Facial Expression Recognition Based on Facial Components Detection and HOG Features

Junkai Chen[1], Zenghai Chen[1], Zheru Chi[1], and Hong Fu[1,2]

[1]Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong
[2]Department of Computer Science, Chu Hai College of Higher Education, Hong Kong
Email: Junkai.Chen@connect.polyu.hk

***Abstract*—In this paper, an effective method is proposed to handle the facial expression recognition problem. The system detects the face and facial components including eyes, brows and mouths. Since facial expressions result from facial muscle movements or deformations, and Histogram of Oriented Gradients (HOG) is very sensitive to the object deformations, we apply the HOG to encode these facial components as features. A linear SVM is then trained to perform the facial expression classification. We evaluate our proposed method on the JAFFE dataset and an extended Cohn-Kanade dataset. The average classification rate on the two datasets reaches 94.3% and 88.7%, respectively. Experimental results demonstrate the competitive classification accuracy of our proposed method.**

Keywords—facial expression recognition, HOG features, facial component detection, SVM

I. INTRODUCTION

Human beings could convey intentions and emotions through some nonverbal ways, such as gestures, facial expressions and involuntary language. Facial expressions may be the most useful nonverbal ways for people to communicate with each other. Facial expressions recognition has gained a growing attention because it could be widely used in many fields such as lie detection, medical assessment, and Human Computer Interface (HCI). In fact, a widely accepted prediction is that computing will move to the background, weaving itself into the fabric of our everyday living spaces and projecting the human user into the foreground [1]. To reach this goal, computer vision and machine learning techniques have to be developed while strengthening psychological analysis of emotion.

However, facial expression recognition is an extremely challenging task. Many factors like illumination, pose, deformation and wild environment could contribute to the complexity. Moreover, facial expressions are subtle facial muscle movements, and it is challenge to detect and represent these kinds of slight changes.

Facial expressions have been studies for a long time and we have witnessed some progress in recent decades. The Facial Action Coding System (FACS), which was proposed in 1978 by Ekman et al. [2] and refined in 2002 [3], is a very popular facial expression analysis tool. FACS attempts to decompose facial expressions into different action units. Based on the combination of the action units, facial expressions could be recognized. Another approach is to recognize facial expressions directly from images.

In a direct approach, two mainstream approaches, called appearance-based and geometry-based [4], are used in facial expression recognition. Appearance-based methods apply the Gabor filters, Local Binary Pattern (LBP) texture descriptors to represent the features of facial expressions. Geometry-based methods focus on capturing the shape of faces. A shape is constituted with a group of fiducial points. These points could be regarded as the geometry features.

Many attempts have been made to recognize facial expressions. Zhang et al. [5] investigated two types of features, the geometry-based features and Gabor-wavelets based features, for facial expression recognition. They applied a two-layer perceptron as the classifier and compared the performance of the two features. Feng et al. [6] provided a coarse-to-fine classification scheme for facial expression recognition. The coarse stage included producing the basic model vectors and computing the distance from the feature vectors to the model vectors. After that, a K-nearest neighbor classifier was employed to do the final classification in the fine stage. In [7], Khandait et al. found that the width and height of the face portions were distinct features in facial expression recognition. Based on the facial elements and muscle movements, Zhang et al. adopted the salient distance features to do the facial expression recognition [8]. They extracted the 3-D Gabor features, selected the "salient" patches and matched the patches to obtain salient distance features. Shan et al. [9] considered that the Local Binary Pattern (LBP) was a good texture descriptor and could be used to represent facial expressions. They adopted a Boosted-LBP to select the most distinguished LBP features. The boosted-LBP features were employed to train the SVM and acquired a prominent recognition rate.

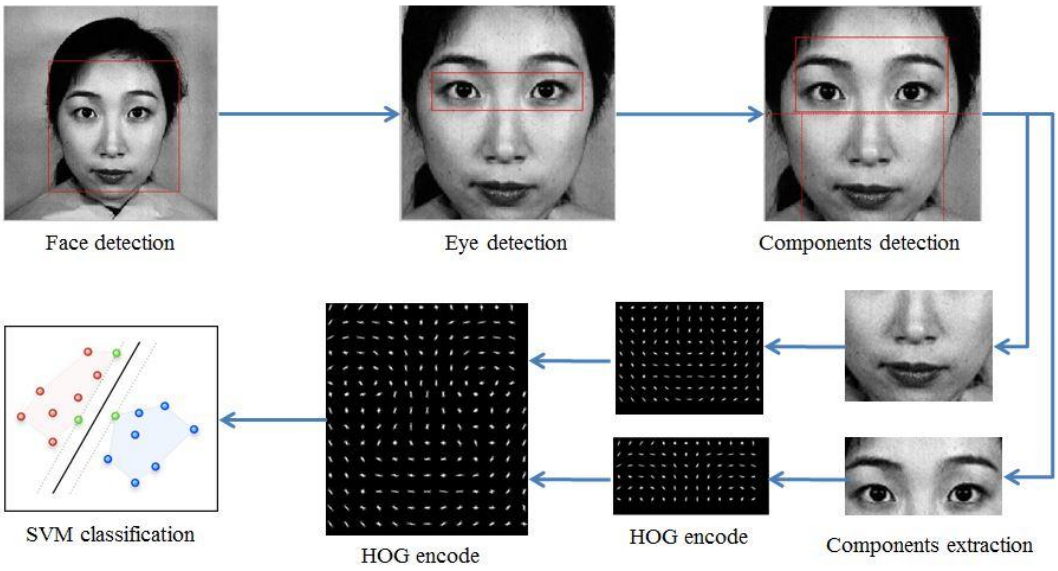

Fig. 1. The schematic overview of the propsoed system.

