2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Facial Expressions Recognition system using Bayesian Inference

Maninderjit Singh, Anima Majumder* and Laxmidhar Behera

***Abstract*—The paper presents a facial expressions recognition system using Bayesian network. We train the network using probabilistic modeling that draws relationship between facial features, action units and finally recognizes six basic emotions. We propose features extraction methods to get geometric feature vector containing angular informations and appearance feature vector containing moments extracted after applying gabor filter over certain facial regions. Both the feature vectors are further used to draw relationships among Action Units (AUs). The angular informations are directly extracted from the facial landmark points. The geometric features extraction approach contains only 22 dimensional angular informations against direct facial landmarks based approach that contains 136 dimensional feature vector. Facial activities are represented by three distinct layers. Bottom level contains landmark measurement data with angular features. Middle level has facial AUs those are coded in facial action coding system (FACS) and the top level, represents emotion node. We also propose a method using k-means clustering to automatically define the states of nodes in anatomical layer that draws relationship among AUs and measurement data. Extended Cohn Kanade Database is being used for our experimental purposes. An average emotion recognition accuracy of 95.7% is achieved using proposed Bayesian network based approach for 22 dimensional angular feature vector. To verify the performance of the proposed approach we apply three different classifiers such as, Support vector machine, Decision tree and Radial basis functions network. The confusion matrices show that the Bayesian network based classification approach outperforms all other applied approaches. The experimental results illustrates the effectiveness of the proposed model.**

***Index Terms*—Bayesian network, Probabilistic inference, Facial expressions recognition, Support vector machine, Radial basis function, Decision tree.**

I. Introduction

FACIAL expressions plays a fundamental role in human communication [1]. Through evolution, humans are very good at recognizing subtle changes in facial expressions which helps them to recognize the emotions and thereby the intent of other persons. But the task of recognizing emotions and extracting facial features is very challenging due to non-rigid motion of facial muscles. Moreover, our sensitivity towards subtle differences in emotions makes it even difficult to have a recognition system that could perform comparable to humans cognition system. Over the last few decades there have been a numerous research in this area to tackle the challenges of feature extraction and emotion recognition [2], [3], [4], [5], [6], [7], [8], [9] The six basic emotions those have been studied widely in this field are: anger, disgust, fear, happiness, sadness and surprise. But apart from these basic emotions, humans are capable of producing a wide variety of facial expressions those can be captured using Facial Action Coding system (FACS) developed by Ekman[10]. The FACS are consisting of codes called Action Units (AUs); they are assigned to different facial regions and their action. For example, AU2 represents outer eyebrow raiser, AU12 represents lip corner pullers etc. The traditional techniques used for AU classification like SVMs [11], [12], Decision trees [13] and Neural Networks [14], [15], [16] use spatial classification techniques those tend to classify each AU independently. This doesnt exploits the mutual dependencies of AUs. For example, occurrence of AU1 (inner brow raiser) significantly increases with the increase in probability of AU2 (outer brow raise) or AU23 (lip tightener) increases the likelihood of AU24 (lip presser). To efficiently model and capture the flow of evidence from top to bottom and bottom to top layer; i.e, among the emotions, AUs and features layers; we need to represent the system as a network that could capture the relationship between different variables. Moreover, by using probabilistic framework we could robustly deal with variety of scenarios, like missing information, occlusion and information fusion through different measurements. Thus, Bayesian Networks serves as an important tool that could capture the dependencies among the different AUs. It also allows us to incorporate prior knowledge into models more easily. We can build a hierarchical model consisting of several layers with data fusion at different levels of abstraction. The lowest layer consist of features extracted from important landmarks, the next upper layer represents the states of anatomical features like nose and eyes. Above this layer we have AUs, which are combination of different states of anatomical features and wrinkle patterns. At the top we have the emotion layer representing six basic emotions. This hierarchal structure allows us to easily evaluate the performance. This approach could also be easily extended to model the temporal relationships and thus it can further improve the accuracy. Also, instead of directly recognizing emotions from image features space, if we have a anatomical level that builds relationships between image features and AUs and then between AUs and emotion space; we can have a more

Maninderjit Singh is bachelor degree student at Department of Electrical Engineering, Indian Institute of Technology Kanpur, PIN 208016, Uttar Pradesh, India (email: maninder@iitk.ac.in)

Anima Majumder is doctoral student at Department of Electrical Engineering, Indian Institute of Technology Kanpur, PIN 208016, Uttar Pradesh, India (email: animam@iitk.ac.in)

