ISSN: 2321-7782 (Online)
Volume 2, Issue 11, November 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com
Special Issue: International Conference on Advanced Research Methodology Held by The International Association for Scientific Research & Engineering Technology, India

A Comprehensive Study on Sequence-based Classifiers for Facial Expression Recognition

J.Suneetha[1]
Research Scholar,
S.P.M.V.V, Tirupathi, INDIA

Dr. K.Sandhyarani[2]
Professor, Department of Computer Science,
S.P.M.V.V, Tirupathi, INDIA

Abstract: Facial expressions are widely used in the behavioral interpretation of emotions, cognitive science, and social interactions. The Facial Expression Recognition process performed by computers which consist of face detection from image, extracting the facial features and classifying the facial expressions such as fear, sad, disgust, anger, surprise, happy. So Facial Expression Recognition has been one of the challenging and active research topic in the recent years. This paper presents classification techniques for fully automatic facial expression recognition in facial image sequences.

Keywords: Facial Expression Recognition, Action Units (AU), Classifiers.

I. INTRODUCTION

A human face carries a lot of important information while interacting to one another. In social interaction, the most common communicative source is one's facial expression. Automatic facial expression recognition and analysis has been an active topic in the scientific community for over two decades [1]. Mainly in psychology, the expressions of facial features have been largely considered. Recent psychological research has shown that facial expressions are the most expressive way in which humans display emotion. As per the study of Mehrabian [2], the verbal part of the message contributes only 7% of the effect of the message as a whole, and the vocal part 38%, while facial expression contributes 55% . Therefore, automated and real-time facial expression recognition would be useful in many applications, e.g., human-computer interfaces, virtual reality, video-conferencing, customer satisfaction studies, face animation, etc. in order to achieve the desired result.

The earliest research on facial expression recognition can trace back to 19 century. In 1971, Ekman and Friesen [3] classified expressions into six universal emotions such as happy, sad, surprise, fear, disgust, anger. Every emotion has a corresponding specific expression. Expression classification requires supervised training, so the training set should consist of labeled data. Once the classifier is trained, it can recognize input images by assigning them a particular class label. The most commonly used facial expressions classification is done both in terms of Action Units, proposed in Facial Action Coding System (FACS) and in terms of six universal emotions: happy, sad, anger, surprise, disgust and fear defined by Ekman [4].

a) *Facial Action Coding System (FACS)*:

In 1978, Ekman et al. [5] introduced the system for measuring facial expressions called FACS – Facial Action Coding System. The FACS was developed to help psychologists with face behavior analysis. The process of expression analysis using FACS, decomposing observed expression into the set of Action Units (AU). There are 46 AUs that represent changes in facial expression. The AUs are classified as lower level and upper level. Each AU is associated with the physical behavior of specific facial muscle, giving an appearance-based description of faces. So facial expression corresponds to the combination of a set of AUs. In 1998 James J. Lien and Takeo Kanade developed Automatic facial expression recognition system based on FACS Action Units[6].

