International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 8 (2018) pp. 6081-6088

Facial Expression Recognition using Moments Invariants and Modified Hidden Markov Model

[1] Mayur Rahul, [2]Dr. Narendra Kohli and [3]Dr. Rashi Agarwal

[1]*Research Scholar, AKTU, Lucknow, India.*

[2]*Professor, HBTU, Kanpur, India.*

[3]*Assistant Professor, University Institute of Engineering and Technology, CSJM University, Kanpur, India.*

Abstract

Facial Expression Recognition is an application used for biometric software that can be used to identify special expressions in a digital image by comparing and analysing the different patterns. These software are popularly used for the purpose of security and are commonly used in other applications such as home security, human-computer interface, credit card verification, surveillance systems, medicines etc.. Recognizing faces becomes very difficult when there is a change occurs in facial expressions. In this paper two layer extension of HMM is used to recognized continuous effective facial expressions. Moments Invariants are used for feature extraction. Two layered extension of HMM consists of bottom layer which 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. In HMM, Baum-Welch, Viterbi and Forward Procedure methods are used for parameter estimation, calculating the optimal state sequence and for probability calculation of the observed sequence respectively. This proposed system consists of three level of classification. Output of the first level is used for the training purposes for the second level and further this level is used for the third level for testing. Six basic facial expressions are recognised i.e. anger, disgust, fear, joy, sadness and surprise. Experimental result shows that Proposed System performs better than normal HMM and has the overall accuracy of 84% using JAFFE database.

Keywords: HMM, Pattern Recognition, Human-Computer Interface, Acoustic state model, Moment Invariants, JAFFE database, Physiognomy

INTRODUCTION

The Facial Expression is the most popular technique for non-verbal communication in human life[25]. Recognition of facial expressions can be used in various areas such as Human-computer interface, Telecommunications, Behavioural Science, Video Games, Animations, Psychiatry, Automobile Safety, Affect sensitive music juke boxes and televisions, Educational Software, etc. Feature Extraction techniques and Classifiers are very important part of the Facial Expression Recognition. They both are responsible for better classification and identification of the object. Better feature extraction method and better classifier significantly improves the performance of the classifiers.

The Hidden Markov Model is very useful in analysis of speech pattern using acoustic state model[9]. The HMM becomes the most powerful in speech and pattern recognition. HMM is also very useful in classifying unknown feature vector sequence due to its handling of time series data and its learning capability. It also requires small amount of data to train full HMM[10,11]. 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 [12,13].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 paper, moments invariants based feature extraction method and two layer extension of HMM are introduced. Basic assumption behind this new modified HMM is that every facial expression is made by the combination of eyes, nose and lips movements. This new modified HMM consists of two layers. Bottom layer represents the atomic expressions made by eyes, nose and lips. Upper layer represents the combination of these atomic expressions. Newly introduced system is capable of handling almost every critical situation such as occlusions, dynamic background, different transformations etc.

Rest of the paper is organized as follows: Basics of HMM is discussed in section-2, Moments invariants has been explained in section-3,Related works are discussed in section-4,Proposed System in section-5, Experiments and Results are discussed in section-6 and finally, concluded in section-7

BASICS OF HMM

Markov process is used for time series data modelling[11]. 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 aspoken sentence, the present pronounced word is depending on the previous pronounced word. Markov process effectively handles such situations.

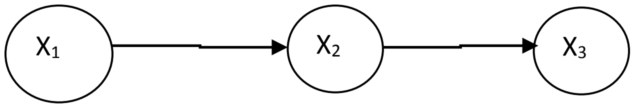


Figure 1: A Markov Model for 3 observations variables

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

$p(X_1,X_2,----,X_N) = p(X_1)$ Л $p(X_N|X_{N-1})$ (1)

where $p(X_1,X_2,------,X_N)$ is the joint probability distribution of states $X_1,X_2,----,X_N$ and p(X1) is the probability of state X_1 and $p(X_N|X_{N-1})$ is 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 (Figure 2). Hidden states are states above the dashed lines.