In this paper, we introduce an effective appearance-based method to handle the facial expression recognition problem. Given a face image, the system detects the face first and then, extracts the facial components from the face image. After that, Histogram Oriented Gradient (HOG) is extracted to encode these facial components and concatenate them into a single feature vector. These feature vectors are used to train a linear SVM.

Our work is somewhat similar with the previous work in [10]. However, there are still some differences between our work and the previous work. The previous work applied the feature descriptors on the whole face, and they explored different features including HOG, LBP and LTP. Our work considered the facial components and employed the HOG feature descriptors on the facial components. The previous work focused on the problem of facial expression recognition with registration errors. Our study paid attention to the facial components which contribute to the facial expression recognition.

The rest of this paper is organized as follows. Section Ⅱ describes our proposed facial expression resonation system, and the details of computing facial components and HOG. Experimental results and analysis are given in section Ⅲ. Concluding remarks are made in Section Ⅳ.

II. PROPOSED FACIAL RECOGNITION SYSTEM

The proposed system includes three function blocks. The first function is face detection and facial components extraction. The second function block is using HOG to encode these components. The last function block is training a SVM classifier. The schematic overview of our proposed facial recognition system is shown in Fig.1.

A. Face Detection and Facial Components Extraction

This part begins with face detection using the Viola-Jones face detector [11]. After the face region is acquired, it is necessary to extract the brows, eyes, nose and mouth from the face. We could detect the eyes first and extract the other components based on the relative positions of these components. The face images of the database we used are all of the frontal view and we know that the brows are above the eyes. We could enlarge the detected eye regions to contain the brows as well. As for the nose and mouth, we know they locate just below the eyes; it is not difficult to locate the region which contains the nose and mouth.

B. Histogram of Oriented Gradients Features

Different features including SIFT [12], Gabor filters [13], Local Binary Patterns (LBP) [14] and HOG (Histogram of Oriented Gradient) [15] have been proposed for facial expression recognition. Facial expressions result from muscle movements and these movements could be regarded as a kind of deformation. For example, the muscle movements of the mouth cause the mouth open or close, and cause brows raiser or lower. These movements are similar to deformations. Considering that HOG features are pretty sensitive to object deformations. In this paper, we propose to use the HOG features to encode facial components. HOG was first proposed by Dalal and Triggs in 2005 [15]. It is well received by computer vision community and widely used in many object detection applications, especially in pedestrian detection. HOG numerates the appearance of gradient orientation in a local patch of an image. The idea

Fig. 2. The seven expressions from one subject.

is that the distribution of the local gradient intensity and orientation could describe the local object appearance and shape [15].

Compared with other features such as LBP and Gabor filters, HOG is also very useful in facial expression recognition. HOG can characterize the shapes of important components constitute facial expressions. So we apply the HOG to encode these facial components. In our experiments, we set cell size to 8×8, the number of bin size to 9, the orientation range to 0 -180.

C. *Support Vector Machine*

Support Vector Machine (SVM) has been widely used in various pattern recognition tasks. It is believed that SVM can achieve a near optimum separation among classes. In our study, we train SVMs to perform facial expression classification using the features we proposed. In general, SVM builds a hyperplane to separate the high-dimensional space. An ideal separation is achieved when the distance between the hyper plane and the training data of any class is the largest. Given a training set of labeled samples:

$$D = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in R^n, y_i \in \{-1,1\}\}_{i=1}^{p} \qquad (1)$$

A SVM tries to find a hyperplane to distinguish the samples with the smallest errors.

$$\mathbf{w} \cdot \mathbf{x} - b = 0 \qquad (2)$$

For a input vector $\mathbf{x_i}$, the classification is achieved by computing the distance from the input vector to the hyperplane. The original SVM is a binary classifier. However, we can take the one-against-rest strategy to perform the multi-class classification. We use the LIBSVM in our experiments [16].

III. Experimental Results and Discussion

In order to evaluate the performance of our proposed approach, we utilize two commonly adopted datasets: The Japanese Female Facial Expression (JAFFE) Database [17] and the Extended Cohn-Kanade Dataset [18].

A. *JAFFE Database*

This database contains 213 images in total. There are 10 subjects and 7 facial expressions for each subject. Each subject has about twenty images and each expression includes two to three images. The seven expressions are angry, happy, disgust, sadness, surprise, fear and neutral respectively. Fig.2 shows the seven expressions from one subject.

