Automatic Facial Feature Recognition and Facial Expression Analysis in Images

Sumalakshmi.C.H, Adithya V

***Abstract*—**The identification of facial activities from images have been of great interest in the field of computer vision. The facial activities are generally characterized by three levels. The bottom level is the facial feature identification, which recognizes the important facial feature points surrounding facial components (i.e., mouth, eyebrow, etc.), which gives the face shape information. The middle level is the facial action recognition, which determines the facial Action Units (AUs) defined in the Facial Action Coding System (FACS), identifying some significant facial activities (i.e., lid tightener, eyebrow raiser, etc.). The topmost level is the facial expression analysis which finds out the facial expressions that represent the human emotional states. The facial feature identification, AU recognition and expression recognition represent the facial activities in three levels, and they are mutually dependent. However, the recent techniques recognize the facial activities in one or two levels, and are identified separately, either discarding their interactions or limiting to one way. Unlike the conventional approaches, a Bayesian network model is proposed which also makes use of a feature based expression recognition model, to simultaneously recognise all the three levels of facial activities. The use of this feature based algorithm helps in increasing the accuracy of the expressions recognised. In the existing system we use the image measurements as well as the Bayesian network in order to recognise the facial features as well as the Action Units. Hence there is a great improvement in the efficiency of feature detection and AU recognition. But there is no separate method for the recognition of facial expression other than the Bayesian network. Moreover, unlike the existing system which uses 15 AUs, here 19 AUs are taken into consideration which will improve the efficiency of the Bayesian network based model. Given the facial action model and image observations, all three levels of facial activities are estimated simultaneously through a probabilistic inference by systematically integrating visual measurements with the proposed model.

Index Terms— **Bayesian network, expression recognition, facial action unit recognition, facial feature tracking, simultaneous feature recognition and expression analysis.**

Manuscript received March, 2014.

***Sumalakshmi.C.H**, Computer Science and Engineering, KMCT College of Engineering, Calicut, India, +919946877539.,*

***Adithya.V**, Computer Science and Engineering, KMCT College of Engineering, Calicut, India,+918129897716.,*

I.INTRODUCTION

The recognition of facial activities in image sequence is a significant and demanding problem. In recent years, a number of computer vision techniques have been developed to identify facial activities in three levels. First, in the bottom level, facial feature identification, which usually recognizes the prominent facial feature points surrounding facial components (i.e., mouth, eyebrow, etc.), captures the detailed face shape information. The middle level is the facial action recognition which recognizes the facial Action Units (AUs) as defined in the Facial Action Coding System (FACS) [2]. This level tracks some meaningful facial activities (i.e., lid tightener, eyebrow raiser, etc.). In the top level, facial expression analysis is performed which attempts to recognize the facial expressions that represent the human emotional states.

The facial feature identification, AU recognition and expression analysis represent the facial activities in three levels and they are interdependent problems. For example, facial feature tracking can be used in the feature extraction stage in expression/AUs recognition, and expression/ AUs recognition results can provide a prior distribution for facial feature points [1]. However, most of the current methods only recognize the facial activities in one or two levels, and track them separately, either ignoring their interactions or limiting the interaction to one way. In this paper, in contrast to the conventional approaches, we build a probabilistic model based on the Bayesian Network (BN) to capture the facial interactions at different levels. Thus the flow of information is two-way, not only bottom-up, but also top-down. In particular, not only the facial feature identification can contribute to the expression/AUs recognition, but also the expression/AU recognition helps to further improve the facial feature performance. Thus all three levels of facial activities are recovered simultaneously through a probabilistic inference by systematically combining the measurements from multiple sources at different levels of abstraction [1]. The proposed facial activity recognition system also uses an algorithm based method in order to find out the expression so that the expression recognized by the proposed system is based on two methods. Thus the result obtained by the proposed model produces a more accurate result. The system has two main stages: offline facial activity model construction and online facial motion measurement and inference. Specifically, using training data and subjective domain knowledge, the facial activity model is constructed offline. During the online recognition, we determine the facial feature points, and the AUs. These measurements are then used as evidence to infer the simultaneously recognize the facial activities.