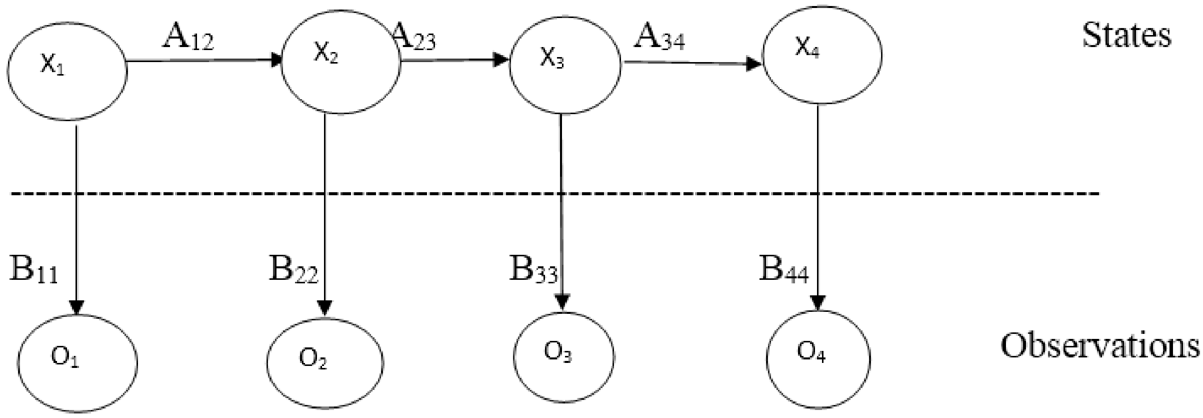


Figure 2: A hidden Markov Model for four variables

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

where $p(X_{t+1} = j \mid X_t = i) = aij$ is the conditional probability of the state X_{t+1} given the state X_t and a_{ij} is the ith row, jth column entry of the transition matrix $A = (a_{ij})$.

$p(O_t = k \mid X_t = l) = b_{lk}$ (3)

where $p(O_t = k \mid X_t = l) = b_{lk}$ is the conditional probability of the observation variable O_t given the state X_t and b_{lk} is the lth row, kth column entry of the emission matrix $B = (b_{lk})$.

An HMM is defined by A,B and Л . The HMM is denoted by ∂ = (A,B, Л)

MOMENTS INVARIANTS:

An important issue in the pattern recognition is the recognition of object and characters regardless of their orientation, position and size. The idea of using moments is very useful in pattern recognition. The moment invariants are invariant under different transformations such as scaling, rotation and shifting. Moments invariants are widely used in the area of pattern recognition. They can be derived with the help of various methods. Moments invariants was first introduced by the Hu[7] by using algebraic invariants and develop different type of moment invariants. Abu-Mostafa and Psatis showed the powerful nature of the moment invariants in pattern recognition[8]. They also showed the robustness and influence of moment invariants.

Two-dimensional (p+q)th order moment are defined by Hu[7]:

$m_{pq}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q \; f(x,y)\; dx\; dy$ p,q=0,1,2,3… (4)

If f(x,y) is a piecewise continuous bounded function, then all moments(m_{pq}) exists and can be calculated by f(x,y) and vice-versa.

The moments in (1) cannot be invariants when f(x,y) changes by rotating, scaling and translating. Invariants can be achieved by central moments, which are given by the equation:

$\mu_{pq}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\bar{x})^p(y-\bar{y})^q\, f(x,y)\; dx\; dy$ p,q=0,1,2,3… (5)

where

$\bar{x} = \frac{m_{10}}{m_{00}}$and $\bar{y} = \frac{m_{01}}{m_{00}}$

The image f(x,y) has the pixel ($\bar{x}$,$\bar{y}$) as the centroid.

The m_{pq}, whosecentre has been shifted to centroid of the image is equal to the centroid moments μ_{pq}computed using the centroid of the image f(x,y). Hence we can say that the central moments are invariants to the image translations.

