Pose and Expression Recognition Using Limited Feature Points Based on a Dynamic Bayesian Network

Wei Zhao[1], Goo-Rak Kwon[2], and Sang-Woong Lee[1]

[1] Department of Computer Engineering
[2] Department of Information and Communication Engineering
Chosun University
Gwangju, Korea 501-759
brandyzhao@gmail.com
grkwon@chosun.ac.kr
swlee@chosun.ac.kr

Abstract. In daily life, language is an important tool during the communications between people. Except the language, facial actions can also provide a lot of information. Therefore, facial actions recognition becomes a popular research topic in Human-Computer Interaction (HCI) field. However, it is always a challenging task because of its complexity. In a literal sense, there are thousands of facial muscular movements many of which have very subtle differences. Moreover, muscular movements always occur spontaneously when the pose is changed.

To address this problem, firstly we build a fully automatic facial points detection system based on local Gabor filter bank and Principal Component Analysis (PCA). Then the Dynamic Bayesian networks (DBNs) are proposed to perform facial actions recognition using junction tree algorithm over a limited number of feature points. In order to evaluate the proposed method, we have applied the Korean face database for model training, and CUbiC FacePix, FEED, and our own database for testing. Experiment results clearly demonstrate the feasibility of the proposed approach.

Keywords: DBNs, Pose and expression recognition, limited feature points, automaticly feature detection, Local Gabor filters, PCA

1 Introduction

Facial actions can provide information not only about affective state, but also about cognitive activity, psychopathology and so on. However, this is always a tough task because of the essence of facial actions. Thousands of distinct nonrigid facial muscular movements have been observed and most of them only differ in a few features. For example, spontaneous facial actions are usually in the term of slight appearance changes. What is more, different facial actions can happen simultaneously. All of these make the recognition difficult. Many methods are proposed by researchers in order to solve this problem. The Facial Action Coding System (FACS) [1] is one of the most popular methods to analyze the facial actions. In FACS system, nonrigid facial nuscular movement is described by a set of facial action units (AUs).

In this paper, we focus on the methods based on DBNs. Many researchers have attempted to build different DBNs to solve this problem. In [2], a unified probabilistic framework is built to

recognize the spontaneous facial actions based on DBNs. The authors assume there are coherent interactions among rigid and nonrigid facial motions. According to this idea, facial feature points can be organized into two categories: global feature points and local feature points. By separating these 28 feature points into two groups, they realize the interactions between pose and expression variables. In this paper, the pose is considered in only pan angle and divided into three state: left, frontal and right. The expression is analyzed by FACS. In another paper [3], a probabilistic measure of similarity is used instead of standard Euclidean nearest-neighbor eigenface matching. The advantage of this improved method is demonstrated by the experiments. In paper [4], the authors use BN for face identification. Some other researchers use hierarchical DBNs for human interactions [5]. In our case, we use DBNs to handle the pose and expression recognition using only 21 feature points on human face. We assume that pose and expression can only be considered as a kind of distribution of the feature points.

The paper is organized as follows: Section 2 illustrates a novel facial features detection method. Section 3 briefly introduces the theories of BN and DBNs. In section 4, our model will be introduced in detail and experiment results will be given too. Section 5 gives the conclusions.

2 Facial Feature Point Detection

Generally, a whole facial actions system includes the detection system and recognition system. The facial feature detection system is very crucial. It decides the performance of the recognition system and the whole system. Here we proposed a low-dimensional facial feature detection system.