The paper is divided as follows: In Sec. II, the steps involved in facial activity analysis are presented; Sec. III describes the details of facial activity modeling, Bayesian Network Based Expression Recognition(Sec.III-A); Algorithm Based Facial Expression Recognition System(Sec. III-B); The paper concludes in Sec. IV with a summary of our work and its future extensions.

.

II.STEPS INVOLVED

In this section, we are going to analyze the various steps involved in the proposed facial activity recognition system.

A. Facial Feature Identification

Facial feature points contain critical information about face shape and face shape deformation. Exact positioning and identification of facial feature points are crucial in the applications such as animation, computer graphics, etc. For extracting the facial features each of the facial component is segmented separately and the facial features are extracted for each of them. For each of the facial component, the skin region is extracted, we then determine the connected region, out of which the facial features are extracted by applying Bezier curves. The various facial points are then marked on the obtained facial features. In the proposed system we are using 25 facial points which are depicted in fig.1. The face model is composed of 27 features which are defined in correspondence with a set of 25 facial points.

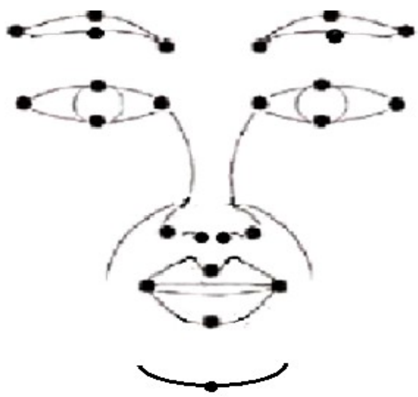

Fig 1. Facial feature points used in the algorithm.

B.AUs Recognition

Over the past decades, there has been extensive research on facial expression analysis. Here we use a method which focuses on recognizing facial actions by observing the representative facial appearance changes, usually try to classify expression or AUs independently and statically. Here we use 19 AUs as defined in FACS, in order to recognize the various facial expressions in a more effective manner[4]. The AUs are responsible for the identification of some significant facial activities (i.e., lid tightener, eyebrow raiser, etc.).

C. Expression Recognition

The expression analysis portion identifies six prototypical facial expressions namely happy, sad, anger, surprise, disgust and fear. In the proposed system we are making use of two methods in order to recognize the facial expression. The two methods employed for finding the expressions are as follows.

1) By using the Bayesian network to simultaneously identify the facial features and recognize the facial expressions. The facial features as well as Action units are used in developing the Bayesian Network. A Bayesian Network is a directed acyclic graph that represents the joint probability distribution among the variables. It is a unified probabilistic framework that simultaneously represents the facial activities. Dependencies are found by using the conditional probability. Thus by using this probabilistic Bayesian network facial expressions are determined. Given the facial action model and image observations, all three levels of facial activities are estimated simultaneously through a probabilistic inference by systematically integrating visual measurements.

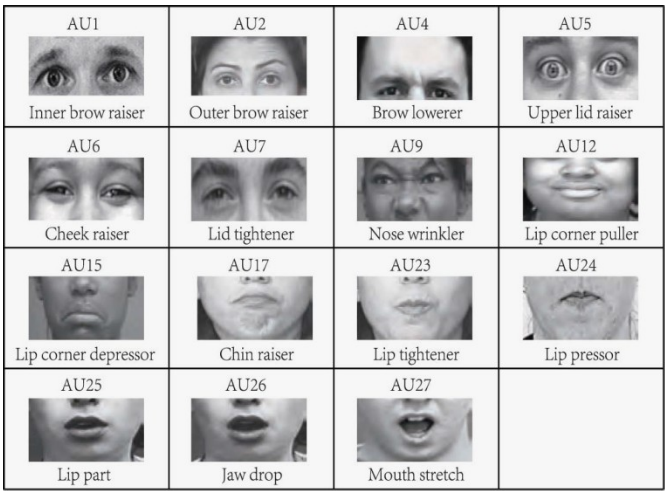

Table I. Some of the AUs and their interpretations

2) By using an algorithm based expression recognition system to identify the expression. In the proposed system we have defined 27 features whose measurement is used to compute the algorithm for expression recognition. The face model is composed of 27 features which are defined in correspondence with a set of 25 facial points[4]. This algorithm based method initially classifies expressions as those coming under 'with teeth' category and those under 'without teeth' category. The five expressions except the expression sad appear under 'with teeth' category whereas the without teeth category has all the six expressions.

