Feature Tracking and Expression Recognition of Face Using Dynamic Bayesian Network

Sonali V.Hedaoo[#1], M.D.Katkar[*2], S.P.Khandait[#3]

[1]*Department of Computer Science and Engineering, Nagpur University,*
G.H. Raisoni Institute of Engineering and Technology for Women,
Nagpur-440034, India

[2]*Department of Computer Science and Engineering, Nagpur University,*
G.H. Raisoni Institute of Engineering and Technology for Women,
Nagpur-440034, India

[3]*Department of Computer Science & Engineering, Nagpur University*
K.D.K. College of Engineering,
Nagpur-440009, India

***Abstract*— The human face plays a central role in social interaction, hence it is not surprising that facial information processing is an important and highly active subfield of cognitive science research. The face is a complex stimulus displaying information about identity, age, gender, as well as emotional and attention state. Here we consider the problem of extracting information about emotional state (facial expression) from single images. Due to the difficulty of obtaining controlled video sequences of standard facial expressions, many psychological and neurophysiologic studies of facial expression processing have used single image motivations. In proposed system, in contrast to the mainstream approaches, we are trying to build a probabilistic model based on the Dynamic Bayesian Network (DBN) to capture the facial interactions at different levels. Hence the proposed system deal with the identification of facial expression on the image captured through camera**.

Keywords— Bayesian network, expression recognition, facial Action unit recognition, facial feature tracking, simultaneous Tracking and recognition.

I. INTRODUCTION

The recovery of facial activities in image sequence is an important and challenging problem. In recent years, plenty of computer vision techniques have been developed to track or recognize facial activities in three levels. First, in the bottom level, facial feature tracking, which usually detects and tracks prominent facial feature points (i.e., the facial landmarks) surrounding facial components (i.e., mouth, eyebrow, etc.), captures the detailed face shape information. Second, facial actions recognition, i.e., recognize facial Action Units (AUs) defined in the Facial Action Coding System (FACS) [1], try to recognize some meaningful facial activities (i.e., lid tightener, eyebrow raiser, etc.). In the top level, facial expression analysis attempts to recognize facial expressions that represent the human emotional states.

The facial feature tracking, AU recognition and expression recognition represent the facial activities in three levels from local to global, 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. However, most current methods only track or 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 addition, the estimates obtained by image-based methods in each level are always uncertain and ambiguous because of noise, occlusion and the imperfect nature of the vision algorithm.

In this paper, in contrast to the mainstream approaches, we build a probabilistic model based on the Dynamic Bayesian Network (DBN) to capture the facial interactions at different levels. Hence, in the proposed model, the flow of information is two-way, not only bottom-up, but also top-down. In particular, not only the facial feature tracking can contribute to the expression/AUs recognition, but also the expression/AU recognition helps to further improve the facial feature tracking performance. Given the proposed model, 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.

The proposed facial activity recognition system consists of 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, as shown in Fig. 1, various computer vision techniques are used to track the facial feature points, and to get the measurements of facial motions, i.e., AUs. These measurements are then used as evidence to infer the true states of the three level facial activities simultaneously.

II. LITERATURE SURVEY

In many paper number of algorithm are used for Facial Expression detection & Tracking . Michel F. Valstar, and Maja Pantic [1] proposed fully automatic method which not only allows the recognition of 22 AUs but also explicitly models their temporal characteristics (i.e., sequences of temporal segments: neutral, onset, apex, and offset). To do so, it uses a facial point detector based on Gabor-feature-based boosted classifiers to automatically localize 20 facial fiducially points. These points are tracked through a sequence of images using a method called particle filtering with factorized likelihoods. To encode AUs and their temporal activation models based on the tracking data, it applies a combination of GentleBoost, support vector machines, and hidden Markov models.

Ch. J. Wen Y. Zh. Zhan [2] had proposed HMM and KNN classifiers, and put forward a combined approach for facial expression recognition. The basic idea of this approach is to employ the classifiers of HMM and KNN in series way. First, DHMM classifier is used to calculate the probabilities of six expressions. Then, based on the most possible two results of classification by DHMM, the KNN classifier is used to make final decision while the difference between the maximum probability and the second is greater than the average difference.

Danijela Vukadinovic and Maja Pantic[2] had proposed Locating facial feature points in images of faces is an important stage for numerous facial image interpretation tasks. In this paper we present a method for fully automatic detection of 20 facial feature points in images of expressionless faces using Gabor feature based boosted classifiers. The method adopts fast and robust face detection algorithm, which represents an adapted version of the original Viola-Jones face detector. The detected face region is then divided into 20 relevant regions of interest, each of which is examined further to predict the location of the facial feature points.

