2008 Seventh International Conference on Machine Learning and Applications

Temporal Exemplar-based Bayesian Networks for Facial Expression Recognition

Lifeng Shang and Kwok-Ping Chan
Department of Computer Science
The University of Hong Kong, Hong Kong
{lfshang, kpchan}@cs.hku.hk

Abstract

We present a Temporal Exemplar-based Bayesian Networks (TEBNs) for facial expression recognition. The proposed Bayesian Networks (BNs) consists of three layers: Observation layer, Exemplars layer and Prior Knowledge layer. In the Exemplars layer, exemplar-based model is integrated with BNs to improve the accuracy of probability estimation. In the Prior Knowledge layer, static BNs is extended to Temporal BNs by considering historical observations to model temporal behavior of facial expression. Experiment on CMU expression database illustrates that the proposed TEBNs is very efficient in modeling the evolution of facial deformation.

1 Introduction

Facial expression recognition has become an active research topic in recent years due to its potential applications in human computer interfaces, image retrieval, data-driven animation, etc. Most facial expression recognition methods attempt to recognize six prototypic expressions (namely, joy, surprise, anger, disgust, sadness and fear) proposed by Ekman [7]. Over the past decade, many techniques (Neural networks [16], Support Vector Machine [1], Local Parameterized Models [4], etc.) have been proposed for still facial images recognition. However, psychological studies suggest that analyzing facial image sequences produces more accurate and robust recognition compared to mug shots [3]. The attention has been moving to model the dynamics for facial deformation by integrating temporal information.

The approaches to model temporal behaviors of facial expressions are generally classified as designing dynamic features (e.g. Dynamic Texture [20]) or introducing sequential data modeling tools (e.g. Dynamic Graphical Model [19]). Yang et al. [17] designed dynamic Haar-like features to represent facial image sequences. Yeasin et al. [18] captured the temporal dynamics of facial sequences by Hidden Markov Models (HMMs). To better model the relative change of the emotional magnitude, Zhang et al. [19] presented a probabilistic framework by integrating the Dynamic Bayesian networks (DBNs) with the facial action units (AUs) [8]. Their methods can reflect the evolution of a spontaneous expression. Meanwhile, their DBNs also involves many latent variables, which makes learning and inference difficult. Recently, Elgammal et al. [9] proposed an exemplar-based approach for view-based recognition of gestures. Their method to utilizing nonparametric exemplar-based density estimation reduced the need for lengthy and non-optimal training of the HMM observation model. This motivated us to well model the evolution of facial expression by exemplar-based model.

This paper presents a three-layer TEBNs to learn the dynamics of facial deformation. In the Exemplars layer, exemplar-based model is integrated with BNs to reduce the need for non-optimal parameters learning. In our exemplar-based model, the similarity function is optimized by Maximum Entropy principle [10], which improves the accuracy of probability estimation without assumptions on prior distributions. To further improve the computational efficiency of the proposed TEBNs, we build it in a local linear subspace constructed by Principal Component Analysis (PCA). In the Prior knowledge layer, we extend static BNs to temporal BNs by considering historical observations.

The rest of this paper is organized as follows. Section 2 presents the features used for the facial expression recognition. In Section 3, we introduce the proposed TEBNs and discuss the computation for the probabilities involved in this model. In Section 4, the performance of proposed method is evaluated by the Cohn-Kanade Database [11]. Section 5 summarizes this paper.

2 Feature Extraction and Indication

Feature extraction is the basis for any recognition system, and the representative features should realistically describe the physical phenomena. In facial expression recognition, there are two types of facial features: permanent and transient features. The permanent facial features are

978-0-7695-3495-4/08 $25.00 © 2008 IEEE
DOI 10.1109/ICMLA.2008.9

Figure 1. The tracking results of one subject's six basic expressions.

the shapes and locations of eyebrows, eye lids, nose, lips and chin. The transient features are the wrinkles and bulges appeared with expressions. In this paper, we use the movement of permanent facial features away from neutral positions to measure facial expression variation.