Normalizations are used to obtained scale invariance. The normalised central moments can be obtained by:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\Upsilon}} \;, \Upsilon = (p+q+2)/2 \;, p+q = 2,3,4\ldots\ldots \quad (6)$$

Hu[1] introduced seven moment invariants based on the normalized central moments

$M_1 = \eta_{20} + \eta_{02}$

$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$

$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \mu_{03})^2$

$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \mu_{03})^2$

$M_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$

$M_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$

$M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$

The above seven moment invariants are the useful properties if there are no change in the scaling, rotation and translation.

Image geometric transformation is very useful in image processing, which consists of scaling, rotation and translation. The translation involves mapping of pixels from an input image to new position in the output image. The scaling is the process of changing the size of an image by applying interpolation to the input image. The rotation is achieved by rotating input image into some specified angle, which produces some non-integer coordinates. Different interpolation methods are used to generate the intensity of each pixel such as bi-linear interpolation, bi-cubic interpolation and nearest-neighbour interpolation.

From the last paragraph we can observe that the translation changes the position of pixels while scaling and rotation not only changes the position of the pixels but also its function. Hence scaling and rotation produces the errors in moment invariants. Further scaling and rotation changes the function itself then it also changes its moment invariants correspondingly. Therefore, moments invariants are very useful in facial expression recognition to detect changes in the face. Further these changes are useful in detecting different facial expressions.

RELATED WORKS:

Facial Expression Recognition is the basic topic in the field of Human-Computer interface. Facial expressions can be created by doing some actions on the facial muscles. The Facial Expression and Physiognomy was first introduced in the early fourth century. The Physiognomy is the study of the person's behaviour from their outer appearance [1]. The studies on the Physiognomy were minimized and Facial expression becomes popular. In 17th century, John Bulwer was given the detail theory about the facial expressions and head movement in his book "Pathomyotomia".Le Brun in 1667 gave a very effective notes on Facial Expression, then it was later published as a book in 1734[2].

Charles Darwin created a benchmark for various expressions and means of expressions in both animals and humans in the early 19th century[3]. The great job has been done by Ekman and his colleagues in 1970. His work has a great influence in today's 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 et. al. 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 introduced a framework in Psychology to recognize facial expression [4]. In 1990, researcher used Ekman's framework to recognize facial expression in video and image[5]. Hidden Markov Model was also used to recognised facial expressions [6]. Further Cohen et. al. proposed a multilevel HMM to recognise emotions and further optimize it for better accuracy as compared to emotion specific HMM[5].

Moment Invariants are widely used in the field of pattern recognition. Several methods have been used to derive moment invariants. Moments invariants were first introduced by the Hu [7] by using algebraic invariants and develop different type of moment invariants. Abu-Mostafa and Psatis showed the powerful nature of the moment invariants in pattern recognition [8]. They also showed the robustness and influence of moment invariants.

PROPOSED METHOD

Pattern Recognition is very important part of any image analysis system. Most of these systems contain four general steps: Image acquisition, pre-processing of the images, feature extraction, and classification[14]. Our goal is to improve classification results under various situations. Many approaches have been applied including feature selection, decision theory, learning etc.[15].An effective shape descriptors is very important in the description of multimedia content, since shape is a basic property of an object. Contour-based and region-based are the two shape descriptors [16]. Moments invariants are the contour-based shape descriptors developed by Hu(1962) consists of set of equations[7]. Further these moment invariants are extended to larger sets[17] and to some other forms[18,19]. Zion et. al. has developed an image processing algorithm, which have been used to discriminate between three fish species in the freshwater fish farms[20]. Shutler et. al. developed Zernike velocity moments which not only used as shape descriptor but also used as a motion descriptor[21].