Fadi Dornaika , Franck Davoine[4] had proposed facial actions/features are retrieved from the images, and then facial expressions are recognized based on the retrieved temporal parameters. Unlike this main stream, this paper introduces a new approach allowing the simultaneous recovery of facial actions and expression using a particle filters adopting multi-class dynamics that are conditioned on the expression. For each frame in the video sequence, their approach is split in two consecutive stages. In the first stage, the 3D head pose is recovered using a deterministic registration technique based on Online Appearance Models. In the second stage, the facial actions as well as the facial expression are simultaneously recovered using the stochastic framework with mixed states.

Fadi Dornaika, Frank Davoine [5] had proposed unconstrained face recognition problem and utilizing multiple cameras to improve the recognition accuracy using a probabilistic approach. They propose a dynamic Bayesian network to incorporate the information from different cameras as well as the temporal clues from frames in a video sequence. The proposed method is tested on a public surveillance video dataset with a three-camera setup and compared their method to different benchmark classifiers with various feature descriptors .This paper also state that modeling the face in a dynamic manner the recognition performance in a multi-camera network is improved over the other classifiers with various feature descriptors and the recognition result is better than using any of the single camera.

Yongqiang Li, Jixu Chen, Yongping Zhao, and Qiang Ji[6] Stated that besides the development of facial feature extraction techniques and classification techniques, prior models have been introduced to capture the dynamic and semantic relationships among facial action units. Previous works have shown that combining the prior models with the image measurements can yield improved performance in AU recognition. Most of these prior models, however, are learned from data, and their performance hence largely depends on both the quality and quantity of the training data. These data-trained prior models cannot generalize well to new databases, where the learned AU relationships are not present. To alleviate this problem, they propose a knowledge-driven prior model for AU recognition, which is learned exclusively from the generic domain knowledge that governs AU behaviors, and no training data are used.

G.Donato, M.S.Bartlett, J.C.Hager, P.Ekman and T.J.Sejnowski [9] explores and compares techniques for automatically recognizing facial actions in sequences of images. These techniques include: analysis of facial motion through estimation of optical flow; holistic spatial analysis, such as principal component analysis, independent component analysis, local feature analysis, and linear discriminant analysis; and methods based on the outputs of local filters, such as Gabor wavelet representations and local principal components. Performance of these systems is compared to naive and expert human subjects. Best performances were obtained using the Gabor wavelet representation and the independent component representation, both of which achieved promising accuracy for classifying 12 facial actions of the upper and lower face.

II. RELATED WORK

In this section, we are going to introduce the related works on facial feature tracking, expression/AUs recognition and simultaneous facial activity tracking/recognition, respectively.

A. Facial Feature Tracking

Facial feature points encode critical information about face shape and face shape deformation. Accurate location and tracking of facial feature points are important in the applications such as animation, computer graphics, etc. Generally, the facial feature points tracking technologies could be classified into two categories: model free and model-based tracking algorithms. Model free approaches [10,11] are general purpose point trackers without the prior knowledge of the object. Each feature point is usually detected and tracked individually by performing a local search for the best matching position. However, the model free methods are susceptible to the inevitable tracking errors due to the aperture problem, noise, and occlusion. Model based methods, such as Active Shape Model (ASM) [3], Active Appearance Model (AAM) [4], Direct Appearance Model (DAM) [5], etc.

In the conventional statistical models, e.g. ASM, the feature points positions are updated (or projected) simultaneously, which indicates that the interactions within feature points are interdependent. Intuitively, human faces have a sophisticated structure, and a simple parallel mechanism may not be adequate to describe the interactions among facial feature points. For example, whether the eye is open or closed will not affect the localization of mouth or nose. Tong *et al*. [8] developed an ASM based two-level hierarchical face shape model, in which they used multi-state ASM model for each face component to capture the local structural details. For example, for mouth, they used three ASMs to represent the three states of mouth, i.e., widely open, open and closed. However, the discrete states still cannot describe the details of each facial component movement, i.e., only three discrete states are not sufficient to describe all mouth movements. At the same time, facial action units inherently characterize face component movements; therefore, involving AUs information during facial feature points tracking may help further to improve performance.