2.1 Facial Feature Points Extraction Based on Local Gabor Filter Bank and PCA

For feature points extraction, there are many popular methods such as Gabor filter-based method, Active Shape Model(ASM), Active Appearance Models (AAM) and so on. ASM performs well in experiments [6]. However, ASM is a statistical approach for shape modeling and feature extraction. In order to train the ASM model, a lot of training data is necessary. AAM [7] has the same problem as ASM. In order to realize the fully automatic feature point detection, several other methods appear. One of the most popular methods is to detect feature points using Gabor filter [8]. Gabor filter is a powerful tool in computer vision field. Gabor filters with different frequencies and orientations can serve as excellent band-pass filters and are similar to human visual system. The Gabor filter function can be written as in equation (1):

$$g(x, y, f, \theta) = \frac{1}{ab} exp[-\pi(\frac{x_r^2}{a^2} + \frac{y_r^2}{b^2})][exp(i2\pi f x_r)] \tag{1}$$

where

$$x_r = x\cos\theta + y\sin\theta$$
$$y_r = -x\sin\theta + y\cos\theta$$

In paper [9], the concept of local Gabor filter bank has been proposed and compared with traditional global Gabor filter bank. Both the theoretical analysis and the experiment results show

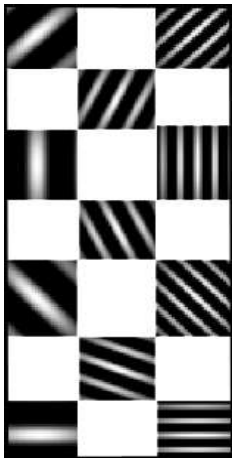

Fig. 1. Local Gabor filter bank with 4 orientations and 3 frequencies

that the local Gabor filter bank is effective. In our experiments, we choose 4 orientations and 3 frequencies from the original 8 orientations and 6 frequencies according to the method introduced in [9]. The 12 Gabor filters are combined together to form a local Gabor filter bank as shown in Fig. 1.

Pupils are the significant features on human faces. Firstly, we apply the method introduced in [10] to detect the pupils and nostrils. Then we take the pupils and nostrils as the reference points to separate the face into several areas. In each small area, we build up the feature vector for each point. The feature vector is extracted from a 11×11 image patch which centered on that point. The image patch is extracted from 12 Gabor filters and the original gray scale image. Thus 1573 ($11 \times 11 \times 13 = 1573$) dimensional vector is used to present one point. In order to describe the method better, we give an example of the feature vector. As shown in Fig. 2 and Fig. 3, there are 12 images each of which is a 11×11 image patch. We can reshape these patches into a vector. This vector represents the point 6 as marked in Fig. 4. In Fig. 2, features of point 6 extracted from two different persons are given. The Fig. 3 shows the features of point 6 and another point. The Gabor filter used in this example is the global Gabor filter. We can also figure out the necessity of using the local Gaobr filter because some patches are similar with each other in these images. The proposed method is similar with the feature extraction method introduced in [8]. However, because we use the local Gabor filter here, the dimension of the feature vector (1573) is much lower than the dimension in [8] which is 8281.

In order to reduce the dimension further, Principal Component Analysis (PCA) is applied. Considering a set of N images $(x_1, x_2, ..., x_N)$, each image is represented by t-dimensional Gabor feature vector. The PCA [9] [11] can be used to transform the t-dimensional vector into a f-dimensional vector, where normally $f \ll t$. The new feature vector $y_i \in \Re^f$ are defined by

$$y_i = W_{pca}^T x_i \quad (i = 1, 2, ..., N) \tag{2}$$

where W_{pca}^T is the linear transformations matrix, i is the number of sample images.

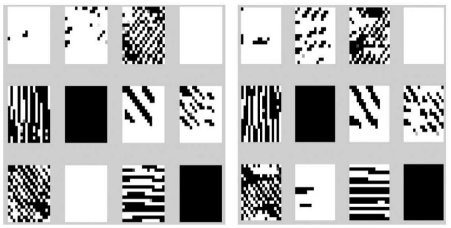

Fig. 2. Features of points 6 from two different persons

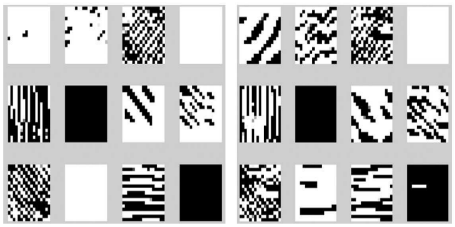

Fig. 3. Features of two points. The left figure is the features of point 6 as marked in Fig.4. The right figure is the features of another point.