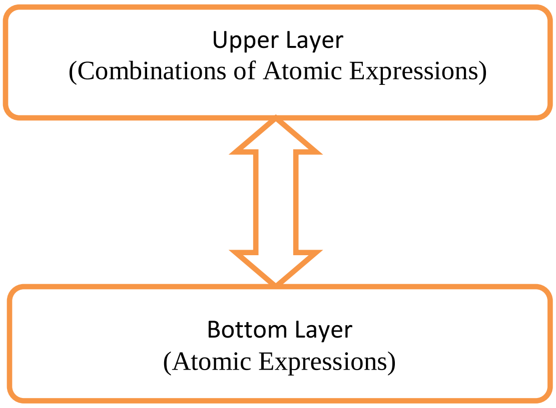


Figure 3: Layered Extension of HMM

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 [5,6].The discriminative strength of normal HMM is less suitable to handle more difficult task. In this paper, we introduce a new framework which uses most powerful moments invariants feature extraction method and modified HMM as a classifier. New extended 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.(see figure 3). 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 figure 4

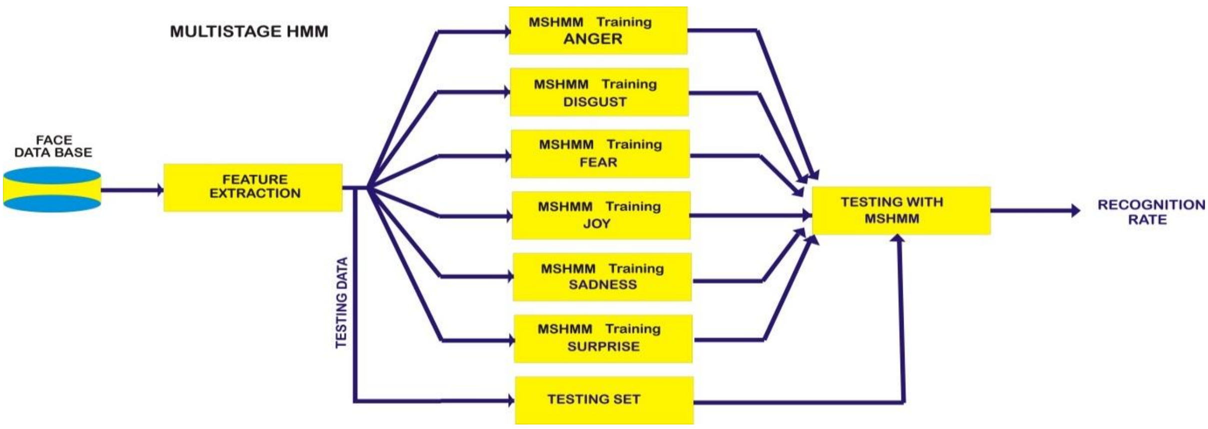


Figure 4: Block Diagram of Proposed System

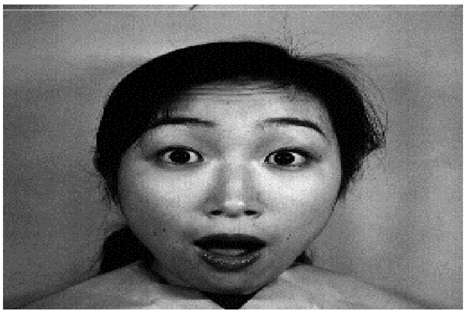
In this method, Moments Invariants are used as a feature extraction method and modified HMM as a classifier. Six universal expressions, i.e. anger, disgust, fear, joy, surprise and sadness, are recognised. HMM was introduced by Rabiner[22], Further it is used as a training HMM and recognition[23] . Baum-Welch method and Viterbi method are used for parameter estimation and finding the optimal state sequence respectively and Forward sequence is used for observed sequence probability. We have tested the original moments invariants with this modified HMM and get the better result as compared to previous work.

EXPERIMENTS AND RESULTS:

The images are considered under the following conditions:

(1) Only the frontal view of the given image is analysed.
(2) No head movement.
(3) Persons are not speaking.
(4) Persons have not facial hair and wearing glasses.

JAFFE database has the 114 images from 13 subjects[26]. They are 9 of anger, 21 of disgust, 13 of fear, 26 of joy, 24 of sadness and 21 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. There are 9 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 results. Also, take the moment invariants as feature vectors.