We applied the well-known Active Appearance Model (AAM) [6] on facial image sequences to track the movement of facial features. Figure 2(a) shows the shape model consisting of 58 facial points which is identical with the one given in AAM-API [14]. Figure 1 displays the facial feature localization results of one subject's six basic expressions. In [5], the (x, y) coordinates of 58 localized facial points forming a 116-dimensional vector are directly used to represent an image. Based on Ekman'facial action code system (FACS) [8], it can be found that the movements of some facial points (e.g. facial points 1 and 13) are not so important to measuring facial deformation. Thus, a subset is selected from these 58 facial points as feature points depicted in Fig. 2(b), in which the solid triangles and rectangles represent that only the X or Y-coordinates are used as feature and the solid circles represent that both X and Y-coordinates are used. The midpoint of the inner corners of the two eyes (facial points 18 and 26) is defined as the origin. To further reduce the inter-personal variations with regard to the amplitudes of facial actions, the movement of feature points are mapped to the range $[-1, 1]$ by

$$F(x;\sigma,T_F) = \begin{cases} 0, & \text{if } |x| \le T_F \\ \text{sign}(x)\left(1 - e^{-(\frac{x}{\sigma})^2}\right), & \text{otherwise} \end{cases}$$

here, $\text{sign}(\cdot)$ is the sign function defined as

$$\text{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0. \end{cases}$$

In this paper, the value of T_F is 1.5, it means that all the translations no larger than 1.5 pixels are regarded as tracking noise and set to 0. The value of σ is determined by $\sigma = \max(|x|)/6$. Figure 3 is the plot of the function $F(x;\sigma,T_F)$. We can see that the value of this function approaches 1 or -1 as the absolute values of movements increase. After this mapping, each frame of video to be recognized is represented by a 52-dimensional feature vector with its elements located in the range $(-1, 1)$.

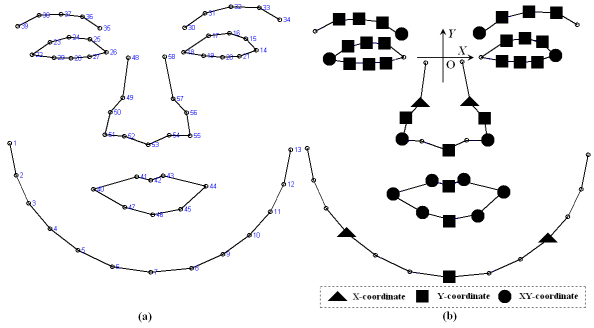

Figure 2. (a)The facial landmarks(58 facial points) and (b) selected feature points.

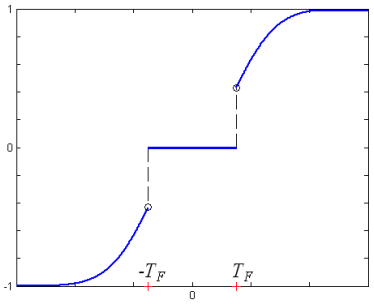

Figure 3. (a)Plot of the function $F(x)$.

3 The proposed TEBNs

In this section, the proposed facial expression classifier-TEBNs and its probabilities estimation are described. Figure 4 shows the proposed TEBNs. It consists of three layers: Prior Knowledge Layer, Exemplars Layer and Observation Layer. In the Observation layer, $X(t)$ is a 52-dimensional feature vector representing the t-th ($t = 1, 2, \cdots, +\infty$) frame of an image sequence. In the Exemplars Layer, E_i ($i \in \{1, 2, \cdots, 6\}$) is the set of training examples of the i-th expression. As in [9], the whole set of training examples is used as the set of exemplars. $L_i(t)$ is the k-nearest neighbors of $X(t)$ in set E_i. In the Prior Knowledge Layer, Y_i is the label of expression i. $H(t)$ represents prior knowledge