Laxmidhar Behera is professor, Department of Electrical Engineering, Indian Institute of Technology Kanpur, PIN 208016, Uttar Pradesh, India (email: lbehera@iitk.ac.in)

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

detailed and accurate system by incorporating the knowledge at different levels of abstraction. The highest level consists of emotions like happiness, anger while at mid-level we have Action units and facial landmarks/ features at the bottom level. In the present work we develop a Bayesian network based facial expressions recognition model that has three different levels. First level is consisting of image features level followed by AUs level that draws relationships between different AUs and finally the top level is the six basic emotions level. The organization of the paper is as follows: Section II gives a brief explanation about the facial alignment method proposed in the work. It includes pose as well as shape alignment. Section III demonstrates an overview of Bayesian network. A brief description about the Bayesian structure applied in this work and the 3 different layer is indicated in this section of the work. Section IV shows the experimental results of the proposed approach. Comparative studies with 3 widely used classifiers such as Support vector machine (SVM), Decision tree and Radial Basis Functions Network (RBFN) are tabulated in this part. Finally, in section V conclusions are drawn.

II. Facial Landmarks alignment

We have used extended Cohn-Kanade dataset [17] which has 327 instances of image sequences from 123 subjects with annotated landmarks, peak emotions and AUs. The dataset consists of sequence of frontal face images where subjects emotions change from neutral to peak emotion state. The facial expression changes are often accompanied by head movements like backward head movement during fear and tilting of head during a smile. The face images also show great variability across different individuals. To effectively compare the facial expressions of different people it is important to compensate head motion and transform the facial landmarks in a common coordinate system. This is achieved by aligning landmarks through pose correction and shape normalization.

A. Pose alignment:

The images in our dataset consist of only frontal faces, so for pose compensation we have to consider only the in-plane rotation and the translation of faces. Four facial landmarks as shown in second Fig. 1 are considered as rigid points (two inner corners of the eyes and two the tip of the nose). They are used to compute a perspective transformation between the corresponding landmark points of a reference image and the image in consideration for features alignment. It is observed that the perspective transform provides better results than an affine transform. This is because, it is observed that in some cases there is out of plane head rotation as well.

The first image in Fig. 1 shows a tilted face image, second one shows the rigid landmarks used for alignment and the third image shows the aligned landmarks after transformation.

B. Shape Normalization:

The facial proportions differ from person to person. This makes it difficult to compare the expressions of people. As an example, a person with an oval face may have mouth landmarks lower than a person with round face. To accommodate this variability in shape, we compute the piece-wise affine transformation of neutral emotion image of each subject with a reference image. This computed set of affine transforms is used to normalize the corresponding image sequences with emotions. Algorithm 1 is used for piece-wise affine transformation:

Algorithm 1 Steps for shape normalization using affine transformation.

1: Make a triangulated mesh connecting different landmarks. For each triangle in neutral face calculate affine transformation in barycentric coordinates with corresponding triangle of reference landmarks.

2: For each set of vector X containing 3 points representing the triangle we calculate the affine transform matrix A that maps it to reference images corresponding set of points vector Y. Initially the inner corners as shown in 2^{nd} image of the Fig. 1 we assign a fixed coordinate position as they are rigid points. Using these 3 points as reference the points in local coordinates of triangle we calculate the position in common coordinate system.

$$AX = Y \tag{1}$$

3: For each subsequent frame, the landmark x is transformed by applying affine transform.

$$Ax + x^{'} = y \tag{2}$$

Where $x^{'}$ is one of the landmarks in triangle which has already been transformed to common coordinate system. The initial choice of x are the inner corners of eyes that have fixed position in common coordinate system.

III. Bayesian network

Bayesian network over variables Z is defined as a pair (G, θ), where: G is a directed acyclic graph over variables Z, called network structure and θ is a set of conditional probability densities (CPDs). θ is assigned for each of the variable in Z called network parameterization. Bayesian networks are compact way to represent probability distribution and utilize graph theoretic algorithms for learning and inference [18]. In this paper we model the system using three different levels such as: feature level, AUs level and emotions level. Thus the complete Bayesian network represents the joint distribution over these variables. This problem can be redefined as maximizing the joint probability of emotions (E) and Action units AU for given measurement data X.

$$E*, AU*, X* = argmax_{E,AU,X} P(E, AU|X) \tag{3}$$

where $E*, AU*$ and $X*$ are the optimal emotions action units for given measurement data; i.e., feature vector containing angular information.