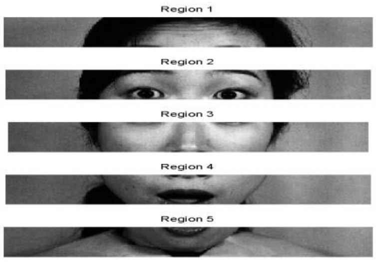


Figure 5: (a) Original Image **(b)** Divided into Regions

Every faces which are taken from JAFFE database can be divided into five parts. All parts are numbered from 1 to 5 and are in rectangular shape (see figure 5). These rectangular shapes can be adjusted to final shape so that every rectangular shape must contain the whole facial feature. The description of feature extraction region is shown in figure 5. The feature vectors are described as follows:

$N = [U_{i,j}]$, $i = 1,2,3,4,5$ $j = 1,2,3,4,5,6,7$

Where i represent the region number

And j represents the Moment Invariant.

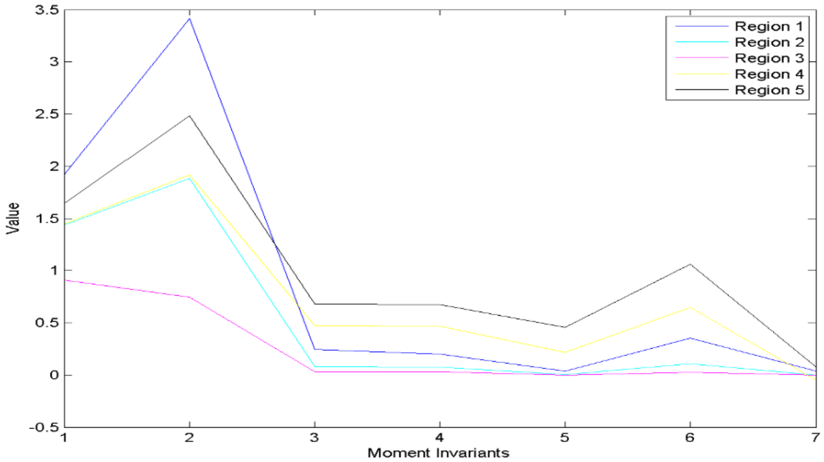


Figure 6: Moment Invariant of five region of same image

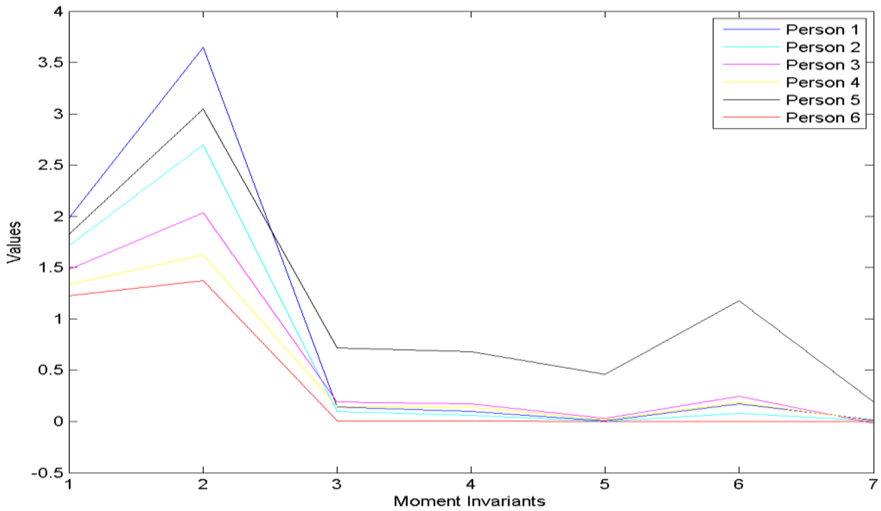


Figure 7: Variations in MI of region 1 with same expression disgust with different persons

