International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

HMM BASED REAL TIME FACIAL EXPRESSION RECOGNITION

A. Punitha[1], M. Kalaiselvi Geetha[2]

[1,2]*Department of Computer Science & Engineering, Annamalai University, Annamalai Nagar, India.*

E-mail: charuka12@yahoo.com

Abstract

The most expressive way humans display emotions is through facial expressions. The aim of facial expression recognition methods is to build a system for classification of facial expressions from continuous video input automatically. The method proposed by Viola and Jones is used to detect the face region. Since the mouth plays a vital role in expressing emotions, the mouth features are used for classifying expressions. The mouth intensity code value (MICV) extracted from the mouth region is used as a feature in this work. This MICV difference between the first and the greatest facial expression intensity frame is used as an input to a Hidden Markov Model (HMM) to recognize facial expression.

***Keywords*-- Face Detection; Mouth Intensity Code Value (MICV); Facial Expression Recognition; Hidden Markov Model (HMM).**

I. INTRODUCTION

Facial Expression Recognition (FER) from video is an essential research topic in computer vision, impacting important applications in areas such as human–computer interaction and data-driven animation. Facial expression recognition is a sort of visual learning process. Applications include video conferencing, forensics, virtual reality, computer games, machine vision etc., In this work four different expressions happy, surprise, disgust and normal are considered.

1.1 Related Work

Detailed review of existing methods on facial expression is seen in[1,2]. Many methods have been proposed for face tracking which include active contours [3], robust appearance filter [4], probabilistic tracking [5], adaptive active appearance model [6] and active appearance model [7]. Facial expression recognition from still image has less precision with respect to video sequence because a single image offers much less information than a sequence of images for expression recognition processing. Bassili [8] conclude that the facial expression was more accurately recognized from dynamic images than from a single static image. Yacoob and Davis [9] used optical flow to track the motion of brows, eyes, nose and mouth. A lookup table to classify six standard facial expressions was presented in their approach.

Otsuka and Ohya [10] presented a feature point tracking approach, where feature points are chosen automatically in the first frame of a given facial expression sequence. This is achieved by acquiring potential facial feature points from local saddle points of illuminance distributions. Difference image-based motion extraction was employed in [11].

A wide range of classification algorithms (e.g. Hidden Markov Models (HMMs) [12], Bayesian network classifier [13], Support Vector Machine (SVM) [14][15], Neural network (NN) [16][17] PCA [18], LDA [19]) have been applied to the facial expression recognition problem.

Section 2 describes the Face Detection procedure and mouth feature extraction using MICV feature. Expression recognition using HMM is presented in Section 3. Section 4 shows the Experimental results of our approach and Section 5 concludes this work.

II. FACE EXPRESSION RECOGNITION

Boosting and cascading detectors have gained great popularity in face tracking applications, due to their efficiency in selecting features for face detection. This work exploits Adaboost algorithm [20] for face detection. Considering the intensity scale of the different facial expressions, each person has his/her own maximal intensity of displaying a particular facial action.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

It is useful to recognize the temporal intensity change of expressions in videos. Based on this idea, this proposed work makes use of the mouth region intensity code value namely MICV as feature for facial expression recognition. Hidden Markov Models (HMMs) have been widely used to model the temporal behaviors of facial expressions from image sequences. This work exploits HMM to recognize facial expression.

2.1 Face/Mouth Region Detection

Face detection is the first stage of an automated face recognition system, since a face has to be located before it is recognized. The face-detection component of our method is based on the well-known Viola-Jones detector (Viola and Jones, 2001). This detector is known for its computational efficiency and performance. Viola and Jones employed efficient machine learning. Viola and Jones further improved the efficiency of the face detector by creating a cascade of boosted classifiers. Patches are processed in a stage-wise fashion through the cascade. If at any stage of the cascade a patch is rejected, it is classified as a non-face area. Only those patches that reach the final stage are classified as a face area.

The combination of the integral image representation and the cascaded AdaBoost algorithm has been shown to be a highly effective face detector that is used in virtually all real-time face detection software. A method called boosting discard the vast majority of features and retain only those that contributes face detection.