Original image Reference rigid landmark Aligned landmarks

Fig. 1: An Example of facial landmarks alignment.

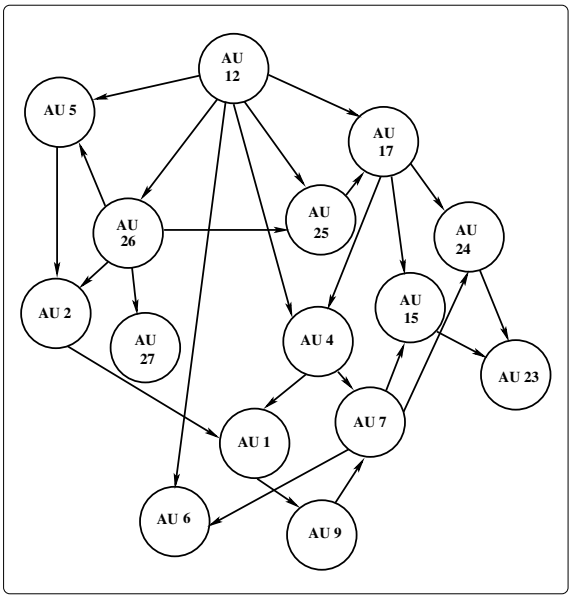

Fig. 2: Bayesian structure for learning action units

A. *Learning Bayesian Networks for inference*

Learning in Bayesian networks involves both learning the structure of the network as well as the parameters or conditional probabilities. In this work we applied the network structure proposed by Qiang Ji [5]. The structure is as shown in the Fig. 2 below. **Geometric Feature extraction** We are taking 68 landmark points for each facial image from Extended Cohn Kanade database. The 68 landmark points gives $2 \times 68 = 136$ dimensional feature vector as they represents the x, y coordinates for each of the point's location. This dimension is quite large and unnecessarily makes the system complex. We propose a new feature extraction method using angular information of certain landmark points as shown in the Fig. 4 below. The figure shows only 22 landmark points using which we are calculating the internal angles formed by polygons representing different facial key regions, like mouth, nose, eyes etc. Fig. 5 depicts the angular representation of the features obtained from 22 landmark points. As we have considerably reduced the dimensionality of our data, we need to ensure that the angular features are descriptive enough to capture the difference among different AUs. Therefore, we plot the distribution of AUs with different pairs of angles and found that AUs forms separable clusters in these subspace. Figure 3 shows some examples of AU plots with only two dimensional subspace AU. It is clearly observed that the AUs are easily separable.

All the images consist of frontal face poses, so angles are invariant features in this constrained environment. Currently we are working with only discrete variables so the angles measurements are discretized by using the distribution scale from 0 to 10 bins. In the entire database we calculate the variation of each angle and assigns bins by uniformly capturing the data

