

Partition Based Feature Extraction Technique for Facial Expression Recognition Using Multi-Stage Hidden Markov Model

¹Mayur Rahul, ²Narendra Kohli and ³Rashi Agrawal

¹A.P.J. Abdul Kalam Technical University (AKTU), Lucknow, Uttar Pradesh (UP), India

²Department of Computer Science and Engineering, HBTU, Kanpur, Uttar Pradesh (UP), India

³Department of IT, University of Engineering and Technology (UIET),
CSJMU, Kanpur, Uttar Pradesh (UP), India

Abstract: Partition based feature extraction is widely used in the pattern recognition and computer vision. This method is robust to some changes like occlusion, background, etc. In this study, partition based technique is used for feature extraction and extension of HMM is used as a classifier. The new introduced multi-stage HMM consists of two layers. In which bottom layer represents the atomic expression made by eyes, nose and lips. Further upper layer represents the combination of these atomic expressions such as smile, fear, etc. Six basic facial expressions are recognised, i.e., anger, disgust, fear, joy, sadness and surprise. Our proposed system is able to get overall accuracy of 82% using JAFFE database.

Key words: JAFFE, Gabor wavelets transform, PCA, local binary patterns, SVM, FAPs, Cohn-Kanade database, Markov process, static modelling

INTRODUCTION

The facial expression is the most useful medium for non-verbal communication in human daily life. Facial expression recognition can be applied in many interesting areas such as in human-computer interface, telecommunications, behavioural science, video games, animations, psychiatry, automobile safety, affect sensitive music juke boxes and televisions, educational software, etc. Better classifier and better feature extraction method is very important because it significantly improves the accuracy and speed for identification and classification.

The hidden markov modelling is useful in spectral and dynamic nature of speech pattern using acoustic state model (Katagiri and Lee, 1993). The HMM becomes the most useful in speech recognition. Due to handling of time series data modelling and learning capability, HMM is very useful in classifying unknown feature vector sequence. Only small number of training samples required to train full HMM (Tsapatsoulis *et al.*, 1998; Rabiner, 1989). The discriminative strength of normal HMM is less suitable to handle more difficult task. Due to dynamic statistical modelling and time sequence pattern matching principle, HMM matches the most similar signal state as recognition results (Atlas *et al.*, 2000; Hatzipantelis *et al.*,

1995). Hidden Markov Model is capable for processing continuous dynamic signal, can effectively use the timing signal moments before the state transition and after the state transition.

In this study, we are using partition based feature extraction method and introduce two layer extension of HMM. New introduced HMM is consists of two layers. Bottom layer represents the atomic facial expressions such as expressions involving eyes, nose and lips separately. Upper layer represents the expressions which are basically the combination of atomic expressions such as smile, fear, etc. Our principle is that every facial expression is the combination of expressions of eyes, nose and lips. Our newly introduced system is capable of handle almost every situation like occlusion, brightness, background, etc.

Basics of HMM: Markov process is used for time series data modelling. This model is applicable for situations where there is a dependability of present state to previous state. For example, bowler bowls the ball in cricket game, ball first hit the ground then hit the bat or pad. This situation can be easily modelled by time series data modelling. In other example such as in a spoken sentence, the present pronounced word is depending on the previous pronounced word. Markov process effectively handles such situations (Fig. 1).

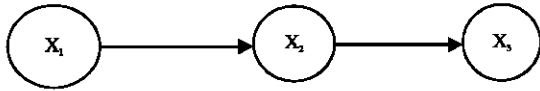


Fig. 1: A markov model for 3 observations variables

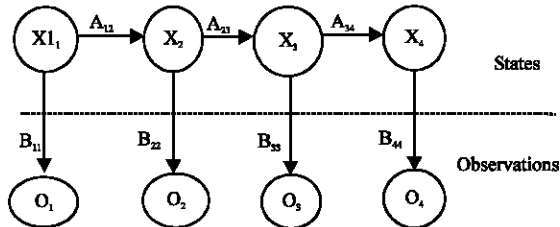


Fig. 2: A Hidden Markov Model for four variables

When an observation variable is dependent only on previous observation variable in a markov model is called first order markov chain. The joint probability distribution is given by:

$$p(X_1, X_2, \dots, X_N) = p(X_1) \pi p(X_N | X_{N-1}) \quad (1)$$

Where:

$p(X_1, X_2, \dots, X_N)$ = The joint probability distribution of states X_1, X_2, \dots, X_N
 $p(X_1)$ = The probability of state X_1
 $p(X_N | X_{N-1})$ = The probability of state X_N given state X_{N-1}

A Hidden Markov Model is a statistical model which is used to model a markov process with some hidden states. This model is widely used in speech and gesture recognition. Hidden markov model is a set of observable states are measured and these variables are assumed to be depend on the states of a markov process which are hidden to the observer (Fig. 2). Hidden states are states above the dashed lines.

$$p(X_{t+1} = j | X_t = i) = a_{ij} \quad (2)$$

Where:

$p(X_{t+1} = j | X_t = i)$ = The conditional probability of the state X_{t+1} given the state X_t
 a_{ij} = The i th row, j th column entry of the transition matrix $A = (a_{ij})$

$$p(O_t = k | X_t = i) = b_{ik} \quad (3)$$

Where:

$p(O_t = k | X_t = i)$ = The conditional probability of the observation variable O_t given the state X_t
 b_{ik} = The i th row, k th column entry of the emission matrix $B = (b_{ik})$. An HMM is defined by A, B and π . The HMM is denoted by $\partial = (A, B, \pi)$

Literature review: The facial expression and Physiognomy was first started in the early 4th century. The physiognomy is the study of the person's character from their outer appearance (Highfield *et al.*, 2009). The interest on physiognomy is minimized and the studies on facial expressions become active. The studies of facial expressions have been found in 17th century back. John Bulwer in 1649 was given the detail theory about the facial expressions and head movement in his book "Pathomyotomia". Le Brun in 1667 gave a detailed notes on facial expression, then it was later reproduced as a book in 1734 (Northwestern University, 2017).

In the beginning of 19th century, Charles Darwin gives a theory which produced the greatest impact in today's automatic facial expression recognition field. Darwin created a benchmark for various expressions and means of expressions in both animals and humans (Darwin, 1904). Ekman and Friesen (1978) done a great job in this field and further their work has great influence for today's new era of Facial Expression Recognition. Today this field is used in almost all area for example in clinical and social psychologists, medical practitioners, actors and artists. However in the end of the 20th century with the advances in the fields of robotics, computer graphics and computer vision, animators and computer scientists started showing interest in the field of facial expressions. Suwa in 1978 creates a framework in automatic facial expression recognition by tracking 20 points to analyse facial expressions from the sequence of images.

Ekman and Freisan (1978) introduced a framework in psychology to recognize facial expression. In 1990, researcher used Ekman's framework to recognize facial expression in video and image (Cohen *et al.*, 2003). Hidden Markov Model was also used to recognised facial expressions (Lien *et al.*, 2000). Further, Cohen *et al.* (2003) proposed a multilevel HMM to recognise emotions and further optimize it for better accuracy as compared to emotion specific HMM. Pardas and Bonafonte (2002) used the automatic extraction of MPEG-4 Facial Animation Parameters (FAP) and also proved that the FAPs provide the necessary information required to extract the emotions. FAPs were extracted using an improved active contour algorithm and motion estimation. They used the HMM classifier. They also used the Cohn-Kanade database for recognition. Their recognition rate was 98% for joy, surprise and anger and 95% for joy surprise and sad. The average recognition rate was 84%. Aleksic and Katsaggelos (2006) showed their performance improvement using MS-HMM. They also used PCA to reduce dimensionality before giving it to HMM. They used MPEG-4 FAPs, outer lip (group 8) and eyebrow (group 4) followed by PCA to reduce dimensionality. They used HMM and MS-HMM as classifiers. They used Cohn-Kanade databases with 284 recordings of 90