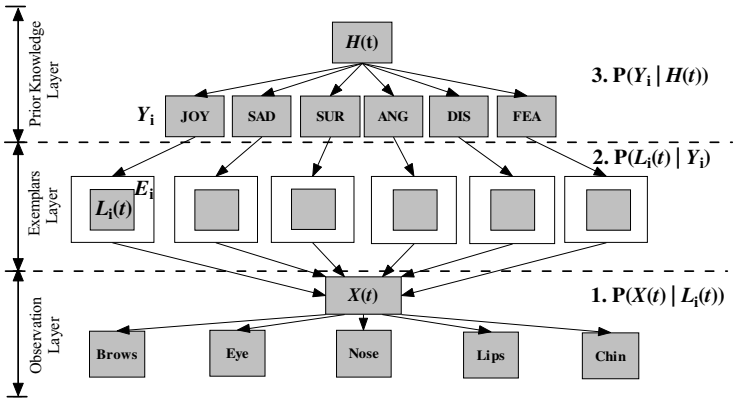

Figure 4. The model of the proposed TEBNs. Note, JOY-Joy, SAD-Sadness, SUR-Surprise, ANG-Anger, DIS-Disgust and FEA-Fear.

of the current time slice. It can be seen that there are three conditional probabilities $P(X(t)|L_i(t))$, $P(L_i(t)|Y_i)$ and $P(Y_i|H(t))$ that should be computed. Following [5], posterior probability is defined as $P(Y_i|X(t), H(t))$. According the Bayes' rule,

$$P(Y_i|X(t), H(t)) = \frac{P(X(t)|Y_i, H(t))P(Y_i|H(t))}{P(X(t)|H(t))}, \quad (1)$$

assuming $X(t)$ is independent on $H(t)$ given Y_i, then

$$P(Y_i|X(t), H(t)) = \frac{P(X(t)|L_i(t))P(L_i(t)|Y_i)P(Y_i|H(t))}{P(X(t)|H(t))}, \quad (2)$$

here $P(X(t)|H(t))$ is independent on class labels, thus it can be viewed as merely a scale factor. In the remainder part of this section, we will give the computation for the other three probabilities involved in (2), respectively.

3.1. The Computation for $P(X(t)|L_i(t))$

The probability $P(X(t)|L_i(t))$ is computed by an exemplar based method. For a classification task, an exemplar model consists of labeled exemplars $X^* = \{x_i^*|1 \leq i \leq N\} \in \mathbb{R}^D$ and similarity function $s(x, x^*)$ measuring how closely a new observation x is related to x^* [9]. The posterior probability of the new observation x classified to class c is

$$P(c|x) = \frac{\sum_{i=1}^{N} s(x, x_i^*)I(y_i^* = c)}{\sum_{i=1}^{N} s(x, x_i^*)} \quad (3)$$

here y_i^* is the label for exemplar x_i^* and $I(S)$ is the indicator function giving 1 if S is true and 0 otherwise. In the computation for $P(X(t)|L_i(t))$, the exemplar set is the union of $X(t)$' k-nearest neighbors, $\bigcup_{i=1}^{6} L_i(t)$. To get an accurate estimation of $P(X(t)|L_i(t))$ without assumptions on prior distribution, similarity function is constructed from the point of maximum entropy as given in [10].

Let $L_{i,j}(t)$ denote the j-th ($1 \leq j \leq k$) element of $L_i(t)$. For simplicity, let $w_{i,j}(t)$ denote the similarity between sample $L_{i,j}(t)$ and the observation $X(t)$, $w_{i,j}(t) \equiv s(X(t), L_{i,j}(t))$. The similarity $w_{i,j}(t)$ is optimized by

$$\text{Maximize} \left(-\sum_{i=1}^{6}\sum_{j=1}^{k} w_{i,j}(t) \ln(w_{i,j}(t)) \right). \quad (4)$$

$$\text{Subject to} \sum_{i=1}^{6}\sum_{j=1}^{k} w_{i,j}(t) = 1. \quad (5)$$

$$\sum_{i=1}^{6}\sum_{j=1}^{k} w_{i,j}(t) L_{i,j}(t) = X(t). \quad (6)$$

This model is to maximize the entropy of $w_{i,j}(t)$ (see (4)) with the constraint that $X(t)$ is the linear combination of its k-nearest neighbors $L_{i,j}(t)$ (see (6)). The analytical solution of this optimization problem is