In our case, each feature point is represented by 1573-dimensional feature vector. Then we apply PCA to reduce the dimension to a lower dimension. After dimension reduction, calculate the Euclidean distances among tested points and the trained points, we can decide which point is the feature point.

2.2 Experiments and Results of Feature Points Detection

The facial feature points detection method proposed here is trained and tested on the Korea Face database. For our study, we use 60 Korea Face database samples. In Korea Face database, there are several different images for one person which are taken under different illuminations or expressions. We choose the images with natural expressions here. The 60 images are divided into two groups. 20 images are used for training and the 40 images are used for testing. To evaluate the performance of our system, the located facial points were compared to the true points which are got manually. If the distance between automatically detected point and the true point is less than 2 pixels, the detection is defined as a success. Table 1 gives the results of 17 facial feature points' detection result based on Korea Face database. Table 2 gives the results of pupils and nostrils detection results using the method introduced in [10]. From the result table, we can find that the corner points are more easier to be detected while the bottom points like point 8 and point 16 are missed more frequently. This is because the characters of corner points are obvious and stable. The feature points described in Table 1 and Table 2 are shown in Fig. 4. Also Fig. 4 gives an example that the feature points are successfully detected and Fig. 5 shows some general mistakes in experiments.

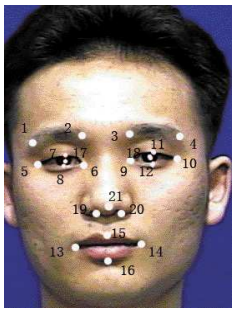

Fig. 4. An example of successfully detected points and the feature points described in Table 1 are marked here too.

Fig. 5. The points 8,15,19 are missed in the left image and the points 14,15,16 are missed in the right image.

Table 1. Facial Feature Point Detection Results Based On Korea Face database

16 **feature points detection results**	
Detected point	Accurate Rate
1: left corner of left eyebrow	95%
2: right corner of left eyebrow	93%
3: left corner of right eyebrow	90%
4: right corner of right eyebrow	90%
5: left corner of the left eye	98%
6: right corner of the left eye	98%
7: top of the left eye	90%
8: bottom of the left eye	86%
9: left corner of the right eye	93%
10: right corner of the right eye	95%
11: top of the right eye	93%
12: bottom of the right eye	90%
13: left corner of mouth	93%
14: right corner of mouth	90%
15: top of the mouth	86%
16: bottom of the mouth	80%
21: center of the nose	90%
Average accurate rate of first 16 points	90.625%

Table 2. Pupils and nostril Detection Results using the method introduced in [10]

Pupils and nostril detection	
Detected point	Accurate Rate
17: left pupil	98%
18: right pupil	98%
19: left nostril	92%
20: right nostril	90%
Average accurate rate of first 16 points	94.5%
Average accurate rate of all points	92.87%

3 The Preliminary of Bayesian Network and Dynamic Bayesian Networks

3.1 An Brief Introduction of BN

A BN was firstly proposed by Pearl[14]. The BN represents the joint probability distribution over a set of random variables in a directed acyclic graph(DAG). The links between these variables represent the causality relationships.

More formally, we can define $Pa(X_i)$ as the parents of variable X_i and the joint probability $P(x)$ over the variables is given by the following equation:

$$P(x) = \prod_{i=1}^{n} P(X_i|Pa(X_i)) \tag{3}$$

where i is the index of variables and n is the total number of variables. From equation (3), we can tell that a BN consists of two crucial respects: the structure and the parameters. So how to learn the structure and the parameters from a actual problem becomes an important issue in BN. We only consider the parameter learning issue here because the structure of BN has already been decided. Parameter learning can be classified into 4 types, depending on the goal is to compute full posterior or just a point, and all the variables are observed or some of them are hidden. In our case, the goal is point estimation and the hidden nodes exist. Hence, Expectation-Maximization(EM) algorithm is applied here.