III.FACIAL ACTIVITY MODELING

A .Bayesian Network Based Expression Recognition

The Bayesian network based expression recognition is the core of the proposed system. Here the input image is pre-processed by improving the contrast of the picture. We then perform the face detection by determining the connected skin region. In the measurement extraction phase the facial features and the Action Units are determined. This is used in developing the Bayesian network which in turn recovers the

facial features, the Action Units and the expression simultaneously. The Simultaneous Facial Activity Recognition using Bayesian Network is given in Fig.2. The Bayesian network is a directed graphical model and is more general to capture complex relationships among variables.

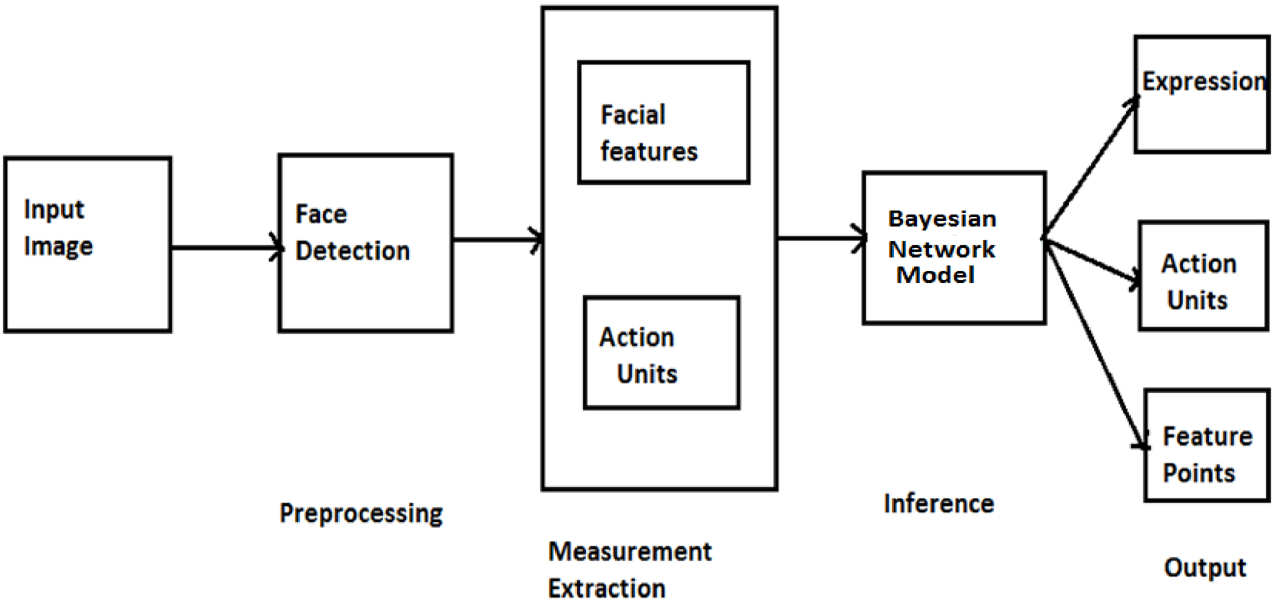

Fig. 2. Flowchart of the BN facial activity recognition system.

We propose to employ Bayesian network to model the spatial dependencies among all three levels of facial activities (facial feature points, AUs and expression) as shown in Fig. 3. Fig. 3 is not the final BN model, but a graphical representation of the causal relationships between different levels of facial activities. The *Et* node in the top level represents the current expression; *AUt* represents a set of AUs; *Xt* denotes the facial feature points we are going to track; *MAUt* and *MXt* are the corresponding measurements of AUs and the facial feature points, respectively. The three levels are organized hierarchically in a causal manner such that the level above is the cause while the level below is the effect. Specifically, the global facial expression is the main cause to produce certain AU configurations, which in turn cause local muscle movements, and hence feature points movements. For example, a global facial expression (e.g., Happiness) dictates the AU configurations, which in turn dictates the facial muscle movement and hence the facial feature point positions [1].

