FACIAL EXPRESSION RECOGNITION USING ENCODED DYNAMIC FEATURES

Peng Yang[1] *Qingshan Liu*[1,2] *Dimitris N. Metaxas*[1]

[1]Computer Science Department, Rutgers University
110 Frelinghuysen Road, Piscataway, NJ 08854, USA
[2]National Laboratory of Pattern Recognition, Chinese Academy of Sciences
Beijing, 100080, China
peyang@cs.rutgers.edu, qsliu@cs.rutgers.edu, dnm@cs.rutgers.edu

ABSTRACT

In this paper, we propose a new approach of facial expression recognition. In order to capture the temporal characteristic of facial expressions, we design dynamic haar-like features to represent the facial images, and code them into binary patterns for the further analysis. Based on the encoded features, Adaboost is employed to learn the combination of optimal discriminant features to construct the classifier. The experiments carried on the CMU database show the promising performance of the proposed method.

1. INTRODUCTION

Automatic facial expression recognition has attracted much attention in the recent years due to its potential applications, and lots of methods have been proposed [1] [2]. The early work mostly assumed that facial expressions are static and recognition is done frame-by-frame without using temporal information [3] [4] [5]. Actually, a natural facial event is dynamic, which evolves over time from the onset, the apex and the offset, including facial expressions. Therefore, static image based expression recognition could not achieve a good performance in practical systems. For video based expression recognition [6] [7] [1], how to extract dynamic features to represent facial image is a key issue. Generally, there are two categories of feature representation: geometric features [8] [9] [10] and appearance feature [4] [3]. Appearance features have been demonstrated to be better than geometric features [11] [12], because geometric features are very sensitive to noises, especially illumination noise. Gabor appearance feature has been widely adopted to describe local appearance [13] [4] [14], but its computation expense is much higher than pixel-grey value based appearance features such as LBP.

In this paper, we propose a new approach of facial expression recognition. In order to capture the temporal characteristic of expression, we design dynamic haar-like features to represent the facial images. Compared to Gabor features, haar-like features are very simpler than Gabor representation in computation cost, since they are just based on simple add or minus operators [15]. The dynamic haar-like features are coded into binary patterns for further effective analysis inspired by [16]. Finally based on the encoded dynamic features, the Adaboost is employed to learn the combination of optimal discriminate features to construct the classifier. The experiments carried on the CMU database show the promising performance of our proposed method.

2. DYNAMIC FEATURES REPRESENTATION

In our work, we design haar-like dynamic features to capture the dynamic characteristic of facial expression, due to theirs good performance and simplicity. In order to facilitate further analysis, we code these dynamic features into binary patterns inspired by [16]. Fig. 1 shows the framework of the feature representation, and the details are given in the following subsections.

2.1. Dynamic Haar-like Features

Haar-like feature achieved a great performance in face detection [15]. Considering its much lower computation expense compared with Gabor features, we exploit the haar-like features to represent face images, and extend it to represent the dynamic characteristic of facial expression. For simplicity, we denote one image sequence I with n frames, and each frame with the label I_i, where i is the index of the frame. We set H as the haar-like feature set which includes all the haar-like features in one face image. For each haar feature in frame I_i, We give it label h_{ij}, where i is the index of the frame and j is the index of the haar-like feature in the feature set H. We call $u_{i,j}$ as a dynamic haar-like feature unit $u_{i,j} = \{h_{i-k,j}, h_{i-k+1,j}, \ldots, h_{i+k,j}\}$.

2.2. Coding the Dynamic Features

As mentioned above, a dynamic feature unit is composed of a set of haar-like features in the same position. Generally, it will be taken as a feature vector. However, in this paper, we further code a dynamic feature unit into a binary pattern inspired by [14] [16]. Our coding method has two advantages:

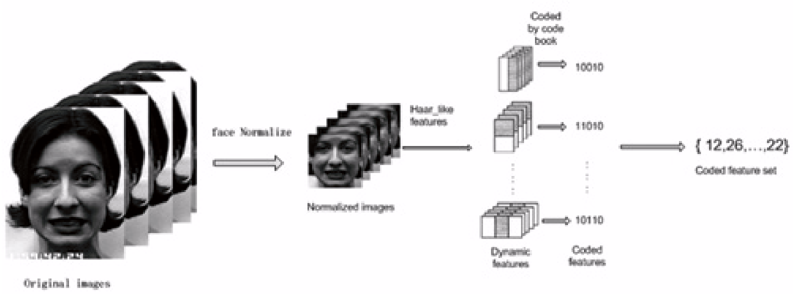

Fig. 1. The frame work of the proposed coded dynamic features representation approach

1) Constructing weak learner for Ababoost learning with one feature is easier than with a feature vector; 2) the proposed binary coding is based on statistical distribution of training samples, so it is robust to noise.