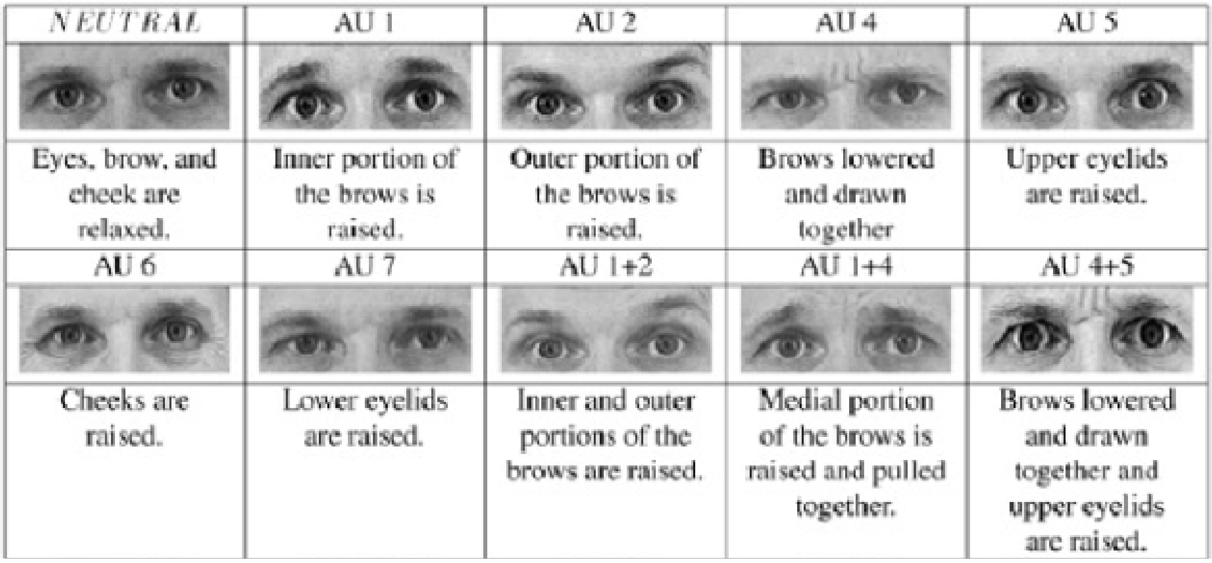

Fig 1: Action units and combinations

b) *Prototypical Facial Expression:*

Instead of describing the detailed facial features most facial expression recognition system attempt to recognize a small set of prototypical emotional expressions. According to the Ekman's theory [4], there are six basic emotion expressions that are universal for people of different nations and cultures. Those basic emotions are disgust, fear, happy, surprise, sad and anger as shown in Fig 2.

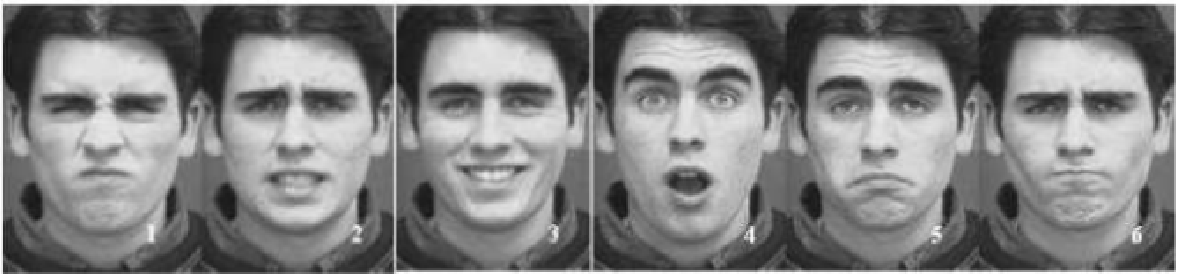

Fig 2: Six Universal Facial Expressions

Challenge in face expression recognition system It has already been stated that face expression recognition techniques have always been a very challenging task for researches because of all difficulties and limitations. The challenges associated with face expression recognition can be attributed to the following factors:

Pose: The images of a face vary due to the relative camera face position such as frontal or non-frontal and side view or others. Face may have a different angle so some of facial features such as an eye or the ears may become partially or wholly occluded. To overcome this challenge implements good pre-processing techniques which are invariant to translation, rotation and scaling.

Illumination: If the images are taken in different lights. Then expression feature can be detected inaccurately and hence recognition rate of facial expression is low.

Motion blur: significant blur can obscure the face if the camera exposure time is set too long or the head moves rapidly.

Occlusion: Faces may be partially occluded by other objects. In an image if face is occlude by some other faces or objects such as mask, hair, glasses etc.

II. Facial Expression Recognition

Facial expression analysis deals with visually recognizing and analyzing different facial motions and facial feature changes. Ekman and Friesen [3] developed the facial action coding system (FACS) to measure the facial behavior. The FACS codes different facial movements into Action Units (AU) based on the underlying muscular activity that produces momentary changes in the facial expression. An expression is further recognized by correctly identifying the action unit or combination of action units related to a particular expression.