We mainly focus on identifying 6 basic expressions, i.e., happiness, surprise, sadness, fear, disgust and anger. The dependency between each level is used to find out the simultaneous facial activity. Though psychologists agree presently that there are ten basic emotions, most current research in facial expression recognition mainly focuses on six major emotions, partially because they are the most basic, and culturally and ethnically independent expressions and partially because most current facial expression databases provide the six emotion labels [1]. Given the measurement sequences, all three level facial activities are estimated simultaneously through a probabilistic inference via Bayesian network.

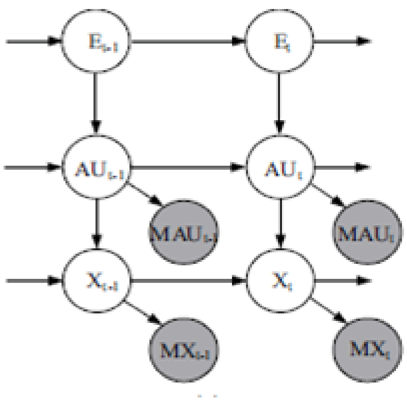

Fig 3. Bayesian Network facial activity model.

1) Modeling the Relationships Between Facial Features and AUs

In this paper, we will track 27 facial features and recognize 19 AUs, i.e., AU1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13,14,15,20, 23, 24, 25, 26 and 27.The selection of AUs to be recognized is mainly based on the AUs occurrence frequency, their importance to characterize the six expressions, and the amount of annotation available [4].

The 19 AUs we propose to recognize are all most commonly occurring AUs, and they are primary and crucial to describe the six basic expressions. They are the most commonly used facial expressions. Even though 19 AUs are used in this paper, the proposed framework is not restricted to recognizing these AUs alone, given an adequate training data set.

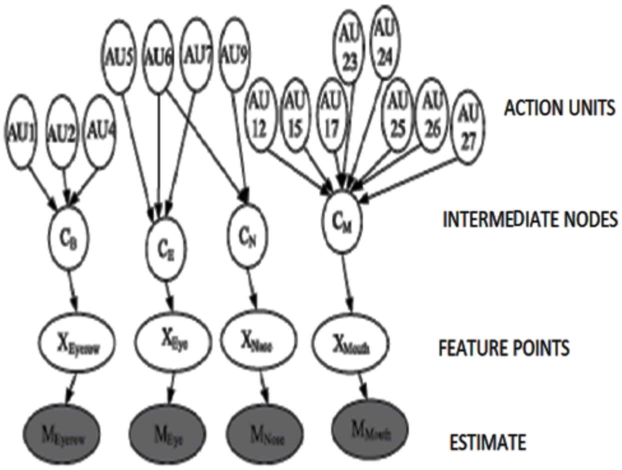

Fig.4. Modeling the relationships between facial feature points and AUs

There should be a minimum of 15 AUs to be chosen for proper recognition of six major expressions. The movement of face components is mainly managed by the Facial Action Units and it also controls the movement of facial feature points.

For instance, activating AU1 (inner brow raiser) results in a raising the inner eyebrow portion; and activating AU6(cheek raiser) raises the cheek upward. At the same time, the deformation of facial feature points reflects the action of AUs. Therefore, we could directly connect the related AUs to the corresponding feature points around each facial component to represent the casual relationships between them. Take *Mouth* for example, we use a continuous node *X*Mouth to represent 8 facial feature points around mouth, and link AUs that control mouth movement to this node[1]. However, directly connecting all related AUs to one facial component would result in too many AU combinations, most of which rarely occur in daily life. As a result, we introduce an intermediate node, e.g., "*CM*," to model the correlations among AUs. Fig. 4 shows the modeling for the relationships between facial feature points and AUs for each facial component. Each AU node has two discrete states which represent the "presence/absence" states of the AU. The intermediate nodes (i.e., "*CB*," "*CE*," "*CN* ," and "*CM*") are discrete nodes, each mode of which represents a specific AU/AU combination related to the face components. The Conditional Probability *P(Ci |pa(Ci))* for each intermediate node *Ci* is set manually based on the data analysis, where *pa(Ci)* represents the parents of node *Ci* . For instance, "*CB*" has five modes, each of which represents the presence of an AU or AU combination related to the eyebrow movement.