For each expression, we learn its own code book in the training set, and then we code the dynamic features into binary sequence. First, we analyze the distribution of each haar-like feature $h_{,j}$ under each expression, and then the mean μ and the variance σ can be estimated from this distribution. For the simplicity, the gaussian distribution $N_j(\mu, \sigma)$ is adopted to estimate the distribution of the haar-like feature $h_{,j}$. After analyzing the distribution of all haar-like features, we get the code book $B_k\{N_1(\mu_1, \sigma_1), N_2(\mu_2, \sigma_2), ...,$ $N_m(\mu_m, \sigma_m)\}$ for the corresponding expression, where k is the label of the different expressions, j is the index of the haar-like feature and m is the size of the feature set H. Based on the code book, we can map each haar-like feature $h_{i,j}$ to $\{1, 0\}$ by formula 1:

$$C_{i,j} = \begin{cases} 0 & : \quad if \quad \frac{\|h_{i,j} - \mu_j\|}{\sigma_j} > T \\ 1 & : \quad if \quad \frac{\|h_{i,j} - \mu_j\|}{\sigma_j} < T \end{cases} \tag{1}$$

where T is the threshold, because the standard normal gaussian distribution $x \sim N(0, 1)$, $Pr(\|x\| \leq 1.65) = 95\%$, therefore we set $T = 1.65$ in our experiments to cover almost all the positive examples.

Based on the formula 1, we can map one haar-like dynamic feature unit $u_{i,j}$ to one binary pattern $T_{i,j}$.

$$T_{i,j} = \{C_{i-k,j}, C_{i-k+1,j}, ..., C_{i,j}, ..., C_{i+k-1,j}, C_{i+k,j}\} \tag{2}$$

Fig. 2 gives the procedure of creating the coded feature $T_{i,j}$.

3. BOOSTING THE CODED DYNAMIC FEATURES

As in [15], there are thousands of coded dynamic features in a facial expression sequence. In order to find a set of discriminating dynamic coded features and combine them together for recognition, we adopt the Adaboost learning to construct the classifier. Fig 3 gives the algorithm. The expression recognition is a typical multi-classes problem, we decompose the multi-classes into multiple two-classes in order to adapt Adaboost learning. For the expression recognition, there are six expressions: happiness, sadness, surprise, disgust, fear, and anger. For each expression, we set its samples as the positive samples, and the samples of other expressions as the negative samples. Fig 4 illustrates the framework.

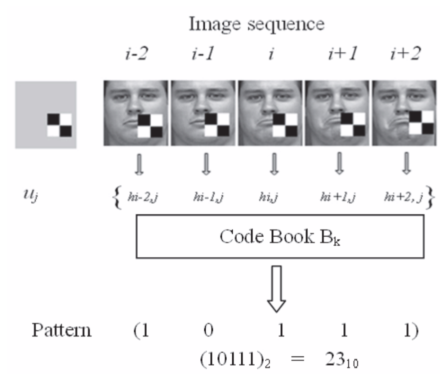

Fig. 2. Example of the coded feature on one haar-like dynamic feature unit $u_{i,j}$

1. Given example image sequences $(x_i, y_i), ..., (x_n, y_n)$, $y_i \in \{1, 0\}$ for specified expression and other expressions respectively.
2. Initialize weight $D_t(i) = 1/N$.
3. Get the dynamic features on each image sequence.
4. Code the dynamic features based on the corresponding code book, and get $T_{i,j}$. Build one weak classifier on each coded feature.
5. Use standard Adaboost to get strong classifier $H(x_i)$.

Fig. 3. Learning procedure based on AdaBoost.

