/stasm-files/locating-facial-features-with-an-extended-asm.pdf.

[28] R. Herpers, G. Verghese, K. Derpains and R. McCready, "Detection and tracking of face in real environments," *IEEE Int. Workshop on Recognition, Analysis and*

Tracking of Face and Gesture in Real- Time Systems, Corfu, Greece, pp. 96-104, 1999.

[29] J. Daugman, "Face detection: a survey," *Comput. Vis. Imag. Underst,* 83, 3, pp. 236- 274, 2001.

[30] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing.* Addison-Wesley, 2002.

[31] Jong-Seok Lee and Cheol Hoon Park, *Speech Recognition, Technologies and Applications,* pp. 275-296, I-Tech, Vienna, Austria, 2008.

[32] P. A. Devijver, **"**Baum's forward-backward algorithm revisited**",** *Pattern Recognition Letter*, 3, pp. 369-373, 1985.

[33] A. Rogozan, P.S. Sathidevi, "Static and dynamic features for improved HMM based visual speech recognition*," 1st International Conference on Intelligent Human Computer Interaction,* Allahabad, India, pp. 184-194, 2009.

[34] J. S. Lee, C. H. Park, "Adaptive Decision Fusion for Audio-visual speech Recognition", Speech Recognition, Technologies and Applications, ed. F. Mihelic, J. Zibert, (Vienna, Australia, 2008), pp. 550, 2008.

[35] A. Adjoudant, C. Benoit, "On the integratio of auditory and visual parameters in an HMM-based ASR," Speechreading by Humans and Machines: Models, Systems, and Speech Recognition, Technologies and Applications, ed. D.G. Strok and M. E. Hennecke, (Springer, Berlin, Germany, 1996), pp. 461-472.

[36] Nayer Wanas, "Feature Based Architecture for Decision Fusion," Ph.D. Dissertation, Dept. of Systems Design Engineering, University of Waterloo, Ontario, 2003.

[37] R. Battiti and A. Colla. Democracy in neural nets: Voting schemes for classification. Neural Networks, vol. 7, no. 4, pp. 691–707, 1994.

[38] C. Ji and S. Ma., "Combinations of weak classifiers," *IEEE Transactions on Neural Networks,* Vol. 8, No. 1, pp. 32–42, 1997.

[39] L. Lam and C. Suen. Optimal combination of pattern classifiers. Pattern Recognition Letters, vol. 16, pp. 945–954, 1995.

[40] L. Xu, A. Krzy´zak, and C. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 22, no. 3, pp. 418–435, 1992.

[41] Dymitr Ruta and Bogdan Gabrys, "An Overview of Classifier Fusion Methods," *Computing and Information Systems,* 7 (2000), University of Paisley, p.1-10, 2002.

[42] N. A. Fox, B. A. O'Mullane and R. B. Reilly, "The Realistic Multi-modal VALID database and Visual Speaker Identification Comparison Experiments," *Proc. of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA-2005),* New York, 2005.

AUTHORS BIOGRAPHIES

Md. Rabiul Islam was born in Rajshahi, Bangladesh, on December 26, 1981. He received his B.Sc. degree in Computer Science & Engineering and M.Sc. degrees in Electrical & Electronic Engineering in 2004, 2008, respectively from the Rajshahi University of Engineering & Technology, Bangladesh. From 2005 to 2008, he was a Lecturer in the Department of Computer Science & Engineering at Rajshahi University of Engineering & Technology. Since 2008, he has been an Assistant Professor in the Computer Science & Engineering Department, University of Rajshahi University of Engineering & Technology, Bangladesh. His research interests include bio-informatics, human-computer interaction, speaker identification and authentication under the neutral and noisy environments.

Md. Fayzur Rahman was born in 1960 in Thakurgaon, Bangladesh. He received the B. Sc. Engineering degree in Electrical & Electronic Engineering from Rajshahi Engineering College, Bangladesh in 1984 and M. Tech degree in Industrial Electronics from S. J. College of Engineering, Mysore, India in 1992. He received the Ph. D. degree in energy and environment electromagnetic from Yeungnam University, South Korea, in 2000. Following his graduation he joined again in his previous job in BIT Rajshahi. He is a Professor in Electrical & Electronic Engineering in Rajshahi University of Engineering & Technology (RUET). He is currently engaged in education in the area of Electronics & Machine Control and Digital signal processing. He is a member of the Institution of Engineer's (IEB), Bangladesh, Korean Institute of Illuminating and Installation Engineers (KIIEE), and Korean Institute of Electrical Engineers (KIEE), Korea.