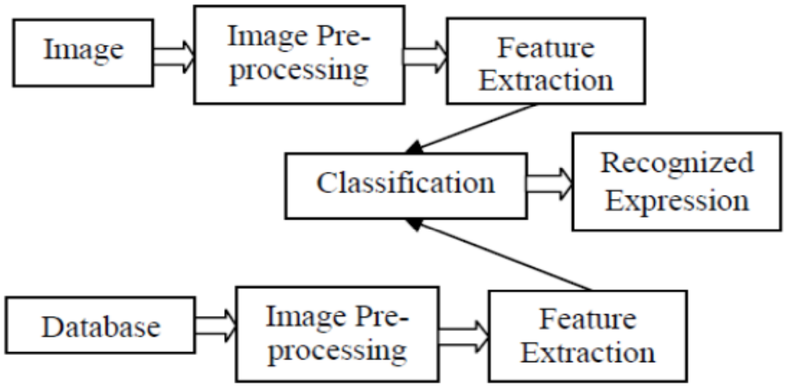

Fig 3: Block Diagram of Facial Expression Recognition Methodology

The Fig 3 shows the block diagram of Facial Expression Recognition (FER) Methodology [14]. In this there are three main factors to construct a Facial Expression Recognition system, namely Face detection, Facial feature extraction, and Emotion classification. An ideal emotion analyzer should recognize the subjects regardless of gender, age, and any ethnicity. The system should be invariant to different lightening conditions and distraction as glasses, changes in hair style, facial hair, moustache, beard, etc[10].

The final phase of automatic facial expression analysis systems is to recognize facial expressions based on the extracted features. The input to the classifier is a set of features which were retrieved from face region in the previous stage. The set of features is formed to describe the facial expression. A good face expression recognition system must consist the features such as it must be fully automatic, able to recognize spontaneous expressions, work in real-time, capability to work with video feeds as well as images, robust against different lighting conditions and able to recognize expressions from frontal, profile and other intermediate angles.

The facial expression recognition methods are divided into two categories such as Frame-based recognition and Sequence-based recognition. The Frame-based recognition method is based on static images and sequence-based recognition method is based on dynamic video images.

2.1 Frame-based Expression recognition

Frame-based Expression recognition uses only the current frame with or without a reference image. It does not use temporal information for the input images. The input image can be a static image or frame. So many methods are used such as Neural networks(NN),Support vector machine(SVM),Bayesian network, Linear discriminant analysis(LDA), rule-based classifiers etc are used for facial expression recognition[10].

2.2 Sequence-based Expression Recognition

Sequence-based Expression recognition uses the temporal information of the sequence to recognize the expressions for one or more frames. Hidden Markov Models (HMM), recurrent neural networks and rule based classifiers use sequence-based Expression Recognition[13].Sequence-based Expression Recognition classification schemes divided into two types such as dynamic and static classification. The static classifiers are classifiers that classify a frame in the video to one of the facial expression categories based on the tracking results of that frame. Dynamic classifiers are classifiers that take into account the temporal pattern in displaying facial expression. A multi-level HMM classifier is used for dynamic classification.

III. Sequence Based Classification Approaches of Facial Expression Recognition

There are mainly two approaches such as static and dynamic are taken in the literature for learning classifiers for emotion recognition.

3.1 The Static Approach:

In the static Approach, the classifier classifies each frame in the video to one of the facial expression categories based on the tracking results of that frame. Bayesian network classifiers were commonly used in this approach. Naive Bayes classifiers were also used often. Because of this unrealistic approach some used Gaussian classifiers and Gaussian Tree-Augmented Naive (TAN) Bayes classifiers.

3.1.1 Bayesian Network Classifiers

Bayes classifier is popular in pattern recognition because it is an optimal classifier. It is possible to show that the resultant classification minimizes the average probability of error. Bayes classifier is based on the assumption that information about classes in the form of prior probabilities and distributions of patterns in the class are known. It employs the posterior probabilities to assign the class label to a test pattern. a pattern is assigned the label of the class that has the maximum posterior probability. The classifier employs Bayes theorem to convert the prior probability into posterior probability based on the pattern to be classified, using the likelihood values.