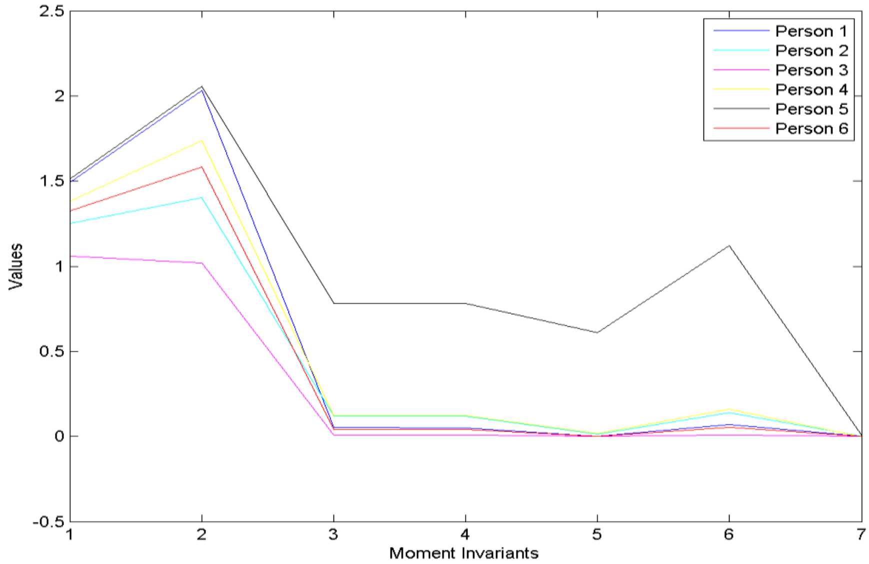


Figure 8: Variations in MI of region 2 with same expression disgust with different persons

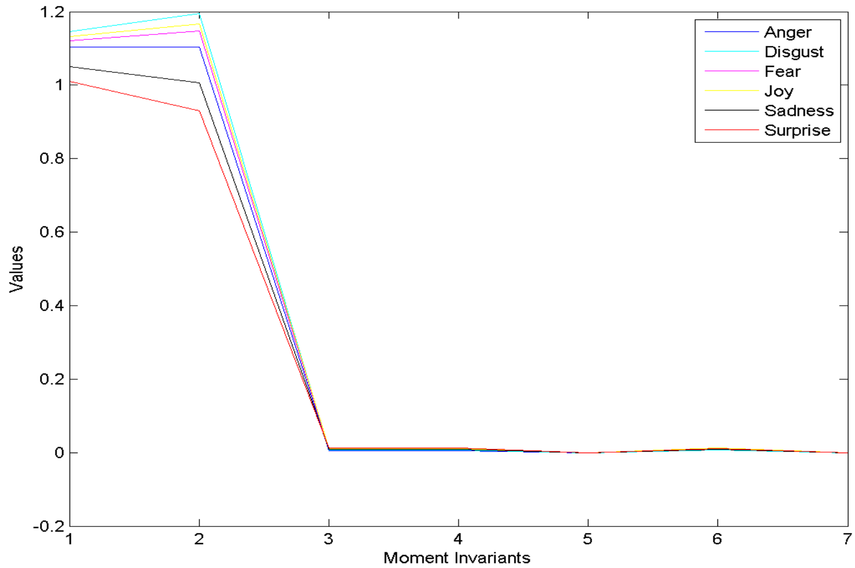


Figure 9: Variations in MI of region1 with six expressions from same person

The origin has no effect on calculating moment invariants while feature vector changes with the change in origin. The origin for the region 2 lies between two eyes. With the help of origin we can calculate ordinary moment invariants. Further this moment invariants are used to calculate feature vectors. The resultant feature vector shows the movements in facial features. They are very helpful in recognition.

Variations in the moment invariants of the various regions are the basic principle in this proposed algorithm (figure 6). Variations in the moment invariants are also seen in the same region of the same expression from the different subjects (figure 7 and 8). Variations are also seen in the same region with different expressions from same subjects(figure 8). These variations are helpful for recognition in facial expressions.

The five stages of every expression are separable easily by given plots(figure 6). The happy sequence consists of five stages i.e. neutral-front, transient-front, peak, transient-back and neutral-back. We use these stages in multistage HMM.

The recognition result using moments invariants are shown in table 1. From the results, we can see that the results from Zhu algorithm are less accurate as compared to proposed algorithm as shown in table 1.