In the face-detection component, the Viola-Jones detector is used to localize the face within each frame of the video. We adopt their standard settings of the detector: the window-size is 24 X 24 pixels and the number of feature types is equal to 3.The three types of haar-like features shown in **Fig. 1** are used to extract the facial features Whenever the Viola-Jones detector fails to localize a face (which happens occasionally due to an occlusion or an uncommon pose), we determine the face location by interpolating the detected face locations in the preceding and succeeding frames. The output of the face-detection component corresponds to square sub-images covering the face with dimensions that range from 1162 to 1262 pixels, depending on the distance of the face to the camera. **Fig. 2** shows a few samples of the detected face, eye and mouth region.

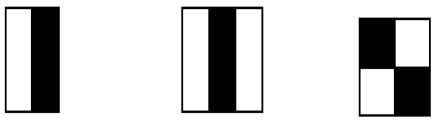

Fig. 1. Types of haar-like features for feature selection

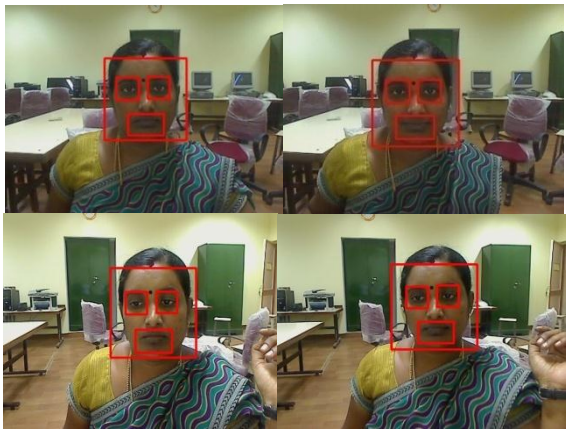

Fig. 2. Images showing detected face, eye and mouth region

2.2 Feature Extraction from Mouth Region

Feature is a descriptive portion extracted from an image or a video stream. Visual data exhibit numerous types of features that could be used to recognize or represent the information it reveals. These features exhibit both static and dynamic properties. Classification or recognizing an appropriate video relies on competent use of these features that provide discriminative information useful for high-level classification. The following subsection present the description of the feature used in this study.

2.2.1 Mouth Intensity Code Value (MICV)

In this section, the method to compute MICV for mouth region is presented, which characterizes the intensity variations between blocks that corresponds to the mouth region in a video frame. The method uses a simple procedure that divides a mouth region into blocks and creates a code called MICV which represents the intensity difference between blocks in a frame. **Eq. (1)** illustrates the generation of proposed MICV feature. i and j represents the i^{th} and j^{th} blocks in a frame. MICV is generated using

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

$$y\left[(i-1)25+j-\frac{i(i+1)}{2}\right]=\begin{cases}1 & if\ x(i)>x(j)\\ 0 & otherwise\end{cases}$$

$$1\le i\le 25, 2\le j\le 25\ and\ i<j, \qquad (1)$$

Where *x(i), x(j)* are the average intensities of the i^{th} and j^{th} blocks respectively. To generate the MICV, for example, the frame is divided into 5 x 5 blocks to generate the feature vector. **Fig. 3** shows the detected mouth region and the 5 x 5 representation of mouth region.

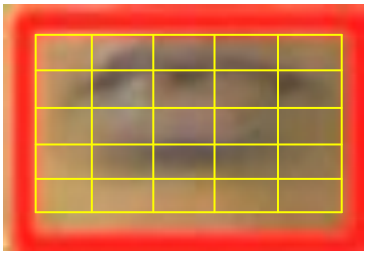

Fig . 3. 5 X 5 representation of mouth region

Each block in a frame is compared with every other block to generate MICV using "Eq. (1)". For example, if the image is divided into 5 x 5 blocks, then "Eq. (1)" generates 300 dimensional feature vectors.