subjects. Their recognition rate using HMM was 88.73% and using MS-HMM was 93.66%. Shang and Chan (2009) proposed an effective nonparametric output probability estimation method to increase the accuracy using HMM. They worked on CMI database and get the accuracy of 95.83% as compared to non-parametric HMMs. Jiang (2011) proposed a method based on code-HMM and KNN further applied some discrimination rules. Their proposed method achieves better accuracy to some extent. Suk and Prabhakaran, 2015 proposed a real-time temporal video segmenting approach for automatic facial expression recognition. They used SVM as a classifier and get the accuracy of 70.6%. Wu *et al.* (2015) incorporated a multi-instance learning problem using CK+ and UNBC-McMaster shoulder pain database. They combine both multi-instance learning and HMM to recognize facial expression and outperforms state-of-the-arts. Sikka *et al.* (2015) proposed a model based similarity framework and combine SVM and HMM (Sikka *et al.*, 2015). They worked on CK+ and OULU-CASIA databases and get the accuracy of 93.89%. Ramkumar and Logashanmugam (2016) incorporated the Active Appearance Model (AAM) to identify the face and extracted its features. They used both KNN and HMM to recognize facial expression and get the better results. Senthil and Vijayalakshmi (2014) proposed a method for recognizing facial expression using HMM. Xufen incorporated some modification in HMM to get better recognition rate (Jiang, 2011). Xiaorong *et al.* (2012) proposed a framework for partially occluded face recognition using HMM and get the better result to some extent. Punitha *et al.* proposed a real-time facial expression recognition frame work and get the better results as compared to existing framework. Pagariya and Bartere (2013) also proposed a system using multilevel

HMM for recognizing facial expression from the video. Islam incorporated both PCA and HMM for appearance and shape based facial expression recognition framework (Islam *et al.*, 2012). Singh *et al.* (2012) introduced a 3 state HMM or recognising faces under various poses.

MATERIALS AND METHODS

Different database requires different feature extraction method and different classifiers. It is not easy to conclude which feature extraction methods and classifiers best suited the situation. HMMs have been used in speech recognition so far. Due to handling of time series data modelling and learning capability, HMM is very useful in classifying unknown feature vector sequence. Only small number of training samples required to train full HMM (Tsapatsoulis *et al.*, 1998; Rabiner, 1989). The discriminative strength of normal HMM is less suitable to handle more difficult task. In this study, we introduce a new framework which uses most powerful partition based feature extraction method and modified HMM as a classifier. New multi-stage HMM is the two layer extension of normal HMM. Bottom layer represents the atomic expressions made by eyes, lips and nose and upper layer represents the combination of these atomic expressions such as smile, fear, etc. The proposed framework uses the partition based feature extraction method and then these extracted features are used to train the classifier. Some dataset are used to test the classifier. This framework is robust to occlusion, background and orientation. In this proposed system, Baum-Welch method is used for parameter estimation. Viterbi method and forward procedure are used for calculating the optimal state sequence and probability of the observed sequence, respectively. The proposed framework is shown in Fig. 3.

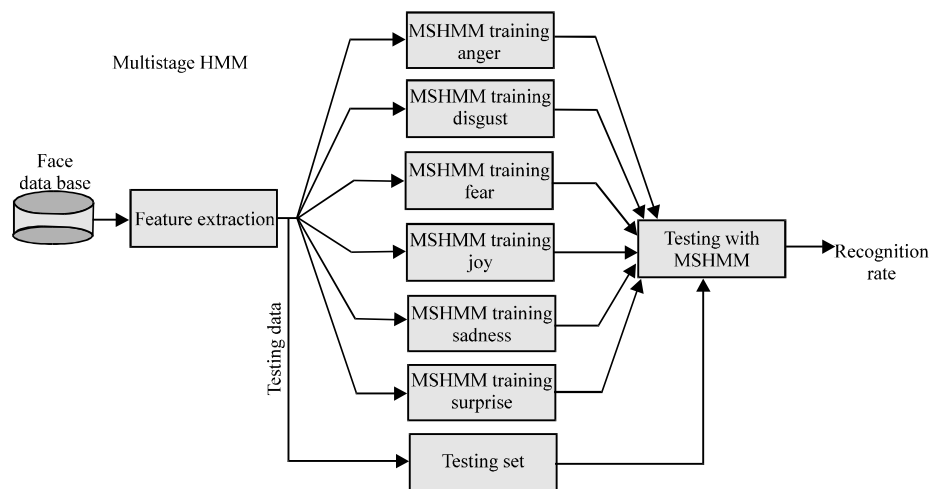


Fig. 3: Block diagram for facial expression recognition using proposed system

RESULTS AND DISCUSSION

Following conditions have been applied during the experiments:

- Faces are analysed from frontal view only
- No movement in the head
- No conversation during image taken
- No glasses during image capturing

JAFFE database has 152 images from 13 subjects. They are 16 of anger, 27 of disgust, 19 of fear, 33 of joy, 30 of sadness and 27 of surprise. To train the HMM, we follow N fold cross validation rule. For example if there is k fold, then k-1 fold are used to train the HMM and remaining 1 fold is used to test HMM. For example in the set of nine images of anger. These 9 images are divided into nine folds. We are using 8 images to train and 1 for test. Further final result is the mean of all the 9 results.

Images are taken from the JAFFE database of facial expression. Each image is partitioned into three parts Fig. 4a, b. Partitioning of image is very effective method as compared to other state-of-the-art feature extraction methods. These partitions are easily distinguishable to achieve the higher accuracy (Fig. 5). Plot shows the Eyes, nose and lips of all six expressions of the same person.

Again next plot shows the eyes, nose and lips partition of same “disgust” expression for different persons (Fig. 6).

These plots show the difference between different expressions of the same person and same expressions of the different persons. This partition based technique

when incorporated with new introduced HMM gives much better results as compared to other state-of-the-art methods. The recognition results using Partition Based feature extraction technique and modified HMM as a classifier is shown in. From the results we can see that using the proposed System will improves the recognition rate.

The receiver operating characteristics curve or ROC curve is a graph used to represent the performance of the classifiers (Fig. 7). It is the plot between True Positive Rate (TPR) versus False Positive Rate (FPR). TPR is the proportion of positive samples identified as positive samples and FPR is the proportion of negative samples identified as positive. The 45° line is called “line of no-discrimination”. The ROC curve above this line represents the goodness of the classifier and also indicates the classifying rate of identifying positive samples as positive is greater than its rate of misclassifying negative samples. This given ROC curve indicates the goodness of our classifier.