A Bayesian network classifier represents the dependencies among features and labels by a directed acyclic graph. This graph is the structure of the Bayesian network. Typically, Bayesian network classifiers are learned with a fixed structure so the paradigmatic example is the Naive Bayes classifier. More flexible learning methods allow Bayesian network classifiers to be selected from a small subset of possible structures — for example, the Tree-Augmented-Naive-Bayes structures [7]. After a structure is selected, the parameters of the classifier are usually learned using maximum likelihood estimation.

Given a Bayesian network classifier with parameter setΘ, the optimal classification rule under the maximum likehood (ML) framework to classify an observed feature vector of n dimensions. $X \in R^n$, to one of |C| class labels, c € {|1,……,|C|},is given as

$$\hat{c} = \underset{c}{argmax} \ P(X|c;\Theta)$$

There are two design decisions when building Bayesian network classifiers. The first is to choose the structure of the network, which will determine the dependencies among the variables in the graph. The second is to determine the distribution of the features. The features can be discrete, in which case the distributions are probability mass functions. The features can also be in continuous, in which case one typically has to choose a distribution, with the most common being the Gaussian distribution.

3.1.2 Naive Bayes Classifiers

A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes theorem where every feature is assumed to be class-conditionally independent. Naive-Bayes classifiers have a very good record in many classification problems, although the independence assumption is usually violated in practice. The reason for the Naive-Bayes success as a classifier is attributed to the small number of parameters needed to be estimated. Abstractly, the probability model for a classifier is a conditional model $p(C|F1, \ldots, Fn)$,over a dependent class variable C with a small number of outcomes or classes, conditional on several feature variables $F1$ through Fn. The problem is that if the number of features is large or when a feature can take on a large number of values then working with such a model using probability tables is not feasible.

If the features in are assumed to be independent of each other conditioned upon the class label(the Naïve Bayes framework), so Bayes classifier equation reduces to:

 Published by: International Journal of Advance Research in Computer Science and Management Studies
Volume 2, Issue 11, November 2014 pg. 1-8

$$\hat{c} = \underset{c}{argmax} \prod_{i=1}^{N} P(x_i|c)$$

The Naive-Bayes classifier was successful in many applications mainly due to its simplicity. Also, this type of classifier is working well even if there is not too much training data. However, the strong independence assumption may seem unreasonable in our case because the facial motion measurements are highly correlated when humans display emotions. Therefore, when sufficient training data is available we want to learn and to use these dependencies.

3.1.3 Gaussian TAN Classifier

TAN classifiers have been introduced by Friedman et al. [7] and are represented as Bayesian networks. Bayesian networks are acyclic graphical models, with the class and features as the nodes, and dependencies are represented by the directed edges in the graph between the nodes. The joint probability distribution is factored to a collection of conditional probability distributions of each node in the graph.In the TAN classifier structure the class node has no parents and each feature has the class node as a parent and at most one other feature, such that the result is a tree structure for the features (see Figure). Friedman et al. [7] proposed using the TAN model as a classifier, to enhance the performance over the simple Naive-Bayes classifier.TAN models are more complicated then the Naive-Bayes, but are not fully connected graphs. The existence of an efficient algorithm to compute the best TAN model makes it a good candidate in the search for a better structure over the simple NB. Learning the TAN classifier is more complicated. In this case, we do not fix the structure of the Bayesian network, but try to find the TAN structure that maximizes the likelihood function given the training data out of all possible TAN structures.

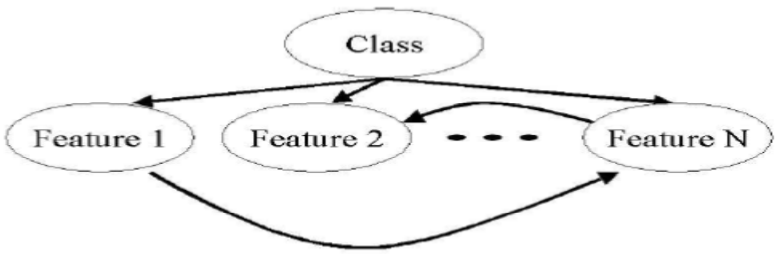

Fig 4: TAN Structure

The five steps of the TAN algorithm are described in Figure 5. This procedure ensures to find the TAN model that maximizes the likelihood of the data. The algorithm is computed in polynomial time with being the number of instances and the number of features).The learning algorithm for the TAN classifier as proposed by Friedman et al. [7] relies on computations of the class conditional mutual information of discrete features. In our problem the features are continuous, and computation of the mutual information for a general distribution is very complicated