First element in the feature vector compares the intensity of 1st and 2nd block; second element compares the intensity of 1st and 3rd block and so on. The distance or error between the two comparison codes P = (p_1, p_2, ⋯, p_n).and Q= (q_1, q_2, ⋯, q_n) can be calculated using

$$d=\sum_{k=1}^{n}(p_k \oplus q_k) \qquad (2)$$

Where ⊕ is the bit-wise exclusive-OR operation

III. Expression Recognition Using HMM

Feature classification is performed in the last stage of an automatic facial expression analysis system. Hidden Markov models (HMM) are commonly used in the field of speech recognition, but are also useful for facial expression analysis as they allow to model the dynamics of facial actions. Several HMM-based classification approaches can be found in the literature [21, 22, 23] and were mostly employed in conjunction with image motion extraction methods.

The idea is that expressions have a unique temporal pattern and recognizing these patterns can lead to improved classification results. For the recognition of facial expressions, 4 HMM one for each expression is adopted.

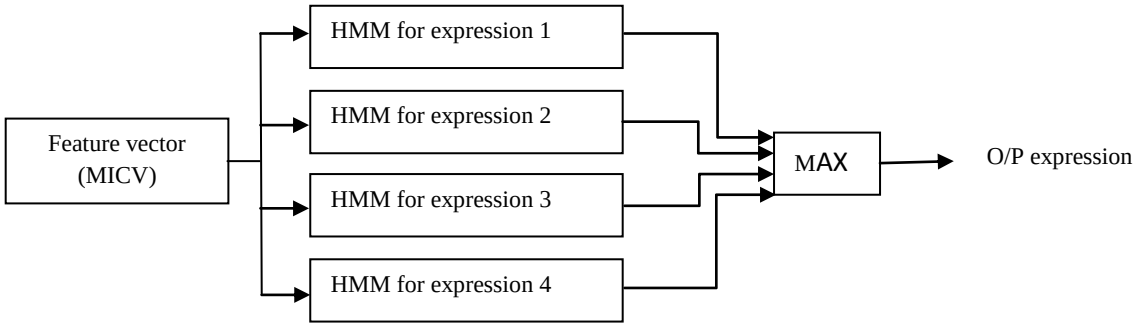

Fig. 4 . HMM classification system

3.1 Hidden Markov model (HMM)

HMM assumes that an input observation sequence of feature vectors follows a multi-state distribution, an output symbol is produced based on the probability distribution It is expressed by the initial state distribution (π), the state transition probabilities (A), and the observation probability distribution in each state (B).

In HMM training, it estimates the parameter set λ = (A, B, π), for each class based on the training sequences. It enters a new state based on the transition probability depending on the previous state. After making the transition depending on the current state, an output symbol is produced based on the probability distribution.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

For HMM training, MICV is extracted from each frame in the video sequence and is given as input to estimate the parameters of HMM. The following steps are implemented using HTK toolkit for the expression recognition.

1. HInit does initialization. It computes an initial set of parameter values using Viterbi alignment to segment the training observations and then recomputes the parameters by pooling the vectors in each segment.
2. To determine the parameters of HMM, the output of HInit is fed as input to HRest. The models are re-estimated using Baum-Welch re-estimation.
3. Testing the data against the model built.

The HMM classification scheme used in the present approach is shown in **Fig. 4.** Initially, separate HMMs are used for each expression. MICV is fed as input to the HMM. Finally, the maximum output obtained is considered as the output expression.

IV. Experimental Results

In this section, we present the experimental results on real-world facial expression dataset. Experiments are conducted on a PC with Logitech webcam Pro9000, 2 megapixel sensors autofocus and mounted on a Intel Xeon® processor in an Ubuntu 12.04 LTS operating system.

Experiments are carried out on 1 hour of video for training and 30 minutes for testing. Precision (P) and recall (R) are the commonly used evaluation metrics and these measures are used to evaluate the performance of the proposed system. The measures are defined as follows:

$$\text{Precision} = \frac{\text{No. of True Positives}}{\text{No. of True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{No. of True Positives}}{\text{No. of True Positives} + \text{False Negatives}}$$

The work used F-score as the combined measure of Precision (P) and recall (R) for calculating accuracy which is defined as follows:

$$F_\alpha = \frac{2PR}{P+R}$$

Where α is weighting factor and α = 0.5 is used.