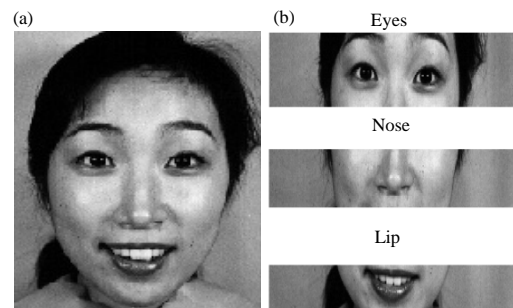


Fig. 4: a) Original image and b) Partitioned image

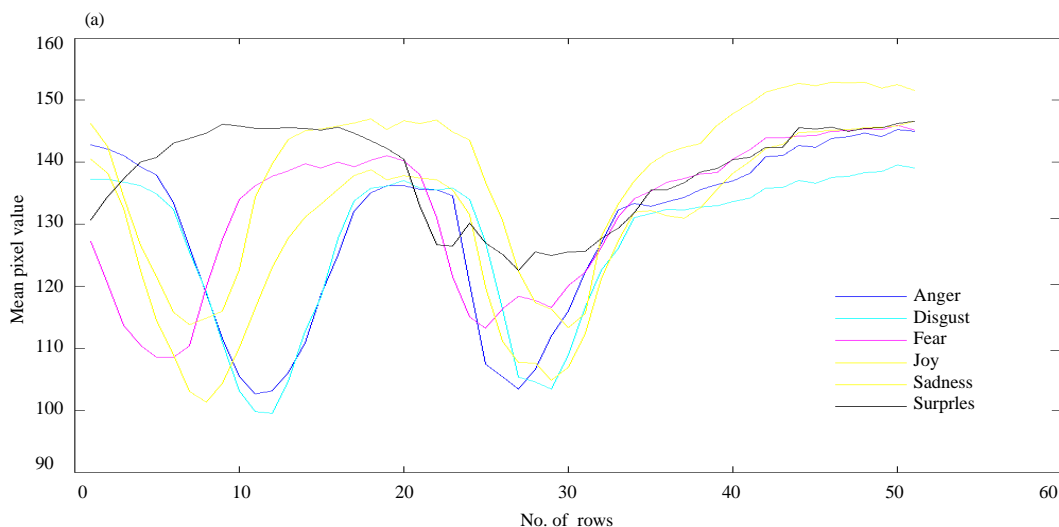


Fig. 5: Continue

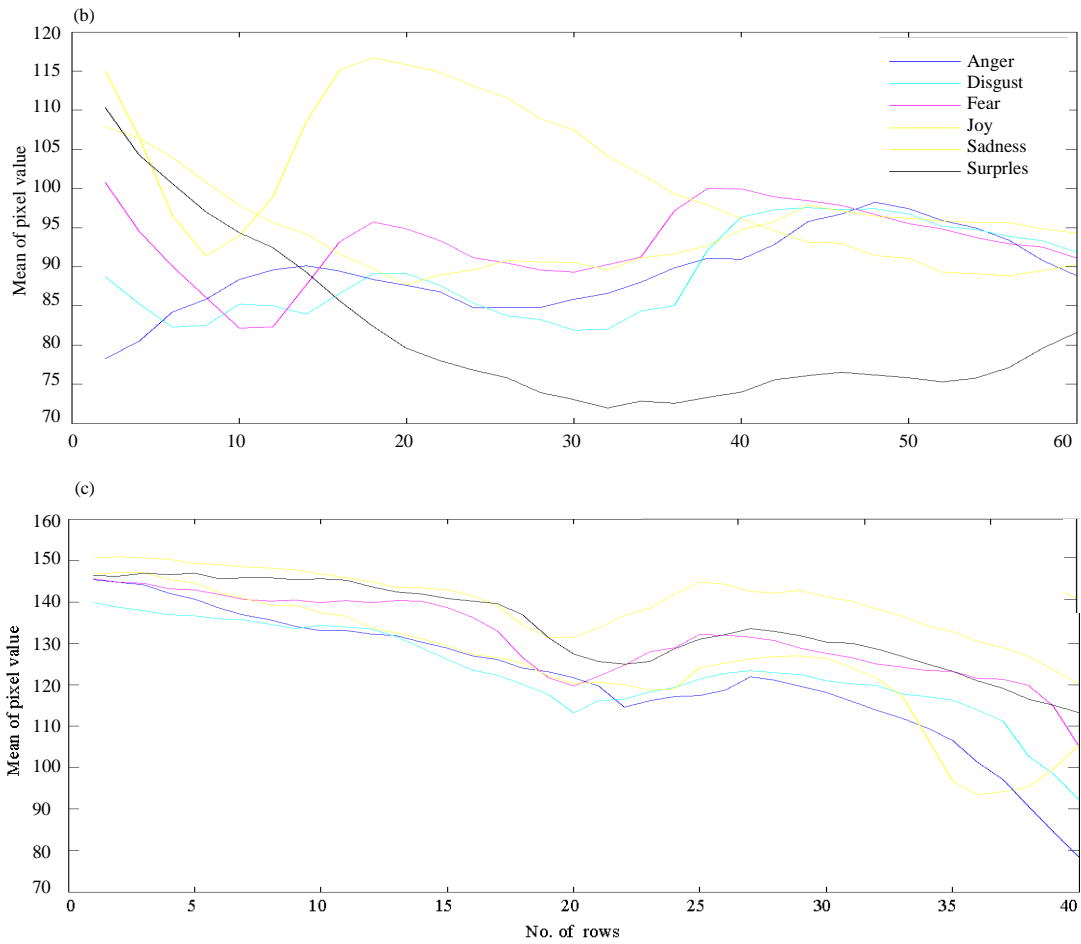


Fig. 5: a) Plot of eyes partition; b) Plot of lips partition and c) Plot of nose partition