$$w_{i,j}(t) = \exp(-\lambda_0 + \lambda^T(L_{i,j}(t) - X(t))), \quad (7)$$

where scalar λ_0 and vector $\lambda = [\lambda_1, \lambda_2, \cdots, \lambda_D]^T$ are the Lagrange multipliers corresponding to the $(D+1)$ constraints, D is the dimension of observation $X(t)$. Substituting (7) into equations (5) and (6) yields the following nonlinear equations

$$\begin{cases} \sum_{i=1}^{6}\sum_{j=1}^{k} \exp(-\lambda_0 + \lambda^T(L_{i,j}(t) - X(t))) = 0; \\ \sum_{i=1}^{6}\sum_{j=1}^{k} \exp(-\lambda_0 + \lambda^T(L_{i,j}(t) - X(t)))L_{i,j}(t) = 0. \end{cases} \quad (8)$$

The values of λ_0 and λ are obtained through solving the $(D+1)$ nonlinear equations (8) with the Newton's method (See the Appendix).

In this work, the value of D is 52, which implies that there are (52 + 1) λ's which values have to be optimized, this is computationally expensive.To overcome this problem, the optimization model (4)-(6) is rebuilt in a d-dimensional linear subspace constructed by PCA. Let C denote the transformation matrix. The number of constraints is reduced from $(D+1)$ to $(d+1)$,

$$\begin{cases} \sum_{i=1}^{6}\sum_{j=1}^{k} w_{i,j}(t) = 1; \\ \sum_{i=1}^{6}\sum_{j=1}^{k} w_{i,j}(t) C^T L_{i,j}(t) = C^T X(t). \end{cases} \tag{9}$$

and the analytical solution becomes

$$w_{i,j}(t) = \exp(-\lambda_0 + \lambda^T C^T (L_{i,j}(t) - X(t))). \tag{10}$$

The determination of d will be given in the experimental part.

By (3), the probability $P(L_i(t)|X(t))$ is computed as

$$P(L_i(t)|X(t)) = \frac{\sum_{j=1}^{k} w_{i,j}(t)}{\sum_{i=1}^{6}\sum_{j=1}^{k} w_{i,j}(t)}. \tag{11}$$

Assume $P(L_i(t)) = \frac{1}{6}$ for $i \in \{1, 2, \ldots, 6\}$, the probability $P(X(t)|L_i(t))$ is computed as

$$P(X(t)|L_i(t)) = 6 \times P(L_i(t)|X(t))P(X(t)). \tag{12}$$

3.2. The Computation for $P(L_i(t)|Y_i)$

Exemplar set E_i contains all the knowledge of the expression Y_i, so $P(L_i(t)|Y_i)$ can be replaced by $P(L_i(t)|E_i)$.

$$P(L_i(t)|Y_i) \triangleq P(L_i(t)|E_i) = \frac{\sum_{x \in L_i(t)} P(x)}{\sum_{x \in E_i} P(x)}. \tag{13}$$

The probability $P(x)$ is estimated by the generally used Kernel probability estimation method,

$$P(x) = \frac{1}{|E_i|} \sum_{y \in E_i} \varphi_h(x - y). \tag{14}$$

here h is the bandwidth and its value is determined by the method given in [13]. Probability $P(L_i(t)|Y_i)$ describes the importance of the k-nearest neighbors $L_i(t)$ with respect to E_i. The larger the value of this probability is, the more representative of $L_i(t)$ is.

3.3. The Computation for $P(Y_i|H(t))$

The probability $P(Y_i|H(t))$ represents the prior probability of the i-th expression at time t. The prior knowledge $H(t)$ is defined as the historical observations from the first time slice to the $(t-1)$-th time slice, $X(1:t-1)$. $P(Y_i|H(t))$ is computed as

$$P(Y_i|H(t)) = \sum_{j=1}^{6} P(Y_i|Y_j, X(1:t-1))P(Y_j|X(1:t-1))$$

Consider the Markov property,

$$P(Y_i|X(1:t-1)) = \sum_{j=1}^{6} P(Y_i|Y_j)P(Y_j|X(1:t-1)). \tag{15}$$

and substitute (15)into (2), we have

$$P(Y_i|X(t), H(t)) = \frac{1}{N(t)} P(X(t)|L_i(t))P(L_i(t)|Y_i) \sum_{j=1}^{6} P(Y_i|Y_j)P(Y_j|X(t-1), H(t-1)). \tag{16}$$

here $P(Y_i|Y_j)$ is the transition probability from expression Y_j to Y_i and $N(t)$ is a scale factor to assure $\sum_{i=1}^{6} P(Y_i|X(t), H(t)) = 1$. From (16), we know that if the transition probabilities $P(Y_i|Y_j)$ is given, the computation for probability $P(Y_i|X(t), H(t))$ only involves summation operations.