Fig. 5 shows the subject expressing different expressions. Recognition accuracy has been measured for the four expressions under various resolutions of the faces. The system is able to handle a relatively wide range of face resolutions. Recognition result for various expressions is shown in **Table 1**. It is observed that among the four expressions "happy" expression is recognized better than the others. **Fig. 6** shows the plot of the accuracy at various face resolution. From **Fig. 7**, it is observed that, as the resolution of face is low then the system fails to detect the features (eye, mouth) as well the expressions at lower resolution.

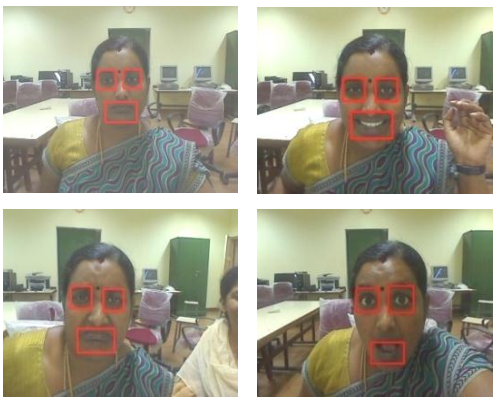

Fig. 5. Real-time test samples showing facial expressions

TABLE 1:
Recognition accuracy for expressions

Expression	Precision (%)	Recall (%)	F- Score (%)
Happy	96	92	94.0
Surprise	93	91	92.0
Disgust	90	84	86.0
Neutral	84	78	81.0

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

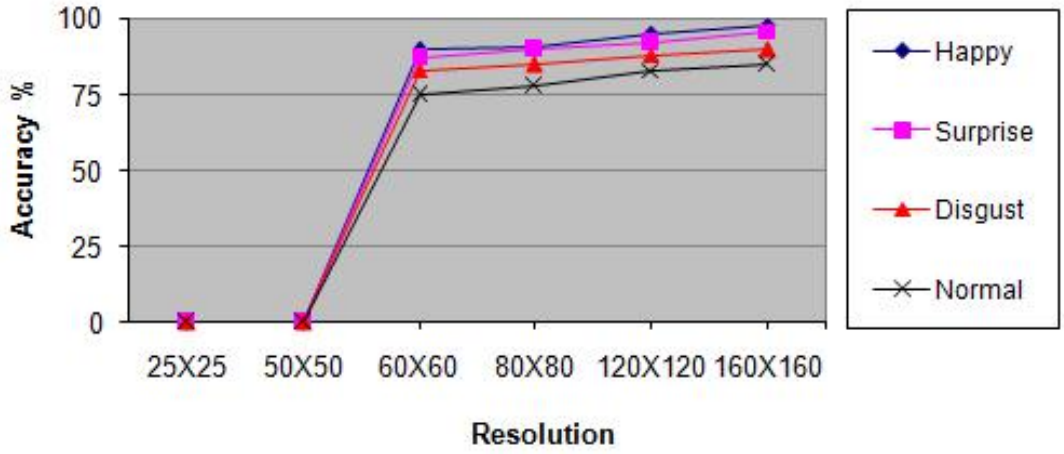

Fig. 6. Accuracy at different face resolution

Images			
Resolution	166 X 166	105 X 105	45 X 45
Face Detected?	Yes	Yes	Yes
Features Detected?	Yes	Yes	No
Expression Recognized?	Yes	Yes	No

Fig. 7. Results for the different resolution faces (Facial features and facial expressions are not recognized at 45 x 45 pixels)

V. CONCLUSION

In this work, a simple approach to automatically recognize facial expression using HMM is presented. The proposed work is able to detect human faces and extract features at different resolutions from the real time video. Among the four expressions, happy expression has been recognized with an accuracy of 94%. Expressions Neutral and disgust cannot be distinguished well. Hence the future work aims to apply the feature extracted in this work to the eye region and also considering more number of expressions.

REFERENCES

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

[2] Fasel, B. and Luettin, J., 2003. Automatic facial expression analysis: a survey, Pattern Recognition 36.

[3] Freedman, D., 2004. Active Contours for Tracking Distributions, IEEE Transactions on Image Processing, Vol. 13, No. 4, pp. 518-526. doi:10.1109/TIP.2003.821445

[4] Nguyen., H.T. and Smeulders, A.W.M., 2004. Fast Occluded Object Tracking by a Robust Appearance Filter, IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 26, No. 8, pp. 1099-1104.