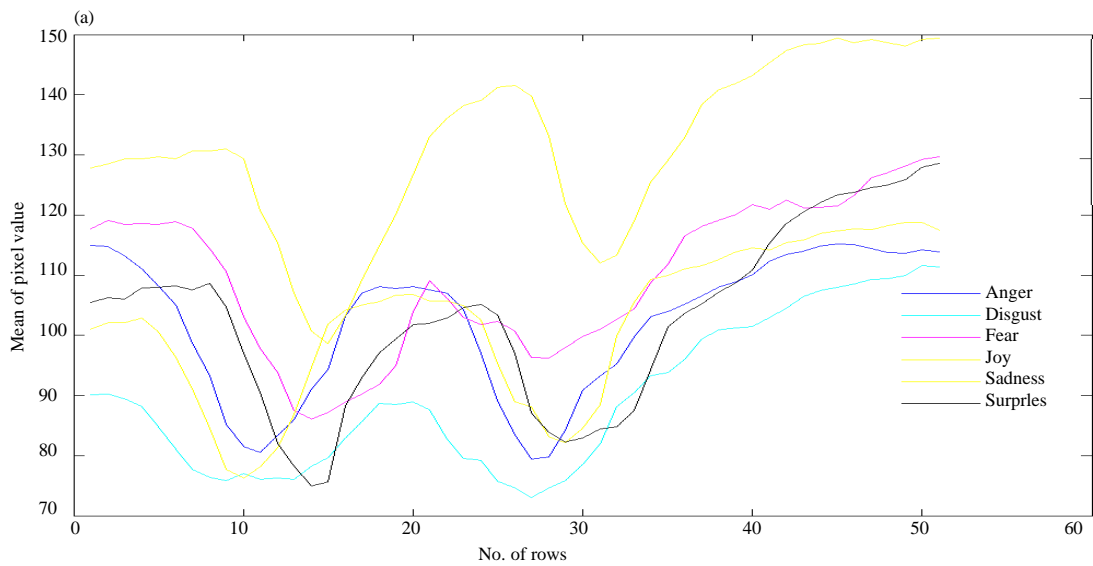


Fig. 6: Continue

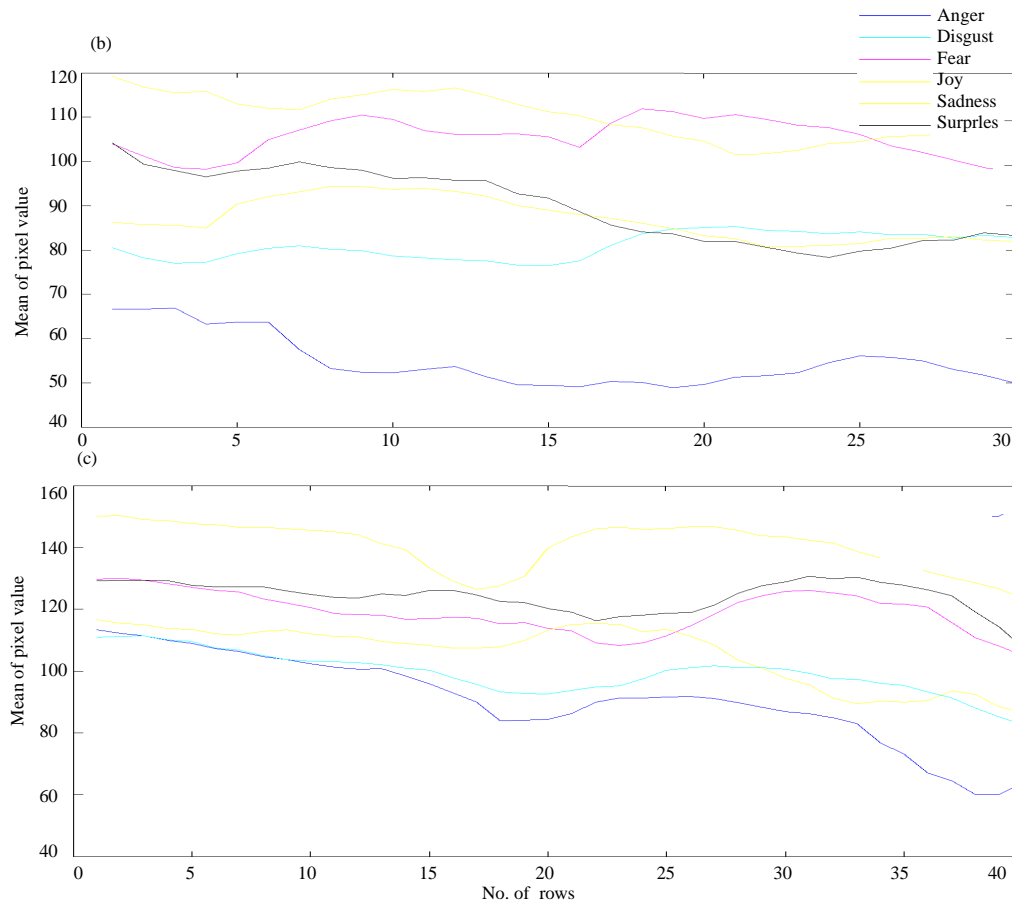


Fig. 6: a) Plot of eyes partition for “Disgust” expression of different persons; b) Plot of lips partition for “Disgust” expression of different persons and c) Plot of nose partition for “Disgust” expression of different persons

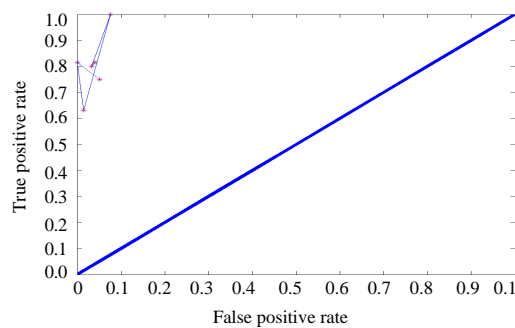


Fig. 7: ROC curve for the given confusion matrix

Table 1: Confusion matrix for FER using proposed system

| Factors | Anger | Disgust | Fear | Joy | Sadness | Surprise | Recognition rate (age %) |
|----------|-------|---------|------|-----|---------|----------|--------------------------|
| Anger | 12 | 0 | 2 | 0 | 0 | 2 | 75 |
| Disgust | 0 | 22 | 0 | 0 | 2 | 3 | 81 |
| Fear | 1 | 0 | 12 | 4 | 2 | 0 | 64 |
| Joy | 0 | 0 | 0 | 33 | 0 | 0 | 100 |
| Sadness | 6 | 0 | 0 | 0 | 24 | 0 | 80 |
| Surprise | 0 | 0 | 0 | 5 | 0 | 22 | 81 |

CONCLUSION

It is not easy to conclude which feature extraction method and classifier is best for this situation but still modification of HMM significantly improves accuracy for the recognition of six basic facial expressions. This study introduced a first framework for partition based method as a feature extraction and modified HMM as a classifier for facial expression recognition. HMM provides some modelling advantages over other feature based methods in facial expression recognition. Due to generative nature of HMM, it is weak classifier as compared to other discriminative classifier such as SVM. The proposed method makes HMM more powerful as a classifier with compared to other discriminative classifiers. We have also showed its strength using confusion matrix and Receiver Operating Characteristics (ROC) curve and found the overall accuracy of 82%.