B) Expression Recognition

Facial expression recognition systems usually try to recognize either six expressions or the AUs. Over the past decades, there has been extensive research on facial expression analysis Image-based approaches, which focus on recognizing facial actions by observing the representative facial appearance changes, usually try to classify expression or AUs independently and statically [1].

C. Simultaneous Facial Activity Tracking and Recognition

Compared to the previous related works, this paper has the following features:

1) First, we build a DBN model to explicitly model the two way interactions between different levels of facial activities. In this way, not only the expression and AUs recognition can benefit from the facial feature tracking results, but also the expression recognition can help improve the facial feature tracking performance.

2) Second, we recognize all three levels of facial activities simultaneously. 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.

III. RESEARCH METHODOLOGY

The proposed system will consists of following modules:

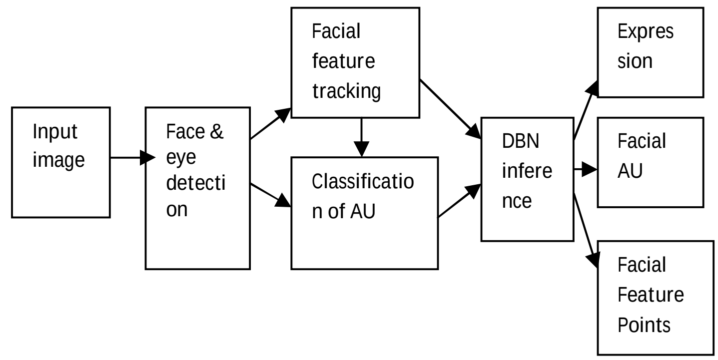

Fig1. Proposed System Architecture

1) *Image Capturing*

Firstly, we capture camera view then get the current frame out of it and load it in to memory for processing and then try to find the specified pixels RGB value then compare this RGB value with predefined value and once the value match then proceed further.

2) *Facial Feature Tracking:*

After capturing the image , the next step is to track the facial feature points. First face is detected in live video automatically then by dividing the face region in two parts i.e. lower & upper part we can determine the facial feature points based on their aligned position.

3) *Facial Action unit Recognition*

After extraction of feature points, the proposed system contains robust vision computer techniques to obtain AU measurement and such AU measurement are then applied as evidence to the DBN for inferring various AUs.AU measurement extraction is done by AdaBoost Classification and DBN inference. AdaBoost is a classification algorithm; it uses weak classifiers (anything that give more than 50% correct result, better than random). AdaBoost classifier combines the wavelet features to produce measurement score for each AU.

4) *Expression Recognition:*

Apart from facial feature tracking & AU recognition, next step is to construct a DBN structure for recognizing six global expression such as happiness, ,surprise, fear, disgust and anger.

5) *Database & Training*:

Lastly database is created for images of supplied face to store the list of feature point and their respective AU's and also training is provided to images to get accurate and fast detection.

IV. FACIAL ACTICVITY MODELING

A) *Image capturing*

As we know that image is made up of pixels and each pixel is made up of bits this bit pattern is depending on the file or image format. Each bit of pixel value for colour. For example in bitmap file format it use one bit for Red, one bit for Green on disk. and this logic is repeated for each frame. As we get the image then we try to extract the pixels from it and as we get the pixel we try to get the RGB colour value from the pixel as pixel made up from the RGB value of the colour. As RGB value get we compare the value & last we have colour.

For skin colour segmentation, first we contrast the image. Then we perform skin colour segmentation. Then, we have to find the largest connected region. Then we have to check the probability to become a face of the largest connected region. If the largest connected region has the probability to become a face, then it will open a new form with the largest connected region. If the largest connected regions height & width is larger or equal than 50 and the ratio of height/width is between 1 to 2, then it may be face.

B) *Face Detection*

For face detection, first we convert binary image from RGB image. For converting binary image, we calculate the average value of RGB for each pixel and if the average value is below than 110, we replace it by black pixel and otherwise we replace it by white pixel. By this method, we get a binary image from RGB image.

Then, we try to find the forehead from the binary image. We start scan from the middle of the image, then want to find a continuous white pixels after a continuous black pixel. Then we want to find the maximum width of the white pixel by searching vertical both left and right site. Then, if the new width is smaller half of the previous maximum width, then we break the scan because if we reach the eyebrow then this situation will arise. Then we cut the face from the starting position of the forehead and its high will be 1.5 multiply of its width.

C) *Eyes Detection*

For eyes detection, we convert the RGB face to the binary face. Now, we consider the face width by W. We scan from the W/4 to (W-W/4) to find the middle position of the two eyes. The highest white continuous pixel along the height between the ranges is the middle position of the two eyes.