After parameter learning, the Bayesian network has been fixed. The next work is to infer the result from evidences. Inference is another important task in BN. In this paper, we use Junction Tree Algorithm as our inference engine. Junction Tree algorithm is a very popular algorithm and can perform exact inference in both directed and undirected graphical models.

3.2 An Brief Introduction of DBNs

A DBN can be defined as $B = (G, \Theta)$ where G is the model structure, and Θ represents the model parameters like the CPDs/CPTs for all nodes. There are two assumptions in the DBN model: First, the system is first-order Markovian. Second, the process is stationary which means that the transition probability $P(X^{t+1}|X^t)$ is the same for all t. Therefore, a DBN can be also defined

by two subnetworks: the static network B_0 and $B_{\rightarrow}$ as shown in Fig. 6. The static distribution $B_0 = (G_0, \Theta_0)$ captures the static distribution over all variables X^0. The transition network $B_{\rightarrow} = (G_{\rightarrow}, \Theta_{\rightarrow})$ specifies the transition probability for all t in finite time slices T.

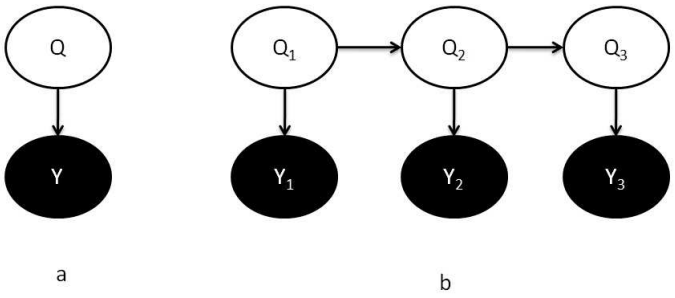

Fig. 6. The left image is a static network B_0 and the right image is the transition network $B \rightarrow$

Given a DBN model, the joint probability over all variables can be factorized by unrolling the DBN into an extended static BN, whose joint probability is computed as follows:

$$P(x^0, ..., x^T) = P_{B_0}(x^0) \prod_{t=0}^{T-1} P_{B_{\rightarrow}}(x^{t+1}|x^t) \tag{4}$$

and transition network $B_{\rightarrow}$ can be decomposed as follows based on the conditional independencies encoded in the DBN:

$$P_{B_{\rightarrow}}(x^{t+1}|x^t) = P_{B_0}(x^0) \prod_{i=1}^{N} P_{B_{\rightarrow}}(x_i^{t+1}|pa(x_i^{t+1})) \tag{5}$$

4 Pose and Expression Recognition Based on the Proposed BN and DBNs

4.1 Introduction of the Proposed BN and DBNs

We build a BN and then extend it to DBNs. We first build a two-layer BN. The first layer contains two discrete variables: pose and smile. Generally, three kinds of angles are used to represent the pose: pan,tilt and roll.

In this paper, we only consider the pan angle. For human being's head, pan angle means turn left or right. Here pan angle is separated into 5 groups according to the angle of the head pose: frontal, left, more-left, right and more-right which are corresponding to five discrete states in BN(Pan angle $\in \{1, 2, 3, 4, 5\}$). As described, we separate left- turn and right-turn angle into two groups. Usually, if the angle is around or larger than $45°$, it will be clustered to the more-left or more-right