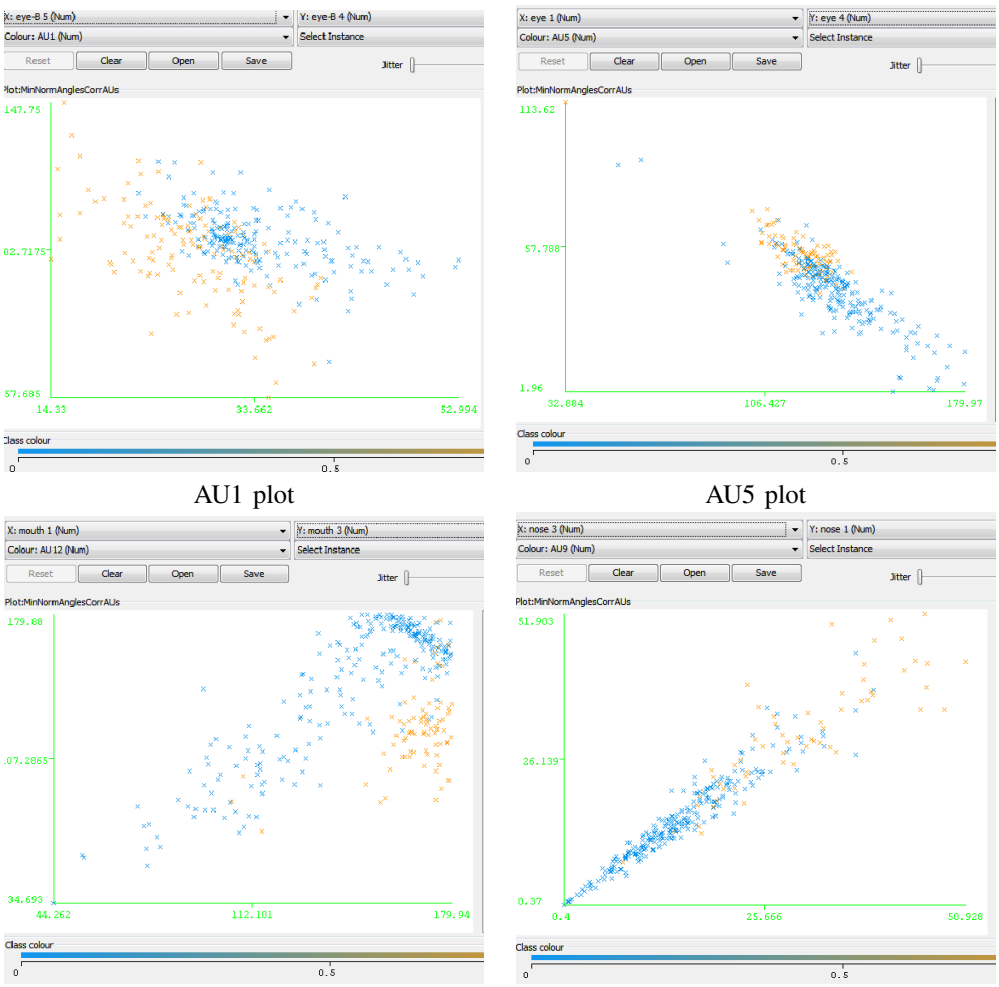

Fig. 3: Examples of plot showing separable clusters of AUs with only two dimensional angular features. We used WEKA software [19] for plotting the data.

in each bin. So we compute the 10^{th}, 20^{th}, …, 90^{th} percentiles to discretize the data into 10 different states.

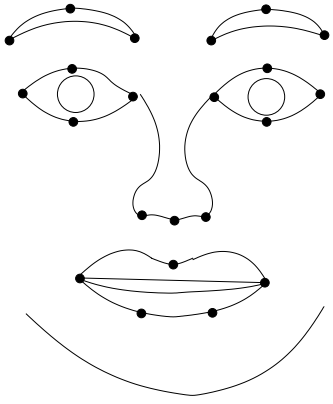

Fig. 4: Significant facial landmark points for features extraction

Appearance Feature extraction To capture the details of facial expressions, we need to capture both the geometric and appearance features. The texture information gives us clues about the skin deformation or wrinkle patterns associated with various expressions. We gathered texture data from certain important regions of face. We choose 12 different facial regions those are sufficient to describe appearance features. Figure 6 shows a pictorial description about the 12 facial regions selected to extract appearance feature.

For each region, a set of four gabor filters at different orientation are applied. To extract features from each of the resultant image after gabor transformation we get 4 different moments those in turn gives summary statistics of each image. We calculate moments $m00$, $m01$, $m10$ and $m11$ for each gabor transformed image. The moments are defined as follows:

$$m(i,j) = \sum I(x,y) \times x_i \times y_i \tag{4}$$

where $I(x,y)$ is the intensity of the pixel at x,y location of

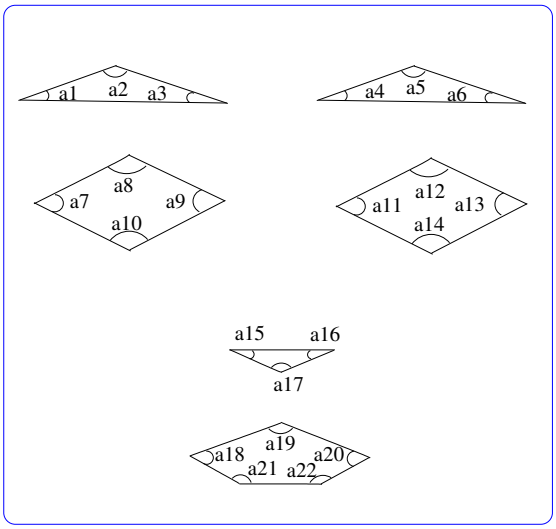

Fig. 5: Angular features extraction from 22 facial points.

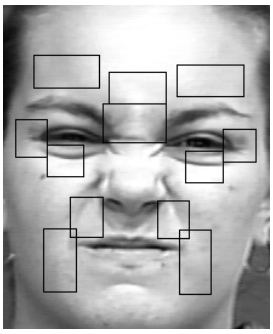

Fig. 6: An example of an image showing 12 selected regions to extract appearance feature.

the image. x and y are normalized image coordinates and summation is over the entire image.