2)Modeling the Relationships Between AUs and Expression

In this section, we will add *Expression* node at the top level of the model [1]. Expression represents the global face movement and it is generally believed that the six basic expressions (happiness, sadness, anger, disgust, fear and surprise) can be described linguistically using culture and ethnically independent AUs, e.g., activating AU6+AU12+AU25 produces happiness expression, as shown in Fig. 5(a).

We group AUs according to different expressions[4]. Several combination of AUs result in different expressions. Generally, grouping of AUs belonging to the same category increases the degree of belief in classifying to that category, as shown in Fig. 5(a) (the combination of AU6 and AU12 increases the likelihood of classifying as happiness).However, combining AUs across different categories may result in the following situations: First, an AU combination belonging to a different facial expression, e.g., when AU1 occurs alone, it indicates a sadness, and when AU5 occurs alone, it indicates a surprise, however, the combination of AU1 and AU5 increases the probability of fear as shown in Fig. 5(b); Second, shows increased ambiguity, e.g., when AU26 (jaw drop), an AU for surprise, combines with AU1, an AU for sadness, the degree of surprise is reduced and the ambiguity of classification may be increased as illustrated in Fig. 5(c).

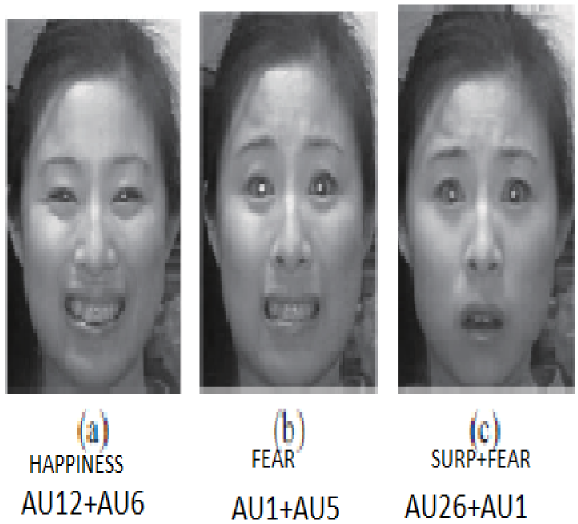

Fig. 5. AU combinations. (a) AU12+AU6 (two AUs from the same category) enhances classification to happiness. (b) AU1+AU5 (two AUs from different categories) becomes a fear. (c) AU26+AU1 (two AUs from different categories) increases ambiguity between a surprise and a fear.

These relationships and uncertainties are systematically represented by our final facial activity model as shown in Fig. 6. At the top level of the final model, we introduce six expression nodes, (i.e., Surp, Sad, Ang, Hap, Dis and Fea), which have binary states to represent "absence/presence" of each expression. We link each expression node to the corresponding AUs. Expressions are inferred from their relationships with AUs and reasoning over time. In principle, a probabilistic combination of any relevant facial AUs is taken in order to determine the corresponding facial expression.