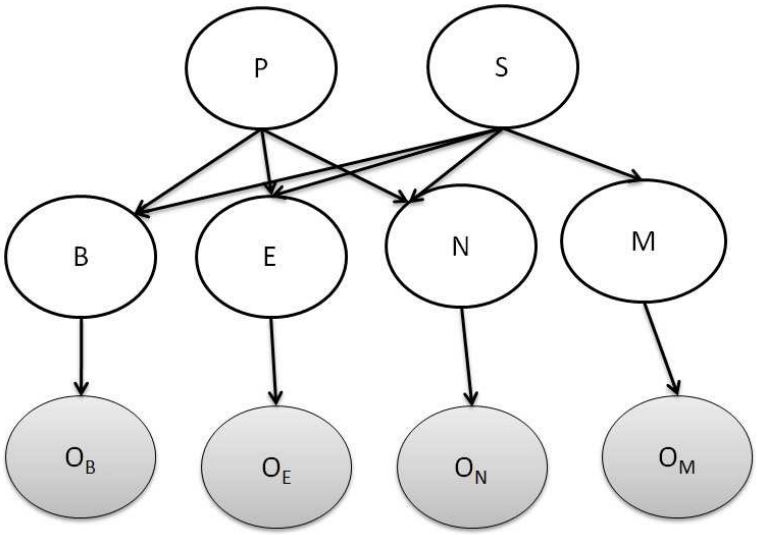

Fig. 7. A simplified Bayesian network of proposed model.

group. If the angle is between 15° and 45°, it will be clustered to left or right group. Of course, the interval here is general.

The second layer consists of four continuous variables: eyebrow, eye, nose and mouth. Each variable is presented by a vector with different length. The joint probability of the first two layers of BN in Fig. 7 is factored into conditional probabilities and prior probabilities as in equation (6). For better expression, we denote P for node pose, S for smile, B for eyebrow, E for eye, N for nose and M for mouth.

$$\begin{aligned} P(B,E,N,M,P,S) &= P(B,E,N,M|P,S) \times P(P,S) \\ &= P(B|P,S)P(E|P,S)P(N|P,S)P(M|P,S)P(P)P(S) \end{aligned} \tag{6}$$

Our aim is to estimate the belief of Pose and Smile nodes given the evidences of the second layer:

$$P(P,S|B,E,N,M) = \frac{P(B,E,N,M,P,S)}{\sum_{pose}\sum_{smile} P(P,S,B,E,N,M)} \tag{7}$$

where the summation is over all possible configurations of the values on the node Pose and Smile.

Then the third layer is built. In this layer, the observation of the second-layer nodes are defined. According to this definition, the nodes in this layer are continuous too. The second and third layers reflet the uncertain relationship between the observations and the real values. The nodes on the third layer can be denoted by O_B, O_E, O_N, O_M. The joint probability of the nodes from these two

layers can be calculated by formula (8):

$$P(B,E,N,M,O_B,O_E,O_N,O_M) = P(B,O_B)P(O_B)P(E,O_E)P(O_E)P(N,O_N)P(O_N)P(M,O_M)P(O_M) \tag{8}$$

According to the definition of DBNs, we unroll the original BN in finite time slices as shown in Figure 8.

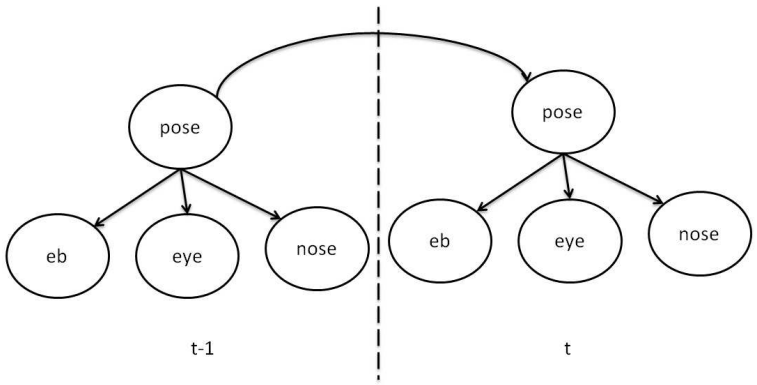

Fig. 8. DBN of pose recognition.

4.2 Parameter Learning of the Proposed Bayesian Network

We only take 21 points on human face as feature points which are shown in Fig 4. When training the BN and DBNs, we choose 50 people from Korea Face database. Each person contains 5 poses and 1 expression. After parameter learning, we can verify the success by sampling some data from the BN and compare them with the training data. From Fig. 9 and Fig. 10, we can figure out that after learning, the similarity of the two figures dramatically increase. This similarity implies the success of parameter learning of BN.

The parameter learning of DBNs can also be verified in the same way. Now we have already get a whole BN and DBNs. Next work is to test this network in experiments.