In the network structure, the emotion layer is slightly modified from the work of Qiang Ji [5] to improve the recognition results. The Bayesian structure is first used for Action unit interaction. The the structure is learned by constraining two parents for each AU. Where AU12 as shown in Fig. 2 is the root node for getting the direction of arrows.

Complete Bayesian Structure consists of following three main layers: a.

1) **Emotion layer:** It consists of a single node that has six different emotion states, namely Anger, disgust, fear, happiness, sadness and surprise. This layer differ from the work of Quang Ji. proposed in [5] that it consists of only single node with six different states rather that six different nodes for each emotion with two states being present and absent. This is because the earlier structure allows for simultaneous presence of multiple emotions. But the new structure do not consider this possibility and therefore gives better recognition results.
2) **Action Unit Layer:** We have considered only the major Action units (AU $1, 2, 4, 5, 6, 7, 9, 12, 15, 17, 23, 24, 25, 26$ and 27 as shown in Fig. 8) as they are sufficient to represent the basic range of emotions that we are trying to identify. The AU network that we shown in Fig. 7 is represented in this layer. To incorporate the texture information for decision making we added new set of nodes to each AU, corresponding the prediction made about that AU from the texture information only. This is combined with the geometric information we get from the Anatomical layer to make the final decision about the type of the AU.
3) **Anatomical Layer:**
If we directly connect all the AUs to one facial component this would result in a network that allows huge number of possible AU components; whereas in reality only few of them could occur together. We could represent the most frequent combination of AUs states through the anatomical layer. So, new nodes have to be created for mouth, eyes, eyebrows and nose to take care of correlation among the AUs for these facial parts. The earlier approach proposed by Qiang Ji [5] has manually defined the CPDs of the nodes in anatomical layer. In this paper we use a new approach based on K-means clustering to define various states in anatomical nodes. We arbitrarily assigned 0 for absent state and 1 to represent present state of each AU. The total number cluster (n) is decided based on the decrease in within-cluster sum of square error as we increase the number of clusters. The n number of centers obtained are used to determine the state of anatomical nodes from the available states of AUs.
4) **Landmark Measurement Layer:** The landmark measurement layer consists of angular features. The angular features are discrete in nature and are distributed into ten states. By providing these measurements as evidence and by using appearance feature vector that extracts 4 different moments for each facial regions, the networks predicts the probability of occurrence of each AU.

IV. Experimental results and discussions

The network parameters were trained using clustering algorithm for learning Bayesian network. The network is initiated with feature vector of dimension 22 for each image. Fig. 9 depicts the prior probability distributions learned by the network. We computed the accuracy of classification by validation using 5 fold cross validation. For training of Bayesian network we use free software developed by Darwiche [18] and GeNIe (Graphical Network Interface) software package developed by Decision Systems Laboratory, University of Pittsburgh. The recognition results of Bayesian network is shown in the Table I below. It has been observed that an average recognition accuracy of 95.7% is obtained with highest recognition rate 100% for dataset containing sad and lowest 84% for fear data.

To validate the recognition accuracy of the proposed Bayesian network based approach, we further classify the 22