To facilitate the computation of $P(Y_i|Y_j)$, define a 6×6 matrix T, and let symbol $T(t)$ represent its value at time t. The (i,j)-th entry of $T(t)$ (noted as $T_{i,j}(t)$) records the times that the expression i transmitted to the expression j in two consecutive time slices before time t. T is initialized to matrix with all 1's. At the time t, the transition probability $P(Y_i|Y_j)$ is

$$P(Y_i|Y_j) = \frac{T_{j,i}(t)}{\sum_{k=1}^{6} T_{j,k}(t)}. \tag{17}$$

All the probabilities involved in (16) are obtained and the detailed procedure for computing $P(Y_i|X(t), H(t))$ is given in Algorithm 1. Given a facial expression sequence, the t-th frame is classified to expression

$$i = \arg \max_{i=1,\ldots,6} P(Y_i|X(t), H(t)), \tag{18}$$

if $\max_{i=1,\ldots,6} P(Y_i|X(t), H(t)) > 0.20$, otherwise it is classified to Neutral expression. The expression label for the whole sequence is

$$i = \arg \max_{i=1,\ldots,6} \sum_{t=1,\ldots,L} P(Y_i|X(t), H(t)), \tag{19}$$

here L is the length of the facial image sequence to be recognized.

4 Experiments and Evaluation

4.1. Data Set

We use the Cohn-Kanade Database to evaluate the performance of the proposed TEBNs. This database consists

Algorithm 1 Compute $P(Y_i|X(t), H(t))$

1: **Input**: Labeled exemplars $\bigcup_{i=1}^{6} E_i$, current observation $X(t)$ and prior knowledge $H(t)$.
2: **Output**: The posterior probability $P(Y_i|X(t), H(t))$.
3: **Initialization**: $P(Y_j|X(0), H(0)) = \frac{1}{6}$, $T_{i,j}(0) = 1$.
4: Step 1: Select the k-nearest neighbors $L_i(t)$ of $X(t)$;
5: Step 2: Compute $P(X(t)|L_i(t))$ by (12);
6: Step 3: Compute $P(L_i(t)|Y_i)$ by (13);
7: Step 4: Compute $P(Y_i|Y_j)$ by (17);
8: Step 5: Compute $P(Y_i|X(t), H(t))$ by (16);
9: Step 6: Update T;
10: Step 7: $t \leftarrow t + 1$; Go back to Step 1;

of 100 university students ranging in age from 18 to 30 years. Sixty-five percent were female, fifteen percent were African-American and three percent Asian or Latino. For our experiment, we selected 72 whole image sequences (totally, 1085 images) from the database. Each expression contains 12 sequences. In [2], they only selected the first frame with neutral expression and the last frame containing the apex expression for each sequence. We selected the whole sequences including the expression with weak intensity, which makes the recognition more difficult. The original frames are normalized to 170×210 pixels facial images based on the positions of two eyes.

4.2. Evaluation and Comparison

In this section, we will first use two examples to illustrate the efficiency of the proposed method in an intuitive way. In the first example, we created a short image sequence as shown in Figure 5(a) in which the subject performs smiling with blinking her eyes in frames 5 and 6. We can observe that from the third frame lip corners begin to be pulled obliquely and cheeks are raised. Fig. 5(b) presents the expression likelihood probabilities for six basic expressions. The probabilities of the six expressions are close in the frames 1 and 2, which implies that the two frames have Neutral expression (See (18)). As the expression progresses with time, the probability of joy increases gradually and decreases in the frames 5 and 6 as a result of blinking eyes action. In the frames 7 and 8, the probability of joy rises to nearly 0.9, which implies that these two frames have the apex joy expression. This experimental result illustrates that our method can well model the evolution of facial expression.