4.3 Experimental Results of Pose and Expression Recognition Based on BN

Evaluation on CUbiC FacePix(30) Database and Our Own Database CUbiC FacePix(30) is a face image database [12][13]created at the Center for Cognitive Ubiquitous Computing (CUbiC) at Arizona State University. It contains face images of 30 people. There are 3 sets of face images for each of these 30 people (each set consisting of a spectrum of 181 images) where each image corresponds to a rotational interval of 1 degree, across a spectrum of 180 degrees. We use the first

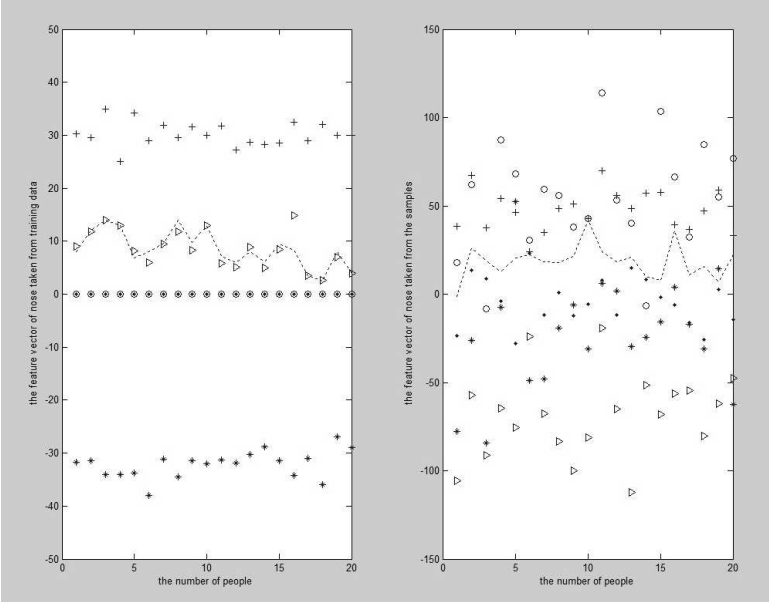

Fig. 9. The left image is from training data of nose. Because the length of nose vector is 6, there are 6 lines in the figure. The right image is the data sampled from model before parameter learning.

set here. In this set, images are taken from the participant's right to left, in one degree increments. Fig. 11 shows some examples from CUbiC FacePix(30) database. We evaluate pose recognition of our system on this database. Firstly, 45 images are picked from database randomly. These 45 images include 6 different angles: right turn $(15°, 30°, 45°)$ and left turn $(15°, 30°, 45°)$. The experiment results are shown in Table 3.

Table 3. Pose recognition with five levels

Pose recognition		
Pose	The group this pose belong to	Accurate Rate
Right(15°)	right	100%
Right(30°)	right	100%
Right(45°)	more-right	100%
Left(15°)	left	93%
Left(30°)	left	100%
Left(45°)	more-left	100%
Images from video	uncertained	95%