4. EXPERIMENTS

To evaluate the performance of our proposed approach, the experiments are done on the CMU Cohn-Kanade facial ex-

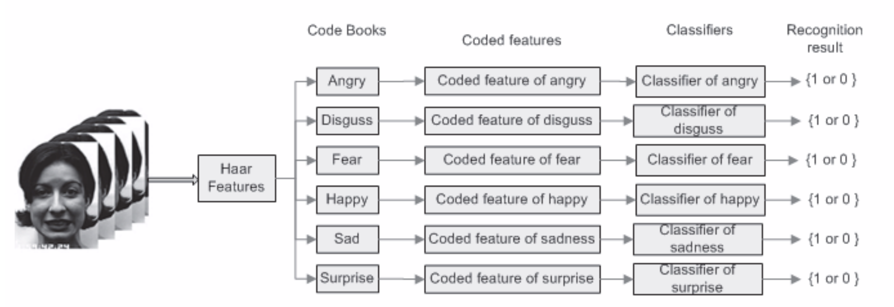

Fig. 4. The framework of the expression recognition

pression database. Psychophysical studies indicates that basic emotions have corresponding universal facial expressions across all cultures, and there is a set of prototypic emotional expressions such as disgust, fear, happiness, surprise, sadness and anger. In our work, we use Viola's [15] method to detect face and normalize the face automatically based on the location of the eyes.

The Cohn-Kanade Facial Expression Database [9] is used here. This database consists of 100 university students aged from 18 to 30 years, of which 65% were female, 15% were African-American, and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were prototypic emotions mentioned above. For our experiments, we selected 300 image sequences from the database. The only selection criterion was that a sequence could be labeled as one of the six basic emotions.The sequences come from 96 subjects, with 1 to 6 emotions per subject. For each sequence, the 8$\sim$12 frames previous peak were used, we randomly select 60 subjects as the training set, and the rest subjects as the testing set. Face image is normalized to 64×64 based on the experiment result of Tian [2], the examples are shown in Fig. 5.

Fig. 5. Examples of six basic expressions.(Anger, Disgust, Fear, Happiness, Sadness and Surprise)

On the training set, we analyzed the distribution of feature units under each expression, thus we get 6 code books for 6 expressions. According to the specified code book, we get the binary sequence for each subject under the corresponding expression. In this experiment, we set the coded feature as

$$T_{i,j} = \{C_{i-2,j}, C_{i,j}, C_{i+2,j}\} \tag{3}$$

and ignore the previous frame $i-1$ and posteriors frame $i+1$ because there is almost no difference between two neighbored frames.

Table 1. The Area under the ROC curves (Expression)

Expression	Static feature + AdaBoost	Coded dynamic feature + AdaBoost
Angry	0.856	0.973
Disgust	0.898	0.941
Fear	0.841	0.916
Happiness	0.951	0.9913
Sadness	0.917	0.978
Surprise	0.974	0.998

In [3] [17], they gave some recognition rate on each expression. In practice, ROC curve is more reasonable and reliable than the recognition rate to be used to evaluate the recognition performance, so we use ROC curve in our experiment to evaluate the proposed method. In order to show the advantage of the coded dynamic feature, we use the Adaboost and static Haar features to do recognition frame by frame on the same training and testing data set. Fig. 6 reports the experiment result, and we can see the proposed coded dynamic feature are much better the static feature. And the area under the ROC curves is listed in the table 1.

5. CONCLUSIONS

This paper presented a novel approach for video_based facial expression recognition by boosting encoded dynamic features based classifiers. Experiments on facial expression database show the coded dynamic features are powerful to discriminate the expressions compared with the static features. The coded dynamic feature could combine the temporal information, describe the temporal pattern and this method can be easily extended to video_based face recognition.

6. REFERENCES

[1] I. Cohen, N. Sebe, L. Chen, A. Garg, and T. Huang, "Facial expression recognition from video sequences Temporal and