Figure 6(a) shows another image sequence in which the subject performs surprise with some frames mis-tracked. In the frames 4 and 6, we can see that the locations of mouth and chin are tracked in error. Fig. 6(b) gives the result of our method, from which we can observe that although the probability of surprise visibly decreases in the frames 4 and

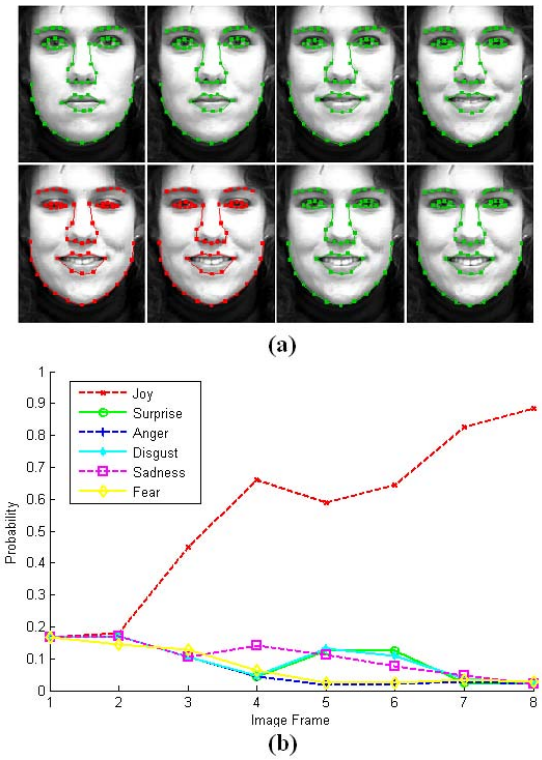

Figure 5. Example 1: (a)An image sequence shows a subject performing smiling with blinking eyes in the frames 5 and 6, (b) the probability distributions of facial expressions.

6 because of tracking error, the facial expression can still be correctly recognized. This example illustrates that our method is robust with respect to tracking error.

Table 1 presents the confusion matrix obtained using a three-fold cross validation. The upper table gives the experimental results for facial expression analysis on still images and the lower table shows the results on videos. It can be seen that the video-based recognition outperforms the still image-based recognition. This result is consistent with the psychological studies that the human ability to recognize facial expression is better in videos than in still images. Table 2 summarizes a comparison to some other approaches. It can be observed that our method outperforms almost all of the other methods for video data recognition and achieves a similar recognition rate (95.83 %) compared with the method [20] (96.26 %) which represents the state-of-the-art recognition technique.

To illustrate the effect of dimension reduction and neighborhood size on average recognition rate, we implemented our experiments with different dimensions and neighbor-

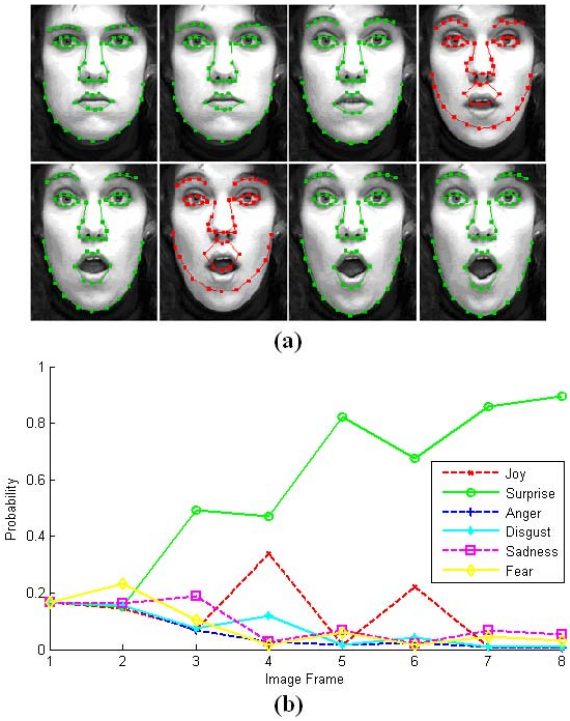

Figure 6. Example 2: (a)An image sequence shows a subject performing surprise with tracking error in the frames 4 and 6, (b) the probability distributions of facial expressions.