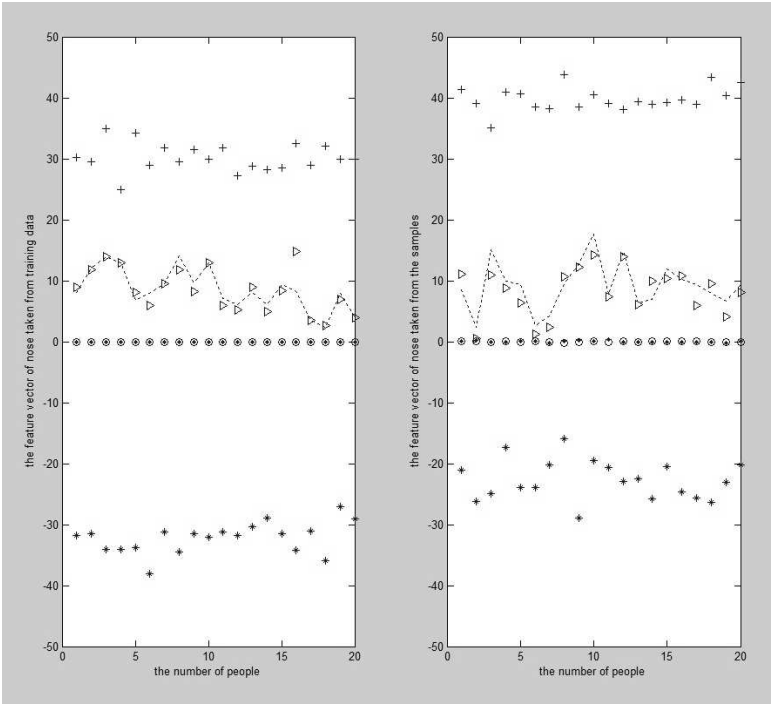

Fig. 10. The left image is training data. The right image is the data sampled from model after parameter learning.

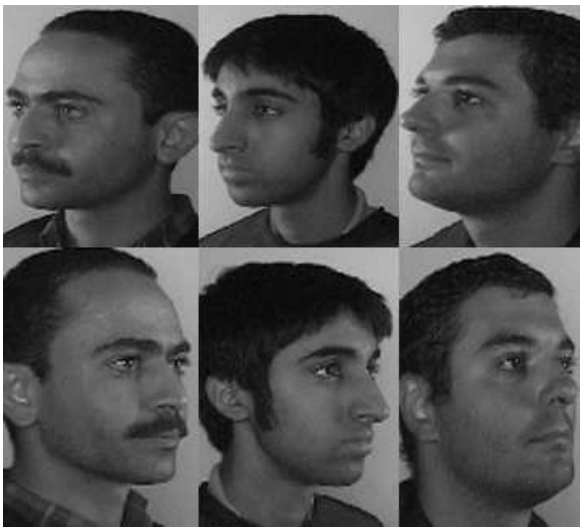

Fig. 11. An example of the test data from CUbiC FacePix(30) database [12] [13] we used in Bayesian Network. The first row is left–turn pose(45 degree). The second row is right–turn pose(45 degree).

After evaluating the network based on CUbiC FacePix(30) database, we generate a small database ourselves. This database is extracted from some short videos in order to evaluate the pose recognition accurate rate in spontaneous situation.

Evaluation on FEED Database In order to evaluate the expression recognition of our system, we use the FEED Database(Facial Expressions and Emotion Database) [15]. This Database with Facial Expressions and Emotions from the Technical University Munich is an image database containing face images showing a number of subjects performing the six different emotions. The database has been developed in order to assist researchers who investigate the effects of different facial expressions. We use the images with happy expressions as shown in Fig 12.

Fig. 12. The images with happy expression from FEED Database [15].

The experiments results based on FEED Database are shown in Table 4.

Table 4. Smile recognition based on the FEED Database

Smile recognition	
Expression	Smile(front)
Accurate Rate	100%

4.4 Experiment Result of Pose and Expression Recognition Based on DBNs

We give the results of pose recognition and smile recognition separately. For each recognition, we define some simple action elements. Firstly, we define four action elements of head pose in two time slices. The definitions are given in Table 5. We choose 50 persons from Korea Face database for training the DBNs. Each person's images contain variety of head poses. We use other 30 persons to test the head pose recognition. Table 5 also gives the results of recognizing these four motions.

For smile recognition, we define four smile's action elements in two time slices whose definitions are given in Table 6. For evaluating the smile recognition, we apply the FEED Database which has introduced previously. We choose 30 persons to train the DBNs and 20 persons to test. The results are shown in Table 6 too.

Table 5. The definition of action elements

$T = t-1$	$T = t$	The name of action elements	Accurate rate
left	left	keep left	90%
left	right	turn right	93%
right	right	keep right	100%
right	left	turn left	90%

Table 6. The definition of action elements

$T = t-1$	$T = t$	The name of action elements	Accurate rate
smile	smile	keep smile	75%
smile	normal	finish smile	90%
normal	smile	smile	95%
normal	normal	keep normal	90%

5 Conclusions and Future Work

In this paper, we propose a whole system for facial actions recognition. There are some conditions for using this method. Firstly, the poses and expressions must have clear and unique feature points' distributions. Secondly, the distribution can reflect the main and universal characters of the poses and expressions we are going to recognize.