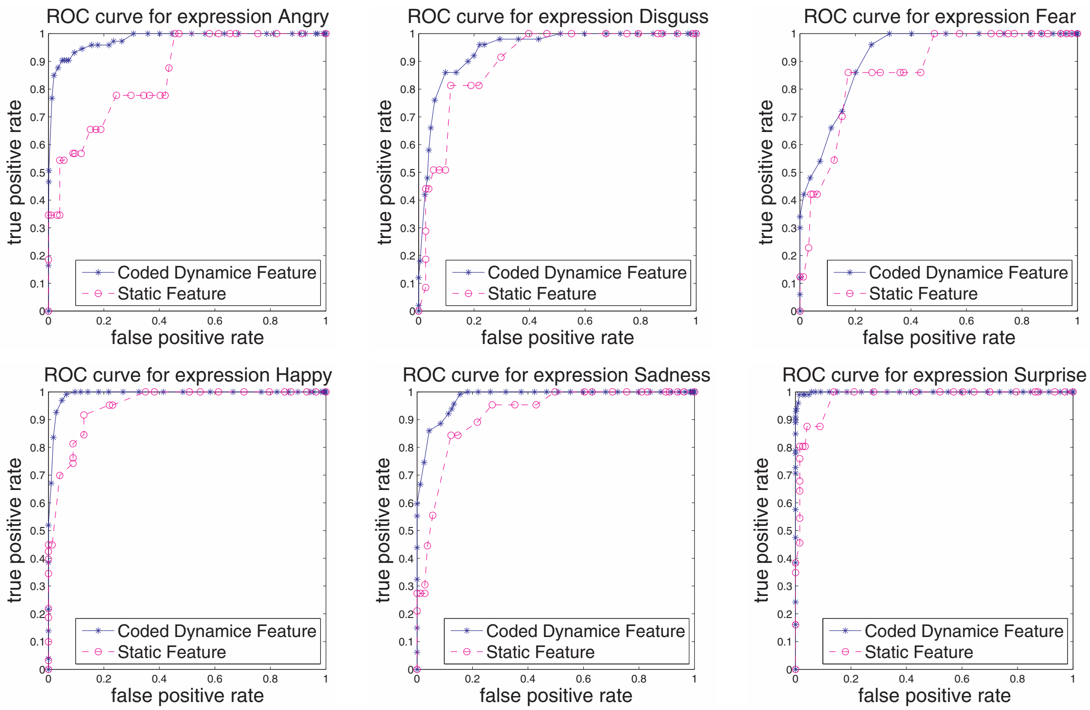

Fig. 6. ROC curves of expressions

static modeling," *Computer Vision and Image Understanding*, 2003.

[2] Y. Tian, "Evaluation of face resolution for expression analysis," *Proc. of Intl Conf. CVPR Workshop on Face Processing in Video (FPIV'04)*, 2004.

[3] Caifeng Shan, Shaogang Gong, and Peter W.McOwan, "Conditional mutual information based boosting for facial expression recognition," *British Machine Vision Conference*, 2005.

[4] M. Bartlett, G. Littlewort, I. Fasel, and J. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction," *Proc. of Intl Conf. CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction*, 2003.

[5] M. Pantic and J. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Transactions on Systems, Man and Cybernetics*, 2004.

[6] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *International Journal of Computer Visione*, 1997.

[7] Yaser Yacoob and Larry Davis, "Computing spatio-temporal representations of human faces," *Proc. of Intl Conf. Computer Vision and Pattern Recognition*, 1994.

[8] Haisong Gu and Qiang Ji, "Facial event classification with task oriented dynamic bayesian network," *Proc. of Intl Conf. Computer Vision and Pattern Recognition*, 2004.

[9] Jenn-Jier James Lien, Takeo Kanade, Jeffrey Cohn, and C. Li, "Detection, tracking, and classification of action units in facial expression," *Journal of Robotics and Autonomous Systems*, 1999.

[10] M. F. Valstar, I. Patras, and M. Pantic, "Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data," *Proc. of Intl Conf. CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction*, 2005.

[11] Zhengyou Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facialexpression recognition using multi-layer perceptrong," *Proc. of Intl Conf. Automatic Face and Gesture Recognition*, 1998.

[12] Y. Tian, T. Kanade, and J. Cohn, "Evaluation of gaborwavelet -based facial action unit recognition in image sequences of increasing complexity," *Proc. of Intl Conf. Automatic Face and Gesture Recognition (FG'02)*, 2002.

[13] Chengjun Liu and Harry Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, 2002.

[14] J. Daugman, "Demodulation by complex-valued wavelets for stochastic pattern recognition," *Int'l J. Wavelets, Multiresolution and Information Processing*, 2003.

[15] Paul Viola and Michael Jones, "Robust real-time object detection," *International Journal of Computer Vision*, 2002.

[16] Timo Ojala and Matti Pietikainen, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002.

[17] Jacob Whitehill and Christian W. Omlin, "Haar features for facs au recognition," *Proc. of Intl Conf. Automatic Face and Gesture Recognition*, 2006.