Table 1. The confusion matrix for our method

	JOY	SUR	ANG	DIS	SAD	FEA	NEU
JOY	**85.35**	2.56	1.10	3.30	6.23	0.36	1.10
SUR	0.00	**96.19**	0.00	0.00	0.35	2.42	1.04
ANG	0.00	0.00	**87.27**	0.91	4.55	5.91	1.36
DIS	5.56	0.00	3.54	**88.38**	0.00	1.01	1.52
SAD	0.40	0.00	4.02	0.00	**87.15**	7.23	1.20
FEA	1.52	6.06	0.00	2.02	4.55	**84.34**	1.51
NEU	0.00	0.00	0.81	0.00	0.00	0.00	**99.19**

	JOY	SUR	ANG	DIS	SAD	FEA
JOY	**91.66**	0.00	0.00	4.17	4.17	0.00
SUR	0.00	**100.00**	0.00	0.00	0.00	0.00
ANG	0.00	0.00	**95.83**	0.00	4.17	0.00
DIS	0.00	0.00	0.00	**100.00**	0.00	0.00
SAD	0.00	0.00	0.00	0.00	**95.83**	4.17
FEA	0.00	0.00	0.00	4.17	4.17	**91.66**

hood size as shown in Figure 7. It can observed that the best results were obtained with dimension 6 and neighborhood size 15. In our experiments we chose the two values for dimension and neighborhood size settings.

Table 2. Comparison with other approaches

	Class Number	Recognition Rate (%)
[18]	6	90.90
[15]	6	92.10
[12]	7	93.80
Ours	6	95.83
[20]	6	96.26

5 Conclusion

This paper proposed a new BNs for facial expression analysis in levels of still images and videos. We introduced exemplar-based model to BNs, which improved the accuracy of probability estimation without no assumption on the prior distribution. Experiments on CMU expression database confirmed the efficiency of the proposed TEBNs in modeling the evolution of facial deformation.

6 Appendix

Algorithm 2 Compute $\lambda_0, \lambda_1, \cdots, \lambda_D$ by Newton's Method

1: **Input**: Labeled k-nearest neighbors $\bigcup_{i=1}^{6} L_i(t)$ and observation $X(t)$.
2: **Output**: The values $\lambda_0, \lambda_1, \cdots, \lambda_D$.
3: **Initialization**:
$\Phi = \begin{pmatrix} 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 \\ L_{1,1} & \cdots & L_{1,k} & \cdots & L_{6,1} & \cdots & L_{6,k} \end{pmatrix}$,
$\lambda = [0, \cdots, 0]^T$, $x = [1, X(t)]^T$, $\epsilon = 10^{-4}$.
4: Step 1: $w_i \leftarrow \exp(\lambda^T \Phi(:, i))$ and stored by vector $w = [w_1, w_2, \cdots, w_{(6\times k)}]^T$; // $\Phi(:, i)$ represents the i-th column of the matrix Φ
5: Step 2: $v_j \leftarrow x_j - \sum_{l=1}^{6\times k} \Phi(j, l) w_l$ and stored by vector $v = [v_1, v_2, \cdots, v_{(D+1)}]^T$; // $\Phi(j, l)$ represents the (j, l)-th entry of the matrix Φ
6: Step 3: $g_{i,j} \leftarrow \sum_{l=1}^{(6\times k)} \Phi(i, l)\Phi(j, l) w_l$, $g_{i,j}$ is recorded by matrix $G = (g_{i,j})_{(D+1)\times(D+1)}$; //$G$ is the Jacobian matrix
7: Step 4: $\lambda \leftarrow \lambda + G^{-1}v$; // G^{-1} is the inverse matrix of G
8: Step 5: Go back to Step1 until $\|G^{-1}v\| < \epsilon$.

References

[1] M.S. Bartlett, G. Littlewort, I. Fasel, and J.R. Movellan. Real time face detection and expression recognition: Development and application to human-computer interaction. *In CVPR Workshop on CVPR for HCI*, 2003.