REFERENCES

- Aleksic, P.S. and K.A. Katsaggelos, 2006. Automatic facial expression recognition using facial animation parameters and multistream HMMs. *IEEE Trans. Inform. Forensics Secur.*, 1: 3-11.
- Atlas, L., M. Ostendorf and G.D. Bernard, 2000. Hidden Markov models for monitoring machining tool-wear. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP Vol. 6, June 5-9, 2000, IEEE, Istanbul, Turkey, ISBN: 0-7803-6293-4, pp: 3887-3890.*
- Cohen, I., N. Sebe, A. Garg, L.S. Chen and T.S. Huang, 2003. Facial expression recognition from video sequences: Temporal and static modeling. *Comput. Vision Image Understanding*, 91: 160-187.
- Darwin, C., 1904. *The Expression of the Emotions in Man and Animals*. 2nd Edn., John Murray Publisher, London, England,.
- Ekman, P. and W. Friesen, 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA., USA.
- Hatzipantelis, E., A. Murray and J. Penman, 1995. Comparing hidden Markov models with artificial neural network architectures for condition monitoring applications. *Proceedings of the 4th International Conference on Artificial Neural Networks Vol. 6, June 26-28, 1995, IET, Cambridge, England, ISBN: 0-85296-641-5, pp: 369-374.*
- Highfield, R., R. Wiseman and R. Jenkins, 2009. How your looks betray your personality. *N. Sci.*, 201: 28-32.
- Islam, M.R., R. Toufiq and M.F. Rahman, 2012. Appearance and shape based facial recognition system using PCA and HMM. *Proceedings of the 7th International Conference on Electrical and Computer Engineering (ICECE), December 20-22, 2012, IEEE, Dhaka, Bangladesh, ISBN: 978-1-4673-1434-3, pp: 1-4.*
- Jiang, X., 2011. A facial expression recognition model based on HMM. *Proceedings of the International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT) Vol. 6, August 12-14, 2011, IEEE, Harbin, China, ISBN: 978-1-61284-087-1, pp: 3054-3057.*
- Katagiri, S. and C.H. Lee, 1993. A new hybrid algorithm for speech recognition based on HMM segmentation and learning vector quantization. *IEEE. Trans. Speech Audio Process.*, 1: 421-430.
- Lien, J.J.J., T. Kanade, J.F. Cohn and C.C. Li, 2000. Detection, tracking and classification of action units in facial expression. *Rob. Auton. Syst.*, 31: 131-146.
- Northwestern University, 2017. *Explore our Many Libraries and Notable Collections*. Northwestern University, Evanston, Illinois,.
- Pagariya, R.R. and M.M. Bartere, 2013. Facial emotion recognition in videos using HMM. *Intl. J. Comput. Eng. Res.*, 3: 111-118.
- Pardas, M. and A. Bonafonte, 2002. Facial animation parameters extraction and expression recognition using hidden Markov models. *Signal Process. Image Commun.*, 17: 675-688.
- Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77: 257-286.
- Ramkumar, G. and E. Logashanmugam, 2016. An effectual facial expression recognition using HMM. *Proceedings of the International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), May 25-27, 2016, IEEE, Ramanathapuram, India, ISBN: 978-1-4673-9546-5, pp: 12-15.*
- Shang, . and K.P. Chan, 2009. Nonparametric discriminant HMM and application to facial expression recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, June 20-25, 2009, IEEE, Miami, Florida, ISBN: 978-1-4244-3992-8, pp: 2090-2096.*
- Sikka, K., A. Dhall and M. Bartlett, 2015. Exemplar hidden Markov models for classification of facial expressions in videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, CVF, Boston, Massachusetts, pp: 18-25.*
- Singh, K.R., M.A. Zaveri and M.M. Raghuwanshi, 2012. Recognizing faces under varying poses with three states hidden Markov model. *Proceedings of the IEEE International Conference on Computer Science and Automation Engineering (CSAE) Vol. 2, May 25-27, 2012, IEEE, Zhangjiajie, China, ISBN: 978-1-4673-0088-9, pp: 359-363.*
- Suk, M. and B. Prabhakaran, 2015. Real-time facial expression recognition on smartphones. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), January 5-9, 2015, IEEE, Waikoloa, Hawaii, ISBN: 978-1-4799-6684-4, pp: 1054-1059.*

- Tsapatsoulis, N., M. Leonidou and S. Kollias, 1998. Facial expression recognition using HMM with observation dependent transition matrix. Proceedings of the IEEE 2nd Workshop on Multimedia Signal Processing, December 7-9, 1998, IEEE, Redondo Beach, California, ISBN:0-7803-4919-9, pp: 89-95.
- Vijayalakshmi, M. and T. Senthil, 2014. Automatic human facial expression recognition using hidden Markov model. Proceedings of the International Conference on Electronics and Communication Systems (ICECS), February 13-14, 2014, IEEE, Coimbatore, India, ISBN:978-1-4799-6638-7, pp: 1-5.
- Wu, C., S. Wang and Q. Ji, 2015. Multi-instance hidden Markov model for facial expression recognition. Proceedings of the 11th IEEE International Conference on Automatic Face and Gesture Recognition (FG) Vol. 1, May 4-8, 2015, IEEE, Ljubljana, Slovenia, ISBN:978-1-4799-6026-2, pp: 1-6.
- Xiorong, P., Z. Zhiku and T. Heng, 2012. Partially occluded face recognition using subspace HMM. Proceedings of the 7th International Conference on Computing and Convergence Technology (ICCT), December 3-5, 2012, IEEE, Seoul, South Korea, ISBN:978-1-4673-0894-6, pp: 720-725.