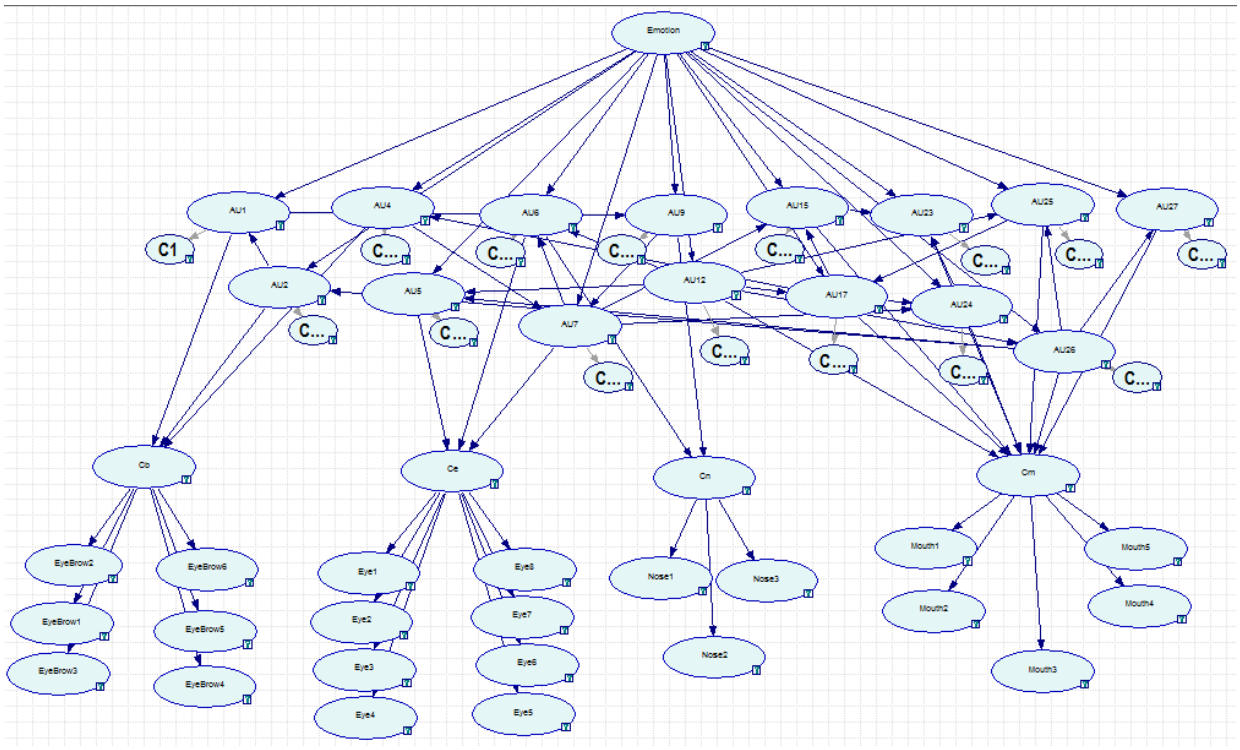

Fig. 7: Bayesian network with three layers: Bottom most layer contains measurement data; anatomical layers contains 15 AUs and the top most layer contains emotion node.

dimensional angular feature vector using 3 more well known classifiers: Support Vector Machine (SVM), decision tree and Radial basis function network (RBFN). We used WEKA software [19] for SVM, RBFN and Decision tree. Tables II, III and IV shows the confusion matrices for SVM, decision tree and RBFN. We use polynomial kernel with degree 2 and cost parameter as 1 for SVM. For decision tree we use *J*48 algorithm, confidence factor as 0.25 and minimum number of instances per leaf as 2. The average recognition accuracies of SVM, decision tree and RBFN are 70.6%, 94.5% and 88.4% respectively. Overall we observed that the recognition performance is best for Bayesian network based learning. It outperforms all the 3 other classifiers.

V. Conlusions

In this paper we present a hierarchical method for modeling emotions from facial expressions images using Bayesian Network. The facial activities are being characterized using three different layers. Bottom layer uses angular features as measurement data. Middle layer contains 15 different AUs and their inter-relationships. Finally, the top layer shows emotion node. We propose a method to automatically draw the relationships among AUs and measurement data using k-means clustering approach. The method thus allows to model the interdependencies among different AUs rather than identifying each AU independently. We also propose a features extraction method using angular information extracted from 22 facial landmarks. This approach reduces the feature dimensions from 136 dimensional data to only 22 dimensional data. The experimentation is performed on extended Cohn-Kanade database and we achieved high recognition accuracy. An average emotion recognition accuracy of 95.7% is achieved in the proposed Bayesian network based approach. We apply three other widely used classifiers such as, Support vector

TABLE I: Confusion matrix of emotions detection for the angular features data using Bayesian network. The emotion classified with maximum percentage is shown to be the detected emotion.

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**97.8**	0	0	2.2	0	0
Disgust	0	**94.91**	0	3.38	1.69	0
Fear	0	0	**84**	0	0	16
Happy	0	1.15	0	**94.26**	4.6	0
Sad	0	0	0	0	**100**	0
	0	0	0	1.2	0	**98.8**