Table 1: Comparison of recognition between Zhu and Proposed Algorithm

	Anger	Disgust	Fear	Joy	Sadness	Surprise	**Recognition Rate Zhu Algorithm[24]**	**Recognition Rate Proposed Algorithm**
Anger	7	0	1	0	0	1	60%	78%
Disgust	0	19	0	0	1	1	86%	91%
Fear	1	0	7	3	2	0	----	54%
Joy	0	0	0	26	0	0	78%	100%
Sadness	2	0	0	0	22	0	----	92%
Surprise	0	0	0	2	0	19	100%	91%

The Receiver Operation Characteristics (ROC) curve is a graph used to represent the performance of the classifier (Figure 5).It is the plot of True Positive Rate against False Positive Rate. TPR is the proportion of positive samples identified as positive and FPR is the proportion of negative samples identified as positive. In the given graph, the 45 degree line is called “line of no-discrimination”. The ROC curve above this line indicates that the classifier’s rate of identifying positive samples as positive is greater than the rate of misclassifying negative samples. Our curve indicates the goodness of our classifier to recognise Facial Expression Recognition.

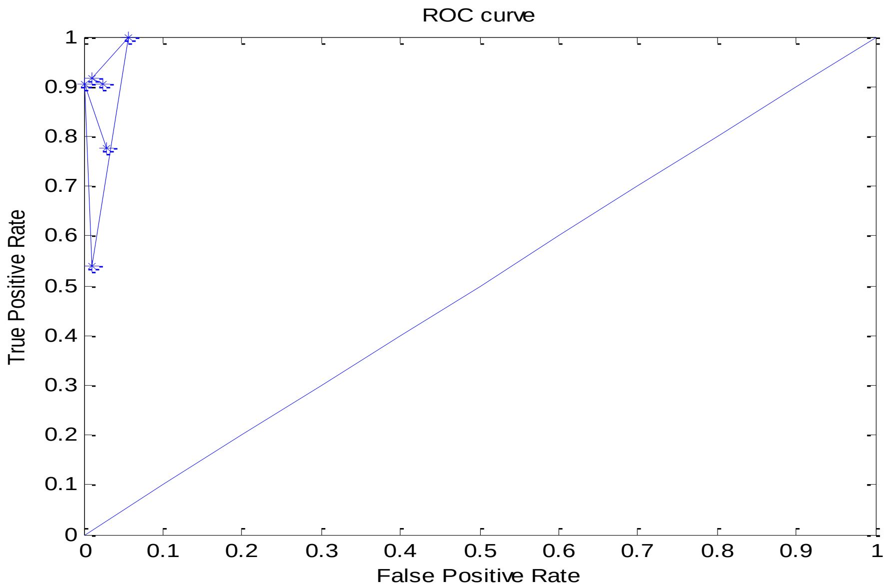


Figure 10: ROC curve for Confusion Matrix

CONCLUSIONS AND FUTURE WORKS:

This proposed system used the moment invariants with modified HMM to recognize six basic facial expressions. This system firstly derived linear Moment Invariants then modified HMM into two stages. This proposed system is able to identify facial movements and deformations of facial features easily. From the recognition results it can be observe that Moment Invariants with discrete HMM is less accurate then this proposed algorithm. This shows that both displacement and deformation of facial features are very important in facial expression recognition. This proposed algorithm makes HMM more powerful as a classifier with compared to other discriminative classifiers, also able to show its strength using Confusion Matrix and Receiver Operating Characteristics (ROC) curve and found the overall accuracy of 84%.

In the future, this modified HMM will be tested with other feature extraction methods such as Zernike moments, Rotational moments and Complex moments and compare it with other state-of-the-art methods and try to improve the accuracy to some extent.

REFERENCES

[1] R. Highfield, R. Wiseman and R.Jenkins, “How your looks betray your personality”, New Scientist, Feb 2009.

[2] http://www.library.northwestern.edu/spec/hogarth/physiognomics1.html

[3] C. Darwin, edited by F. Darwin, “The Expression of the Emotions in Man and Animals,”2nd edition, J. Murray, London, 1904.