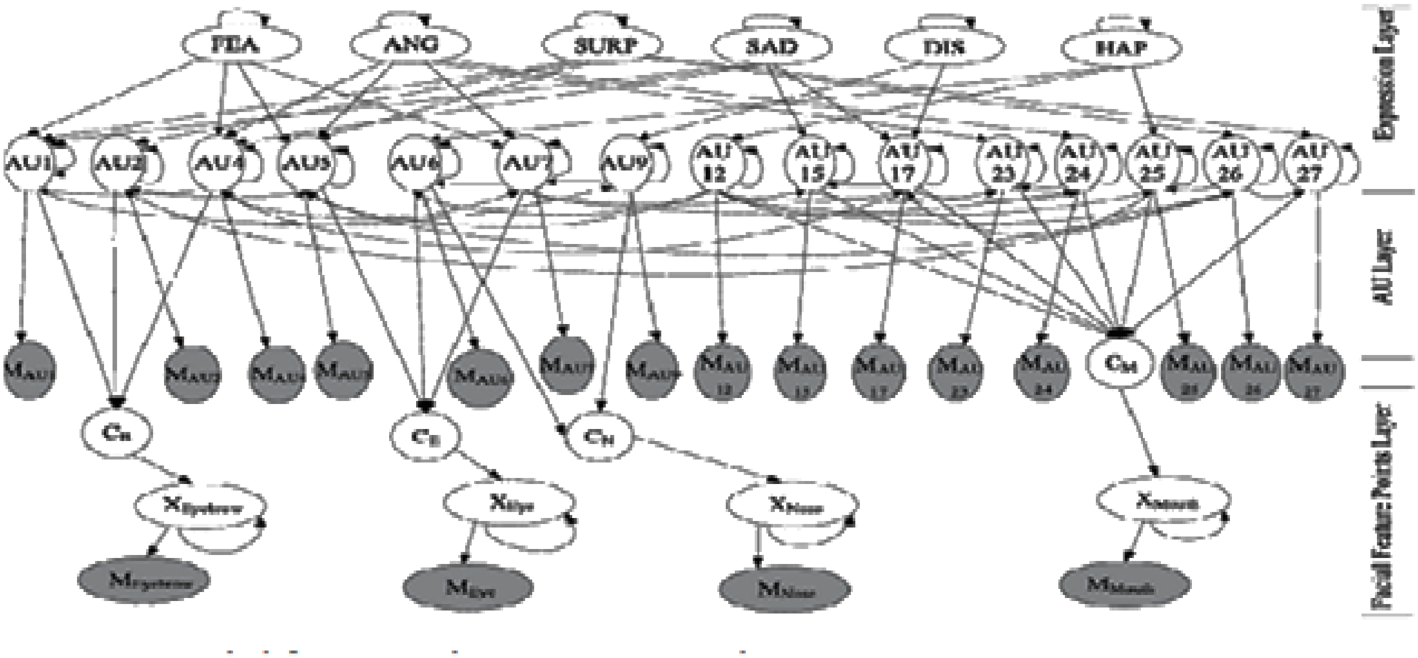

Fig. 6. Complete BN model for simultaneous facial activity recognition.

B .Feature based Algorithm for facial expression recognition

The value of the various facial features is taken in order to determine the various expressions. This algorithm based method classifies expressions as those coming under 'with teeth' category and those under 'without teeth' category. The five expressions except the expression sad appear under with teeth category whereas the without teeth category has all the six expressions. The changes happening in various facial features, over the different facial expressions are studied by comparing their values on different facial images. Study is also carried out on the images of the people without any particular expression. This learning helps to attain a detailed idea regarding the feature point values. Depending upon these feature values we recognize the various expressions. For example, for the expression surprise to be active, the condition chosen are raised eyebrows, increased eye width, the eye and eyebrow distance also increases. Likewise we have feature conditions for each of the expressions. The value of the features is compared against the threshold values which are obtained from the experimental study. In this method the expression determination depends upon the facial feature values. Some of the significant changes in the features for each of the different expression is Table II.

Expression	Feature Changes
Happy	Teeth presence, Increased lip length
Anger	Frowned brows, Raised brows, Increased eye width, Lip tightened
Fear	Frowned brows, Mouth opened, Increased eye width
Sad	Frowned brows, Lowered brows
Disgust	Frowned brows, Lowered brows, Reduced eye width
Surprise	Mouth opened, Increased eye width, Raised brows

Table II .Some Of the Feature changes implemented in algorithm based system.

IV.EXPERIMENTS AND RESULTS

The experiment for evaluating the expression was conducted in 3 steps. At first we conduct the experiment using the Bayesian based network model alone. We have considered an image set containing 100 images with all six expressions. These 100 images are given as input , one by one to the Bayesian network based expression recognition system. It was noted that, out of the 100 images, 90 of them turned out to produce correct expression results, whereas 10 of them showed variation from the actual expression condition.

The same experiment was carried out by taking the feature based expression recognition system alone. Here also the same 100 images were given as input to the system one by one. Out of the 100 images that we supplied 86 of them produced correct expression results, whereas 14 of them showed deviations from the actual expression condition.

From the first two steps it could be inferred that the efficiency of the recognition based on Bayesian network worked better compared to that of feature based algorithm. Therefore in the expression comparison procedure of the proposed system we give more preference to the Bayesian network based method. Here in the Bayesian system instead of 15 Aus (as in the existing system), here 19 Aus have been selected, providing a better efficiency. In the proposed system we initially recognize the feature based algorithm method followed by the Bayesian network based method. If the result of both the methods are the same then the output is chosen as this expression, and is considered to be 100 percent accurate. In cases where both the methods are producing different results we choose the output of the Bayesian network based system since it showed better performance than the other, this result is hence considered to be 90 percentage efficient.