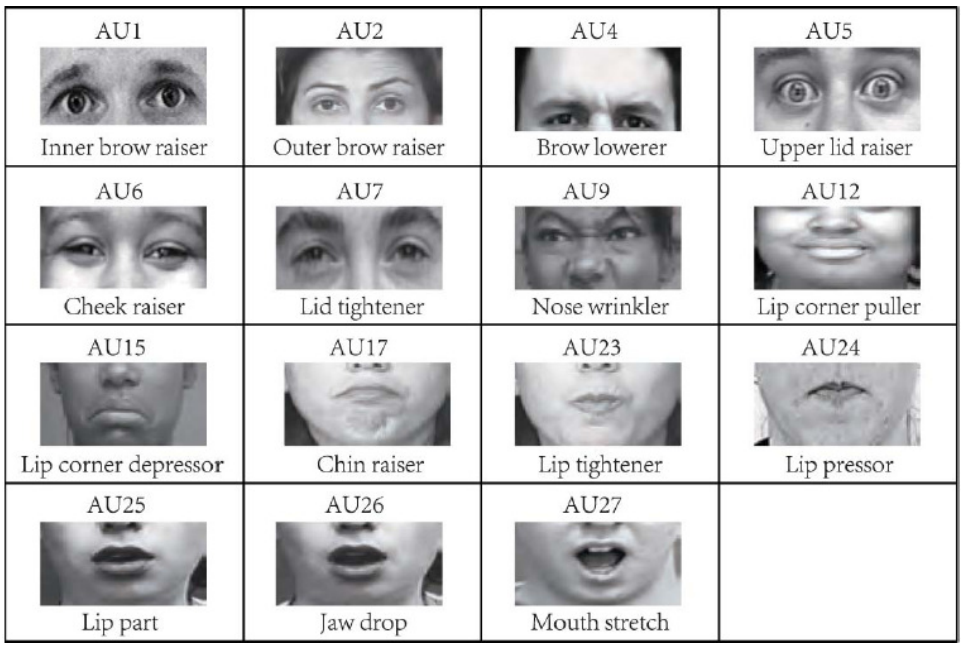

Fig. 8: List of AUs and their interpretations (Taken from [7])

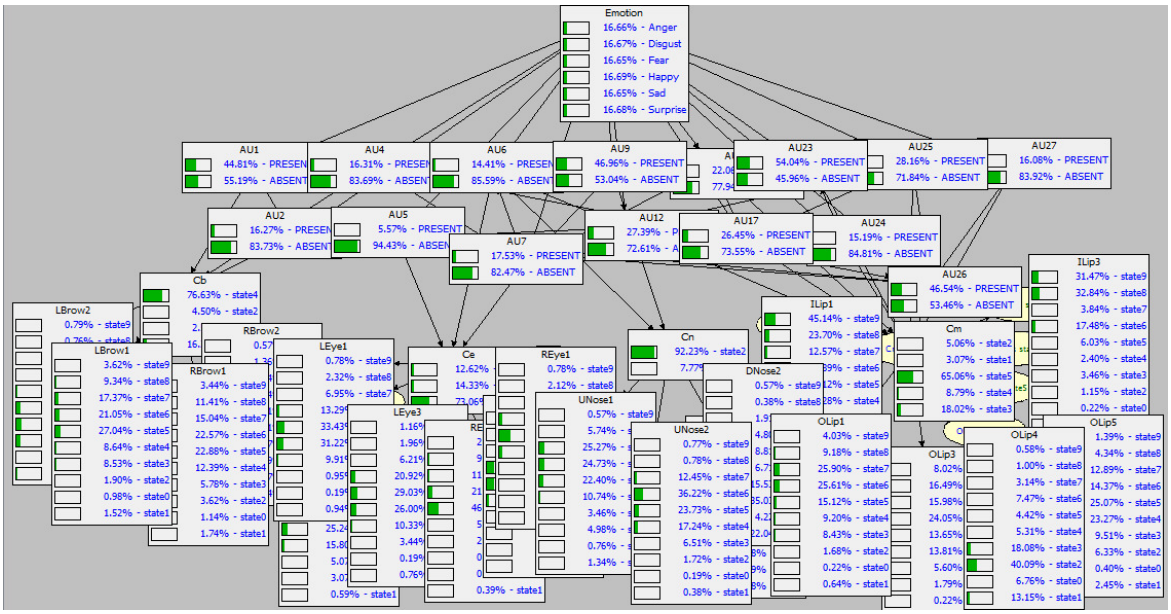

Fig. 9: Distribution of prior probabilities learned by the network.

machine, Decision tree and Radial basis functions network to have a comparative study. The Bayesian network based classification approach is found to be outperforming all other applied approaches. The improved performance of Bayesian network could be attributed to following two factors. Firstly, in the Bayesian approach we look at subsets of data (data with smaller dimensions) rather than the using whole high dimensional data set which is generally used by other classifiers. As an example, we use only six dimensional data to estimate the state of eyebrows which is combined with states of eyes (estimated using only 8 dimensional data); whereas other classifiers use the entire 22 dimensional data. Thus makes the classification job inherently more difficult. Secondly, Instead of independently predicting the AUs, it utilizes the classification results of successively classified AUs to determine remaining AUs states.