[5] Chen, H.T., Liu, T.L. and Fuh, C.S., 2004. Probabilistic Tracking with Adaptive Feature Selection, Proceedings of International Conference on Pattern Recognition Washington DC, Vol. 2, pp. 736-739. doi:10.1109/TPAMI.2004.45

[6] Batur, A.U. and Hayes, M.H., 2005. Adaptive Active Appearance Models, IEEE Transactions on Image Processing Vol. 14, No. 11, pp. 1707-1721doi:10.1109/TIP.2005.854473

[7] Corcoran, P., Ionita, M.C., and Bacivarov, I., 2007. Next Generation Face Tracking Technology Using AAM Techniques, Proceedings of International Symposium on Signals, Systems and Circuits, Vol 1, pp. 1-4. doi:10.1109/ISSCS.2007.4292639

[8] Bassili, J.N., 1979. Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of face, Journal of Personality and Social Psychology, Vol. 37, 2049-2059.

[9] Yacoob, Y. and Davis, L., 1994. Recognizing human facial expressions from long image sequences using optical flow, IEEE Trans. Pattern Anal .Machine Intel. 16 (6), 636–642.

[10] Otsuka, T. and Ohya, J., 1998. Extracting facial motion parameters by tracking feature points, Proceedings of First International Conference on Advanced Multimedia Content Processing, Osaka, Japan, pp. 442–453.

[11] Fasel, B. and Luettin, J., 2000. Recognition of asymmetric facial action unit activities and intensities, Proceedings of the International Conference on Pattern Recognition, Barcelona, Spain.

[12] Pardàs, M., Bonafonte, A. and Landabaso, J., 2002. Emotion recognition based on MPEG4 facial animation parameters, in: Proceedings of IEEE ICASSP.

[13] Cohen, I., Sebe, N., Cozman, F., Cirelo, M. and Huang, T., 2003. Learning Bayesian network classifiers for facial expression recognition using both labelled and unlabeled data, in: Proc. of the IEEE CVPR.

[14] Susskind, J.M., Littlewort, G., Bartlett, M.S., Movellan, J., and Anderson, A.K., 2007. Human and computer recognition of facial expressions of emotion, Neuro psychologia, vol. 45, pp. 152–162.

[15] Geetha, A., Ramalingam, V. and Palanivel, S., 2009. Facial expression recognition, A real time approach, Expert Systems with Applications, vol. 36, pp. 303–308

[16] Ma, L., and Khorasani, K., 2004. Facial expression recognition using constructive feed forward neural networks, IEEE Trans. Syst. Man Cybern.

[17] Seyedarabi, H., Aghagolzadeh, A., and Khanmohammadi, S., Recognition of Six Basic Facial Expressions by Feature-Points Tracking.

[18] Dubuission, S. Davoine, F., and Masson, M., 2002. A solution for facial expression representation and recognition, Signal Processing, Image Communication, vol.17, pp.657-673.

[19] Chen, X. and Huang, T., 2003. Facial expression recognition: A clustering-based approach, Pattern Recognition Letters, vol. 24, pp.1295–1302.

[20] Viola, P., and Jone, M., 2001. Rapid object detection using a boosted cascade of simple features Proceedings of IEEE computer society conference on computer vision and pattern recognition Kauai, 8-14 December, Vol. 1, p. 511.

[21] Otsuka, T., and Ohya, J., 1998. Spotting segments displaying facial expression from image sequences using HMM, IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition (FG'98), Nara, Japan, pp. 442–447.

[22] Cohn, J., Zlochower, A., Lien, J., Wu, Y., and Kanade, T., 1997. Automated face coding: a computer-vision based method of facial expression analysis, Seventh European Conference on Facial Expression Measurement and Meaning, Salzburg, Austria, pp. 329–333.

[23] Oliver, N., Pentland, A., Berard, F., 1997. LAFTER: a real-time lips and face tracker with facial expression recognition, Proceedings of the IEEE Conference on Computer Vision (CVPR97), S. Juan, Puerto Rico.