Firstly, we build a facial points detection system. The proposed method decreases the dimensions of features dramatically and speed up the system. This method can realize the fully automatic feature detection. However, it is not a realtime method. In practice, it can be used to prepare the training data for other realtime tracking methods like ASM and AAM.

Secondly, we introduce our BN and DBNs for facial actions recognition without 3D information. In BN, The different poses and expressions are presented by different distributions of feature points. Some researchers reconstruct 3D face from 2D images and then project the feature points to a 2D plane. However, sometimes it is difficult to get 3D information. Through these experiments, we find that for inexact pose recognition, 2D information can be applied directly and the results are satisfied.

Extending the BN into DBNs, a set of simple gestures are defined and recognized in experiments. These gestures recognized here are defined between two adjacent time slices. In practice, some gestures can appear instantaneously and disappear suddenly. In order to catch these gestures, the interval between two time slices must be tiny. This requires a very fast feature tracking system. What is more, in order to guarantee the performance, more other features should be added.

Future research directions are realtime face tracking and more other kinds of facial actions recognition. Finally, we want to realize the realtime communication between human and computer.

References

1. Gianluca Donato, Marian Stewart Bartlett, Joseph C. Hager,Paul Ekman, and Terrence J. Sejnowski, *Clasifying Facial Actions* IEEE Transactions on Pattern Aanlysis And Machine Intelligence, Vol. 21, No. 10, October 1999
2. Yan Tong, Jixu Chen and Qiang Ji, *A Unified Probabilistic Framework for Spontaneous Facial Action Modeling and Understanding* IEEE Transactions on Pattern Aanlysis And Machine Intelligence, Vol.32, No.2, February 2010.
3. Baback Moghaddam, Tony Jebara and Alex Pentland*Bayesian face recognition*, Pattern Recognition 33 (2000) 1771-1782
4. Guillaume Heusch, *Bayesian Networks as Generative Models for Face Recognition*, IDIAP RESEARCH INSTITUTE
5. Sangho Park, J.K.Aggarwal*A hierarchical Bayesian network for event recognition of human actions and interactions*, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA
6. Li Dang, Fanrang Kong, "Facial Feature Point Extraction Using A New Improved Active Shape Model", 3rd International Congress on Image and Signal Processing (CISP2010), 2010
7. Ching-Ting Tu and Jenn-Jier James Lien, "Automatic Location of Facial Feature Points and Synthesis of Facial Sketches Using Direct Combined Model", IEEE transactions on systems, man, and cyberneticspart b: cybernetics, Vol. 40, No. 4, August 2010
8. Danijela Vukadinovic, Maja Pantic, "Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers", IEEE International Conference on Systems, Man and Cybernetics Waikoloa, Hawaii October 10-12, 2005.
9. Hong-Bo Deng, Lian-Wen Jin, Li-Xin Zhen, Jian-Cheng Huang, "A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA", International Journal of Information Technology, Vol. 11, No. 11 2005
10. Anima Majumder, L. Behera and Venkatesh K Subramanian, "Automatic and Robust Detection of Facial Features in Frontal Face Images," 2011 UKSim 13th International Conference on Modelling and Simulation
11. R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification. Wiley, New York (2001)
12. Black J, Gargesha M, Kahol K, Kuchi P, Panchanathan S, *A framework for performance evaluation of face recognition algorithms* , ITCOM, Internet Multimedia Systems II, Boston, July 2002.
13. Little G, Krishna S, Black J, Panchanathan S, *A methodology for evaluating robustness of face recognition algorithms with respect to changes in pose and illumination angle*, ICASSP 2005, Philadelphia, March 2005.
14. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.
15. Frank Wallhoff, Facial Expressions and Emotion Database http://www.mmk.ei.tum.de/∼waf/fgnet/feedtum.html, Technische University Mnchen 2006