1. Compute the class conditional pair-wise mutual information between each pair of features, (X_i, X_j) for all $i, j \in \{1, ..., n\}$,
$$I_P(X_i, X_j|C) = \sum_{X_i, X_j, C} P(x_i, x_j, c) log \frac{P(x_i, x_j|c)}{P(x_i|c)P(x_j|c)}, i \neq j.$$
2. Build a complete undirected graph in which each vertex is a variable, and the weight of each edge is the mutual information computed in Step 1.
3. Build a maximum weighted spanning tree (MWST).
4. Transform the undirected MWST of Step 3 to a directed graph by choosing a root node and pointing the arrows of all edges away from the root.
5. Make the class node the parent of all the feature nodes in the directed graph of Step 4.

Fig 5: TAN Learning Algorithm

For facial expression recognition, the learned TAN structure can provide additional insight on the interaction between facial features in determining facial expressions. In Figure 6 shows a learned tree structure of the features (Motion units) learned using

 ISSN: 2321-7782 (Online)
Special Issue: International Conference on Advanced Research Methodology Held by The International Association for Scientific Research & Engineering Technology, India

our database of subjects displaying different facial expressions. The arrows are from parents to children MUs. From the tree structure the TAN learning algorithm produced a structure in which the bottom half of the face is almost disjoint from the top portion, except for a weak link between MU 4 and MU 11.

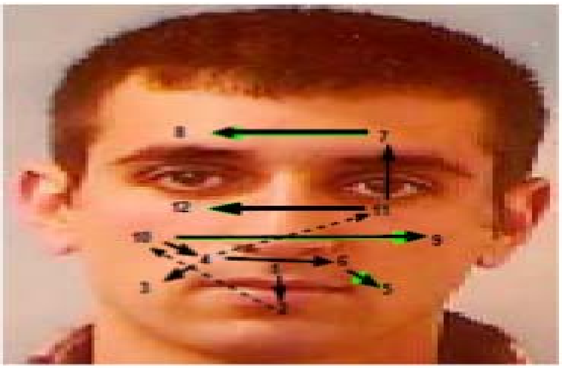

Fig 6: The Learned TAN Structure for thefacial features

3.2 The Dynamic Approach:

In the Dynamic Approach, these classifiers take into account the temporal pattern in displaying facial expression. When recognizing expressions from video, using the temporal information can lead to more robust and accurate classification results compared to methods that are static. Hidden Markov Model (HMM) based classifiers used for facial expression recognition.

3.2.1 Hidden Markov Models

Hidden Markov models (HMMs) are important in pattern recognition because they are ideally suited to classify patterns where each pattern is made up of a sequence of sub-patterns. For example, assume that a day is either *sunny*, *cloudy*, or *rainy* corresponding to three different types of weather conditions. Perhaps the most common application of HMM is in speech and speaker recognition tasks. One of the main advantages of HMMs is their ability to model nonstationary signals or events. Dynamic programming methods allow one to align the signals so as to account for the non stationarity. The HMM finds an implicit time warping in a probabilistic parametric fashion. It uses the transition probabilities between the hidden states and learns the conditional probabilities of the observations given the state of the model. In the case of emotion expression, the signal is the measurements of the facial motion. This signal is non stationary in nature, since an expression can be displayed at varying rates.