[4] P. Ekman and W. Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, 1978.

[5] I. Cohen, N. Sebe, L. Chen, A. Garg, and T. S. Huang. Facial expression recognition from video sequences: Temporal and static modelling. In Computer Vision and Image Understanding, pages 160–187, 2003.

[6] J. J.-J. Lien, T. Kanade, J. F. Cohn, and C.-C. Li. Detection, tracking, and classification of action units in facial expression. Robotics and Autonomous Systems, 31(3):131–146, 2000.

[7] Hu. M. ”Visual Pattern Recognition by Moments Invariants”, IRE Trans. Inform. Theory , IT8,179-187,1962

[8] A. Mostafa, Psatis D., ”Recognition aspects of Moments Invariants” IEEE Trans. In Pattern Analysis and Machine Intelligence 6, 698-706, 1984

[9] Shigeru Katagiri; Chin-Hui Lee. A new hybrid algorithm for speech recognition based on HMM segmentation and learning vector quantization [J]. Speech and Audio Processing, IEEE Transactions on Volume 1, Issue 4, Oct. 1993 Page(s):421 - 430

[10] Tsapatsoulis.N; Leonidou.M.;Kollias.S.Facial expression recognition using HMM with observation dependent transition matrix [J]. Multimedia Signal Processing, 1998 IEEE Second Workshop on 7-9 Dec. 1998Page(s):89-95

[11] Lawrance R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceedings of the IEEE, 1 989, 77 (2) : 257—286.

[12] Atlas L, Ostendorf M, Bernard G D. Hidden Markov Models for Monitoring ,Machining Tool-Wear [C], In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000,6: 3887—3890

[13] Hatzipantelis E, Murray A, Penman J. Comparing hidden Markov models with artificial neural network architectures for condition monitoring applications[C], Fourth International Conference on Artificial Neural Networks, 1995,6:369—374.

[14] Khotanzad A, Lu J-H,”Classification of invariant image representations using a neural network.”, IEEE Trans. Acoust., Speech Signal Process. 38: 1028–1038,1990

[15] Kim Y K, Han J H Fuzzy K-NN algorithm using modified K-selection. Proc. Int. Joint Conf. 4th IEEE International Conf. on Fuzzy Systems (FUZZ-IEEE/IFES’95) 3: 1673–1680,1995

[16] Kim W-K, Sung Y, “A region-based shape descriptor using Zernike moments”, Signal Process. Image Commun.16: 95–102,2000

[17] Wong W H, Siu W C ,” Improved digital filter structure for fast moment computation”, IEEProc. Vision, Image Signal Process. 46: 73–79,1999

[18] Dudani S A, Breeding K J, Mcghee R B,”Aircraft identification by moment invariants”, IEEE Trans. Comput.C-26: 39–46,1977

[19] Liao S X, PawlakM,”On the accuracy of Zernike moments for image analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20: 1358–1364,1998

[20] Zion B et al ,”Sorting fish by computer vision”,Comput. Electron. Agric. 23: 175–187,1999

[21] Shutler J D, Nixon M S,”Zernike velocity moments for description and recognition of moving shapes”,Proc. BMVC, pp 705–714,2001

[22] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, Vol. 77, No. 2, February 1989, http://www.cs.ucsb.edu/~cs281b/papers/HMMs%20-%20Rabiner.pdf

[23] Rabiner L.R.,Juang B.H.,” Fundamentals of Speech Recognition”Prentice-Hall, Eaglewood Cliffs, NJ,1993.

[24] Y. Zhu, L.C. DeSilva, C.C.Ko,” Using Moments Invariants and HMM in Facial Expression Recognition,” Pattern Recognition Letters,23(2002),83-91.

[25] M.Rahul, N.Kohli, R.Agrawal,” Facial Expression Recognition using Hidden Markov Model: A Review”, National Conference on RAICT, HBTU, Kanpur, March 25-26,2017,ISBN:978-93-86256-82-9.

[26] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, JiroGyoba.”Coding Facial Expressions with Gabor Wavelets”, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998).