[2] M.S. Bartlett, G. Littlewort, J. Movellan, and M.S. Frank. Auto FACS Coding. http://mplab.ucsd.edu/grants/project1/research/fully-auto-facs-coding.html, 2007.

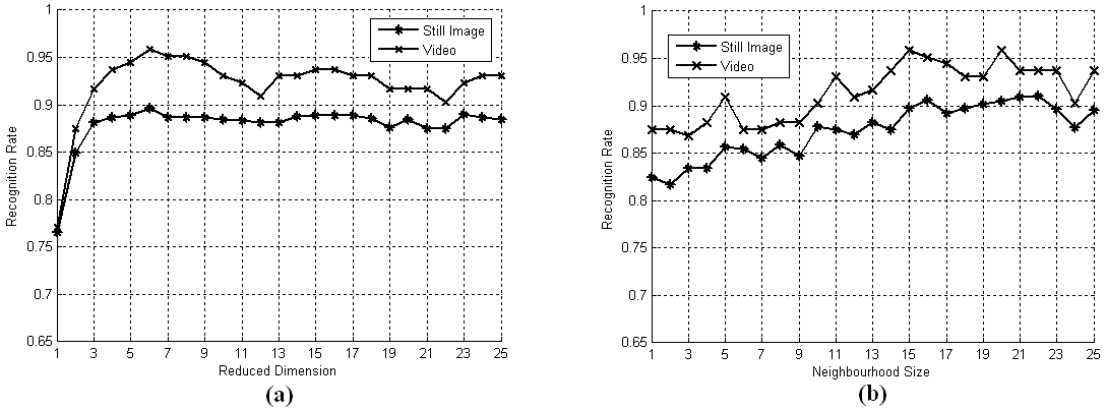

Figure 7. Recognition under different (a) neighborhood size with fixed dimension 6 and (b) reduced dimensionality with fixed neighborhood size 15 for video and still images.

[3] J. Bassili. Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of the Face. *J. Personality and Social Psychology*, 37:2049–2059, 1979.

[4] M. J. Black, and Y. Yacoob. Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion. *International Journal of Computer Vision*, 25(1):23–48, 1997.

[5] Y. Chang, C. Hu, and M. Turk. Probabilistic expression analysis on manifolds. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 520–527, June 2004.

[6] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, June 2001.

[7] P. Ekman, and R. Davidson. *The Nature of Emotion: Fundamental Questions*. Oxford University Press, New York, 1994.

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

[9] A. Elgammal, V. Shet, Y. Yacoob, and L.S. Davis. Learning dynamics for exemplar-based gesture recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 571–578, June 2003.

[10] M. R. Gupta, R. M. Gray, and R. A. Olshen. Nonparametric Supervised Learning by Linear Interpolation with Maximum Entropy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):766–781, May 2006.

[11] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 46–53, 2000.

[12] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of Facial Expression Extracted Automatically from Video. In *Proceedings of the IEEE Workshop Face Processing in Video*,June 2004.

[13] D. Loftsgaarden, and C. Quesenberry. A Nonparametric Estimate of a Multivariate Density Function. *Annals Math. Statistics*, 36:1049–1051, June 1965.

[14] http://www2.imm.dtu.dk/∼aam/.

[15] C. Shan, S. Gong, and P.W. McOwan. Robust Facial Expression Recognition Using Local Binary Patterns. In *Proceedings of the IEEE Intl Conf. Image Procession*, 370–373, 2005.

[16] Y. Tian, T. Kanade, and J. Cohn. Recognizing Action Units for Facial Expression Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115, Feb. 2001.

[17] P. Yang, Q. Liu, and D.N. Metaxas. Boosting Coded Dynamic Features for Facial Action Units and Facial Expression Recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 1–6, June 2007.

[18] M. Yeasin, B. Bullot, and R. Sharma. From facial expression to level of interest: a spatio-temporal approach. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 922–927, June 2004.

[19] 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*, 27(5):699–714, May 2005.

[20] G. Zhao, and M. Pietik*ä*inen. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):915–928, June 2007.