HMM Parameters The various elements are N=(number of states), M =number of distinct observation symbols, L = length of the observation sequence; the size of the training data set in terms of the number of states Using the training data for each class, So compute the corresponding HMM specified by $\lambda = (A,B, \pi)$, where

1. the matrix $A = [aij]$; $aij = P(St+1 = j|St = i)$, which stands for the probability of transition from state i to state j at time t
2. the observation probability $bj(k) = P(Ot = k|St = j)$, $1 \le k \le M$, is the probability of observing symbol k at time t ($Ot = k$) when the state is j
3. the initial state probability $\pi i = P(St1 = i)$, $1 \le i \le N$, the probability of starting in state i,

Assume that for each class under consideration, we are given a training data set of observations $O = (O1, O2, \ldots.., OL)$. We use these observations to learn λ. Some of the related problems are:

1. Estimate the HMM $\lambda = (A,B, \pi)$ that maximises $P(O|\lambda)$.
2. Given O and λ, find the optimal state sequence $S = (St_1, St_2, \ldots. , St_L)$.
3. Given λ, compute $P(O|\lambda)$, the probability of occurrence of O given the model.

3.2.2 Automatic Segmentation and Recognition of Emotions Using Multilevel HMM.

A Multi-level HMM classifier, combining the temporal information which allows not only to perform the classification of a video segment to the corresponding facial expression, as in the previous works on HMM based classifiers, but to also automatically segment an arbitrary long video sequence to the different expressions segments without resorting to heuristic methods of segmentation.

The main problem with the approach taken in the previous section is that it works on isolated facial expression sequences or on pre-segmented sequences of the expressions from the video. In reality, this segmentation is not available, and therefore there is a need to find an automatic way of segmenting the sequences. Concatenation of the HMMs representing phonemes in conjunction with the use of grammar has been used in many systems for continuous speech recognition.

Dynamic programming for continuous speech has also been proposed in different researches. It is not very straightforward to try and apply these methods to the emotion recognition problem since there is no clear notion of language in displaying emotions. Otsuka and Ohya [8] used a heuristic method based on changes in the motion of several regions of the face to decide that an expression sequence is beginning and ending. After detecting the boundaries, the sequence is classified to one of the emotions using the emotion-specific HMM. This method is prone to errors because of the sensitivity of the classifier to the segmentation result.

To solve the segmentation problem and enhance the discrimination between the classes, a different kind of architecture is needed. Figure 7 shows the proposed architecture for automatic segmentation and recognition of the displayed expression at each time instance. The motion features are continuously used as input to the six emotion specific HMMs[9]. The state sequence of each of the HMMs is decoded and used as the observation vector for the high level Markov model. The high-level Markov model consists of seven states, one for each of the six emotions and one for *neutral*.

The *neutral* state is necessary as for the large portion of time, there is no display of emotion on a person’s face. In this implementation of the system, the transitions between emotions are imposed to pass through the *neutral* state since our training data consists of facial expression sequences that always go through the *neutral* state. In unconstraint situation, it is possible (although less likely) for a person to go from one expression to another without passing through a neutral expression. In this case, the higher level Markov model will have non-zero transition probabilities of passing from all states to all states (which appear as arcs between the different states). The recognition of the expression is done by decoding the state that the high-level Markov model is in at each point in time since the state represents the displayed emotion [11, 12].