Then we find the starting high or upper position of the two eyebrows by searching vertical. For left eye, we search w/8 to mid and for right eye we search mid to w – w/8. Here w is the width of the image and mid is the middle position of the two eyes. There may be some white pixels between the eyebrow and the eye. To make the eyebrow and eye connected, we place some continuous black pixels vertically from eyebrow to the eye. For left eye, the vertical black pixel-lines are placed in between mid/2 to mid/4 and for right eye the lines are in between mid+(w-mid)/ 4 to mid+3*(w-mid)/ 4 and height of the black pixel-lines are from the eyebrow starting height to (h- eyebrow starting position)/4. Here w is the width of the image and mid is the middle position of the two eyes and h is the height of the image. Then we find the lower position of the two eyes by searching black pixel vertically. For left eye, we search from the mid/4 to mid - mid/4 width. And for right eye, we search mid + (w-mid)/ 4 to mid+3*(w- mid)/ 4 width from image lower end to starting position of the eyebrow. Then we find the right side of the left eye by searching black pixel horizontally from the mid position to the starting position of black pixels in between the upper position and lower position of the left eye. And left side for right eye we search mid to the starting position of black pixels in between the upper position and lower position of right eye. The left side of the left eye is the starting width of the image and the right side of the right eye is the ending width of the image. Then we cut the upper position, lower position, left side and the right side of the two eyes from the RGB image.

D) *Lip Detection*

For lip detection, we determine the lip box. And we consider that lip must be inside the lip box. So, first we determine the distance between the forehead and eyes. Then we add the distance with the lower height of the eye to determine the upper height of the box which will contain the lip. Now, the starting point of the box will be the ¼ position of the left eye box and ending point will be the ¾ position of the right eye box. And the ending height of the box will be the lower end of the face image. So, this box will contain only lip and may some part of the nose. Then we will cut the RGB image according the box.So, for detection eyes and lip, we only need to convert binary image from RGB image and some searching among the binary image

E) *Further plan of work*

The further plan of work will

i) Model the relationship between facial features and AU's.
ii) Model the Semantic relationship among AU's
iii) Model relationship between AU's and Expression
iv) Modeling dynamic relationship using DBN dynamic structure

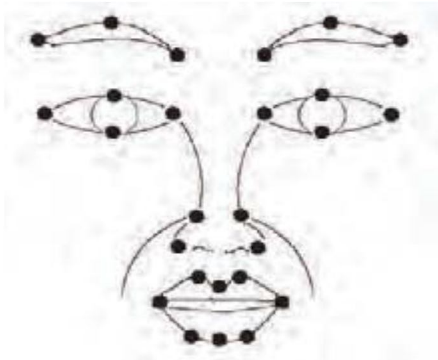

Fig2. Facial feature points used in the algorithm [1]

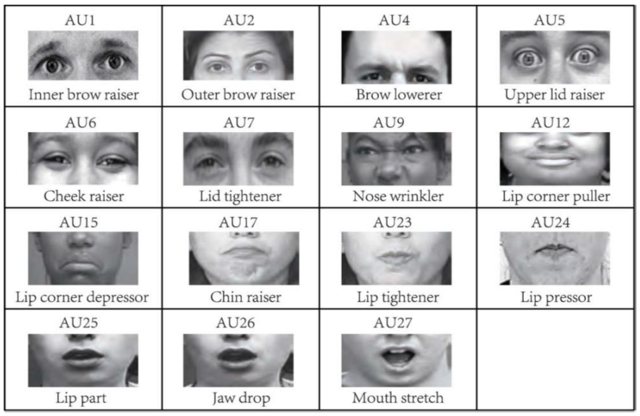

Fig 3. LIST OF AUS AND THEIR INTERPRETATIONS [1]