As a future work, we can extend the work to study the sensitivity of the network towards partial occlusions like, person wearing glasses, with beard etc.

TABLE II: Confusion matrix of emotions detection for angular features data using SVM. The emotion classified with maximum percentage is shown to be the detected emotion.

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**71.11**	6.67	2.22	6.67	1.33	0
Disgust	6.78	**67.80**	1.69	10.17	5.08	8.47
Fear	4	4	**83.13**	0.12	0.2	0.32
Happy	4.60	4.49	4.49	**79.31**	3.37	3.37
Sad	21.43	7.14	17.86	7.14	**46.42**	0
Surprise	1.20	1.20	8.43	4.82	1.20	**83.13**

TABLE III: Confusion matrix of emotions detection for the angular features data using Decision tree. The emotion classified with maximum percentage is shown to be the detected emotion.

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**97.78**	0	0	0	2.22	0
Disgust	5.08	**94.91**	0	0	0	0
Fear	10.71	0	**64.28**	0	7.14	7.14
Happy	0	0	0	**98.85**	1.15	0
Sad	7.14	0	3.57	0	**89.28**	0
Surprise	0	0	1.20	2.41	0	**96.3**

TABLE IV: Confusion matrix of emotions detection for the angular features data using RBFN. The emotion classified with maximum percentage is shown to be the detected emotion.

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**86.67**	4.44	0	2.22	6.67	0
Disgust	0	**94.92**	1.69	0	1.69	1.69
Fear	0.04	0.04	**24**	0.08	0.16	45.83
Happy	1.15	1.15	0	**96.55**	1.49	0
Sad	1.07	3.57	0	0	**85.71**	0
Surprise	0	0	1.20	2.41	0	**96.38**

References

[1] A. Mehrabian, *Nonverbal communication*. Aldine, 2007.

[2] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 21, no. 10, pp. 974–989, 1999.

[3] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[4] A. Majumder, L. Behera, and V. K. Subramanian, "Facial expression recognition with regional features using local binary patterns," in *CAIP (1)*, 2013, pp. 556–563.

[5] Y. Li, S. Wang, Y. Zhao, and Q. Ji, "Simultaneous facial feature tracking and facial expression recognition," 2013.

[6] A. Majumder, L. Behera, and V. K. Subramanian, "Emotion recognition from geometric facial features using self-organizing map," *Pattern Recognition*, vol. 47, no. 3, pp. 1282 – 1293, 2014.

[7] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 27, no. 5, pp. 699 –714, may 2005.

[8] Y. Tong, W. Liao, and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1683–1699, 2007.

[9] Y. Tong, J. Chen, and Q. Ji, "A unified probabilistic framework for spontaneous facial action modeling and understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 258–273, 2010.

[10] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial action coding system*. A Human Face, Salt Lake City, 2002.

[11] M. Valstar, I. Patras, and M. Pantic, "Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops.*, june 2005, p. 76.

[12] A. S. M. Sohail and P. Bhattacharya, "Classifying facial expressions using level set method based lip contour detection and multi-class support vector machines," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 06, pp. 835–862, 2011.

[13] N. Sebe, M. Lew, Y. Sun, I. Cohen, T. Gevers, and T. Huang, "Authentic facial expression analysis," *Image and Vision Computing*, vol. 25, no. 12, pp. 1856 – 1863, 2007.

[14] H. Kobayashi and F. Hara, "Recognition of six basic facial expression and their strength by neural network," in *IEEE International Workshop on Robot and Human Communication, 1992. Proceedings.* IEEE, 1992, pp. 381–386.

[15] M. N. Dailey, G. W. Cottrell, C. Padgett, and R. Adolphs, "Empath: A neural network that categorizes facial expressions," *Journal of cognitive neuroscience*, vol. 14, no. 8, pp. 1158–1173, 2002.

[16] D. Lin, "Facial expression classification using pca and hierarchical radial basis function network," *Journal of information science and engineering*, vol. 22, no. 5, pp. 1033–1046, 2006.

[17] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[18] A. Darwiche, *Modeling and reasoning with Bayesian networks*. Cambridge University Press, 2009.

[19] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.