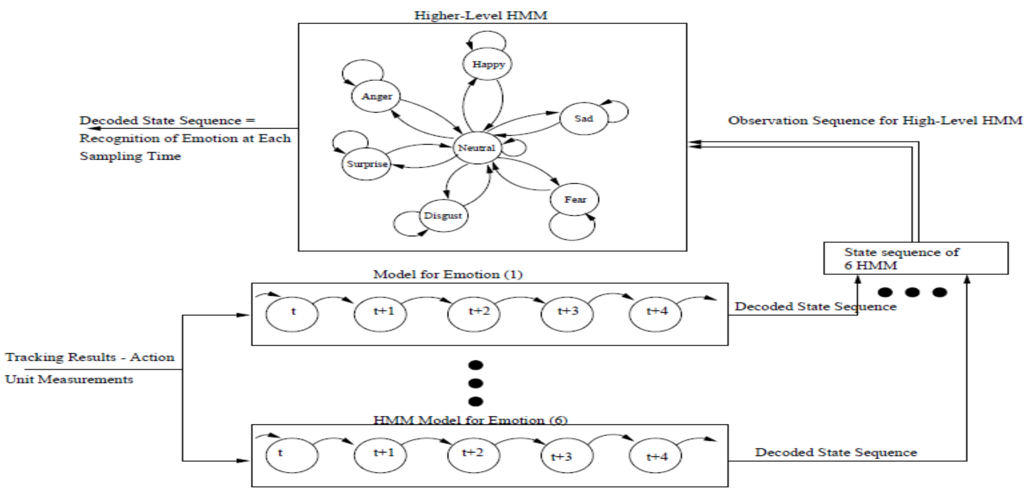

Fig 7: Multilevel HMM architecture for automatic segmentation and recognition of emotion.

In the case of dynamic classifiers the temporal information is used to discriminate different expressions. The idea is that expressions have a unique temporal pattern and recognizing these patterns can lead to improved classification results. So the multi-level HMM architecture and compared it to the straight forward Emotion-specific HMM. By using multi-level HMM architecture gives good results although it does not rely on any pre-segmentation of the video stream.

IV. Conclusion

Facial expressions play an important role in human interactions and non-verbal communication. The main objective of this paper is to summarizes sequence based classifiers for Facial Expression Recognition. In static approach the TAN classifier gives efficient result compare with Naive Bayes & Bayesian network classifiers.

Similarly in Dynamic approach the Multilevel HMM classifier gives good result compare with Single HMM classifier.

References

1. Valstar, M.F Mehu, M Jiang, B.Pantic, M.Scherer, "K. Meta-analysis of the first facial expression recognition challenge ", IEEE Trans. Syst. Man. Cybern. B Cybern. 2012, 42, 966–979.
2. Mehrabian, "A. Communication without words", Psychol. Today 1968, vol 2, pages 53–56.
3. P. Ekman, W.V. Friesen, "Constants across cultures in the face and emotion", J. Personality Social Psychol. 17(2) (1971) ,124-129.
4. Ekman, P, "Strong evidence of universal in facial expressions: A reply to Russell's mistaken critique", Psychol. Bull. 1994, 115, 268–287.
5. James J. Lien and Takeo Kanade, "Automated Facial Expression Recognition Based on FACS Action Units," IEEE published in Proceeding FG 98 in Nara Japan.
6. P. Ekman and W.V. Friesen., "Facial Action Coding System: Investigator's Guide", Consulting Psychologists Press, Palo Alto, CA, 1978.
7. N. Friedman, D. Geiger, and M. Goldszmidt., "Bayesian network classifiers Machine Learning ", 29(2):131–163, 1997.
8. T. Otsuka and J. Ohya. "Recognizing multiple persons' facial expressions using HMM based on automaticextraction of significant frames from image sequences". In Proc. Int. Conf. on Image Processing (ICIP-97),pages 546–549, Santa Barbara, CA, USA, Oct. 26-29, 1997.
9. T. Otsuka and J. Ohya., "A study of transformation of facial expressions based on expression recognition from temproal image sequences". Technical report, Institute of Electronic, Information, and Communications Engineers (IEICE), 1997.
10. M. Pantic and L.J.M. Rothkrantz. "Automatic analysis of facial expressions: The state of the art", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12):1424–1445, 2000.
11. Chellappa, R. Kruger, V.Shaohua Zhou, "Probabilistic recognition of human faces from video", 2002 International Conference on ImageProcessing, Vol 1, 2002, pp. 41-45.
12. J. Lien. "Automatic recognition of facial expressions using hidden Markov models and estimation of expression intensity". PhD thesis, Carnegie Mellon University, 1998.
13. M.Uma selvi and Professor Mr.P.Kannan , " A Survey on Robust Technique for Human Facial Expression Recognition", International Journal of Emerging Technologies and Engineering (IJETE)Volume 1 Issue 1 January 2014, ISSN 2348 – 8050
14. Nidhi N. Khatri , Zankhana H. Shah ,Samip A. Patel, " Facial Expression Recognition: A Survey", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 149-152.