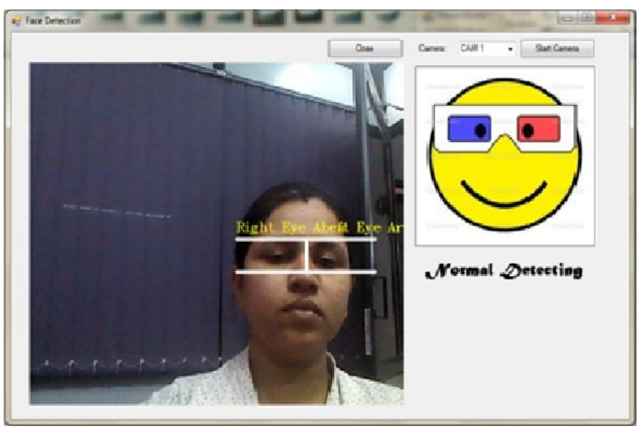

Fig 4. Face & Eye Detection

V.CONCLUSION

In this paper, we use Dynamic Bayesian Network(DBN) for simultaneous facial feature tracking and facial expression recognition. By systematically representing DBN structure we can model the relationship between different AU's. Thus we can get the expected outcome as Facial feature points and their respective Action Unit (AU) & Recognize Expression by modeling relationship between facial feature point & AU.

RERFERENCES

1) Yongqiang Li , Shangfei Wang,Yongping zhao , Qiang Ji, "*Simultaneous Facial Feature Tracking & Facial Expression Recognition*", IEEE Transaction On Image Processing, VOL. 22,NO. 7,PP. 2559-2571,July 2013

2) Michel F. Valstar,Maja Pantic, "*Fully Automatic Recognition of Temporal phases of Facial Action"*, IEEE Transactions On Systems, Man, And Cybernetics—part B: Cybernetics, Vol. 42, No. 1, PP. 28-42, February 2012

3) Ligang Zhang , Dian Tjondronegoro, "*Facial Expression Recognition using Facial Movement Features"* , IEEE Transactions On Affective Computing, Vol. 2, No. 4, PP.219-229,October-December 2011

4) Ch. J. Wen Y. Zh. Zhan, "*HMM+KNN Classifier for Facial Expression Recognition*", IEEE Transaction 2008,PP.260-263,2008

5) Fadi Dornaika , Frank Davoine , " *simultaneous Facial Action Tracking & Expression Recognition Using a Particle Filter"* ,*10th* IEEE International Conference on Computer Vision(ICCV'05), 1550-5495,2005

6) Le An, Mehran Kafai, Student Member, IEEE and BirBhanu, Fellow, IEEE, "*Dynamic Bayesian Network for Unconstrained Face Recognition in Surveillance Camera Networks*", IEEE Journal On Emerging And Selected Topics In Circuits And Systems, Vol. X, No. X, June 2013

7) Yongqiang Li, Jixu Chen, Yongping Zhao, and Qiang Ji, "*Data-free Prior Model for Facial Action Unit Recognition*",Submit To IEEE Transactions On Affective Computing

8) TimoAhonen, Student Member, IEEE,AbdenourHadid, and Matti Pietika¨ inen, Senior Member, IEEE, "*Face Description with Local Binary Patterns:Application to Face Recognition"*, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 28, No. 12, December 2006

9) G.Donato, M.S.Bartlett, J.C.Hager, P.Ekman and T.J.Sejnowski, "Classifying Facial Action", IEEE Transaction Pattern Anal. Mach. Intell , Vol.21,No.10,pp. 974-989,Oct. 1999.

10) Cassio P.de Campos and Q.Ji, " Efficient Structure learning of Bayesian Networks using Constraints", J.Mach Learn , Res,Vol.12,pp. 663-689,March 2011

11) M.Valstar and M.Pantic , "Fully automatic recognition of the temporal phases of Facial Action",IEEE *Trans.Syst., Man,Cybern. B, Cybern.*, Vol. 42, no. 1, pp 28-43,Feb.2012.

12) I.Patras and M.Pantic , "Particle Filtering with factorized likelihoods for tracking Facial Features," in *Proc. IEEE Int.Conf.Autom.Face Gesture Recognition,May 2004,p 97-102.*

13) S.P.Khandait , Dr.R.C.Thool, "*automatic facial feature extraction and expression recognition based on neural network"* (IJACSA) International Journal of Advanced Computer Science and Application , vol.2,no.1,January2011,pp. 113-118, ISSN : 2156-5570(Online) and ISSN : 2158- 107X(Print)

14) S.P. Khandait, Dr. R.C.Thool, P.D.Khandait, "*Comparative Analysis of ANFIS and NN Approach for Expression Recognition using Geometry Method*", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE) Vol.2,Issue 3,March -2012,pp. 169-174, ISSN: 2277 128X

15) S.P.Khandait, Dr. R.C.Thool and P.D.Khandait , "*ANFIS and Neural Network based Facial Expression Recognition using Curvelet Features*", International Journal of Computing, (ISSN 1727-6209) 2012, vol. 11, issue 3, p.255-261 .