Finally we conduct experiments using the proposed system. The test is conducted using 100 images of which 90 of them showed correct results. This helps us to infer that 90 percentage of the output produced by the system is true.

V.CONCLUSION

A Bayesian network model is proposed which also makes use of a feature based expression recognition model, to simultaneously recognise all the three levels of facial activities. The use of this feature based algorithm helps in increasing the accuracy of the expressions recognised. This feature based expression recognition system identifies the expression based on the facial feature values. The results of both the system are compared to generate the output, which is more accurate. Unlike the existing system which uses 15 AUs, here 19 AUs are taken into consideration which will improve the efficiency of the Bayesian based model. By systematically representing and modelling inter relationships among different levels of facial activities; the Bayesian network based model achieves significant improvement for both facial feature identification and AU recognition as well.

Future Research

Multiple face detection and feature extraction could be further improved. Since the current system can deal with some degree of lighting and orientation variation, the resolution of the image would be the main problem to concur for multi-person expression analysis. One direction to advance our current work is to combine the human speech and make both virtual and real robotic talking head for human emotion understanding and intelligent human computer interface, and explore virtual human companion for learning and information seeking.

REFERENCES

[1] Yongqiang Li, Shangfei Wang, Yongping Zhao, and Qiang Ji, "*Simultaneous Facial Feature Tracking and Facial Expression Recognition"* IEEE Transactions on Image Processing, Vol. 22, NO. 7, July 2013.

[2] P. Ekman and W. V. Friesen, "*Facial Action Coding System (FACS)":Manual*. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.

[3] Yan Tong, Wenhui Liao, Qiang Ji," *Facial Action Unit Recognition by Exploiting Their Dynamic and Semantic Relationships", "* IEEE Transactions on Pattern Analysis and Machine Intelligence .Vol. 29, No. 10, October 2007.

[4] M. Pantic*, L.J.M. Rothkrantz,"*Expert system for automatic analysis of facial expressions"* Image and Vision Computing 18 (2000).

[5] Maja Pantic,, and Leon J. M. Rothkrantz," *Facial Action Recognition for Facial Expression Analysis From Static Face Images" "* IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, Vol. 34, No. 3, June 2004.

[6] H. Yamada, Visual information for categorizing facial expression of emotion, Japan Psy-chology Review, vol. 35, pp. 172181, 1993.

[7] James A. Russell\Is There Universal Recognition of Emotion From Facial Expression? AReview of the Cross-Cultural Studies"Psychological Bulletin 1994, Vol. 115, No. 1, 102-141.

[8] P. Ekman, W.V. Friesen, Unmasking the Face, Prentice Hall, New Jersey, 1975.115 (1) (1994) 102141.emotions, Applied Cognitive Psychology 7 (1993) 257270.

[9] M. Valstar and M. Pantic, Fully automatic recognition of the temporal phases of facial actions, IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 42, no. 1, pp. 2843, Feb. 2012.

[10] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, Facial expression recognition from video sequences: Temporal and static modeling, Comput. Vis. Image Understand., vol. 91, nos. 12, pp. 160187, 2003.

.

Sumalakshmi.C.H. Pursuing M.Tech in Computer Science and Engineering from KMCT College of Engineering, Calicut,. Had worked as Lecturer in AWH Engineering College, Calicut for 5 years. Completed B.E in Computer Science and Engineering from Paavai Engineering College,Salem in the year 2006. Research Area is Image Processing and Computer Vision.

Adithya.V Working as Assistant. Professor in KMCT College of Engineering, Calicut. Completed M.Tech in Computer Science with Specialization in Image processing from College of Engineering, Chengannur. Completed B.Tech from Govt..Model Engineering College, Thrikkakara,Cochin. Publications are 1) 1. Adithya V, "Artificial Neural Network Based Method for Indian Sign Language Recognition", IEEE Conference on Information and Communication Technologies, ICT-2013.
2) Adithya V, "Contour Potential Energy Based Method for Indian Sign Language Recognition" ,International Journal of Scientific & Engineering Research (IJSER), Volume 4, Issue 6, June 2013.