In this experiment, images have size of 256×256. After acquiring the face region from the face image, we adjust the size to 156 × 156. And then we apply the techniques mentioned above to detect and extract the facial components and adjust them to the same size. In our experiments, size of the eye-brows is 52×106, the dimensionality of the corresponding HOG encoded feature is 1×2160. Size of the nose-mouth is 78×104, and the dimensionality of the corresponding HOG encoded feature is 1×3456. We concatenate the two feature vectors into a single one. The final feature is a 1 ×5616 vector.

We adopt the leave-one-sample-out strategy to test our method and compare with the other methods. There are 10 subjects in this database. Each subject has a few images. From each group, we randomly select two or three images as the test data set and the remaining images as the training set. At last, there are 23 images in the test set and 190 images constitute the training set. The results are shown in Table Ⅰ.

TABLE I. Classification Results Of Four Methods On The Jaffe Datase.

Method	Classification Rate
Gabor+FSLP [19]	91.0%
LBP [9]	89.1%
Patch-based Gabor [8]	92.3%
Our method	94.3%

In [9], they applied the Local Binary Pattern (LBP) descriptors to represent the facial expression and used the Adaboost to select the optimal features. The average classification rate was about 89.1%. In [19], 18 Gabor filters were convolved with the face images to get the filtered images, and only the amplitudes of selected fiducial points were used as feature vectors. They tested different classifiers and the best performance was about 91%. In [8], Zhang et al. adopted the salient distance features to do the facial expression recognition. They extracted the 3-D Gabor features, selected the "salient" patches, and matched the patches to obtain the salient distance features. The classification rate that they obtained was about 92.3%. From the results, we could find that our method outperforms the other three methods tested.

B. The Extended Cohn-Kanade Dataset

The dataset has 123 subjects and 593 sequences. There are seven expressions and neutral in this dataset. The seven expressions are angry, happy, sad, surprise, contempt, fear, and disgust. Fig.3 shows the 8 expressions with each from a different subject. Among 593 sequences, only 327 sequences have expression labels. We used the peak frame of each labeled sequences as the sample image. The frequency of each expression is shown in Table Ⅱ.

TABLE II. THE FREQUENCY OF EACH EXPRESSION IN THE EXTENDED COHN-KANADE DATASET.

Expression	Frequency
Angry	45
Contempt	18
Disgust	59
Happy	25
Surprise	69
Sad	28
Fear	83

Note that the neural expression is excluded from the experiments. We follow the similar procedure applied in the JAFFE dataset experiments. The original size of the image is 640 ×490. We detect the face first, and adjust the size of face to 256×256. Once we obtained the face region, we could detect the eyes and extract the facial components. The final size of the eye-brow is 74×150 and the nose-mouth 130×128. Down sampling is used for the extracted facial components before applying the HOG to reduce the dimensionality. At last, the HOG encoded features of the eye-brow component are a 1× 864 vector and the HOG encoded features of the nose-mouth component are a 1 × 1764 vector. The final feature is a 1×2628 vector.

In this experiment, we divide the images into two sets. One is the training set and the other is the test set. About one-fifth images of each group are randomly selected for the test set. The remaining images form the training set. At last, there are 59 samples for the test and 268 samples for the training. In order to eliminate the influence of the randomness, we repeated the process 10 times and compute the average classification rate. We achieved an average of 88.7 with a variance of ±2.3% classification rate at last.

In order to compute the classification rate of each expression, we follow the baseline method and adopt the leave-one-subject out strategy. This strategy promises each subject can be evaluated once. There are 118 subjects. Each time, the expression images of one subject are picked out for the test and the images of the other subjects are used for training. We repeat 118 times and compute the average. The results are shown in table Ⅲ. The diagonal values are the hit rates. We could find that the expression "contempt" has the lowest hit rates. This may be this expression is easy to be mixed with the other expressions. The "surprise", "disgust" and "happy" expressions get high hit rates. The three kinds of expressions are more distinct than the other expressions.

We also compare our method with three baseline methods with the results shown in Table Ⅳ. The baseline methods use different features: SPTS, CAPP and SPTS+CAPP, respectively. From Table Ⅳ, we can see that the performance of our method is much better than SPTS and CAPP, especially for the expression "contempt". The hit rate can be improved nearly by 40%. As for the combination of SPTS and CAPP, the hit rate of the expression "contempt" is higher than our method, However, the hit rates of the "anger" and "fear" expressions are lower than our method. Compare with the baseline methods, our method achieves a good performance under a more strict condition. Note that the neutral faces are not used as the reference in our system.

TABLE III. THE CONFUSION MATRIX OF THE EXPRESSIONS.

	AN	**CO**	**DI**	**FE**	**HA**	**SA**	**SU**
AN	0.84	0.04	0.07	0.00	0.02	0.00	0.02
CO	0.06	0.61	0.00	0.11	0.11	0.11	0.00
DI	0.02	0.00	0.95	0.00	0.03	0.00	0.00
FE	0.08	0.04	0.00	0.72	0.12	0.00	0.04
HA	0.01	0.03	0.00	0.00	0.96	0.00	0.00
SA	0.07	0.04	0.00	0.04	0.00	0.82	0.04
SU	0.00	0.01	0.00	0.00	0.00	0.00	0.99

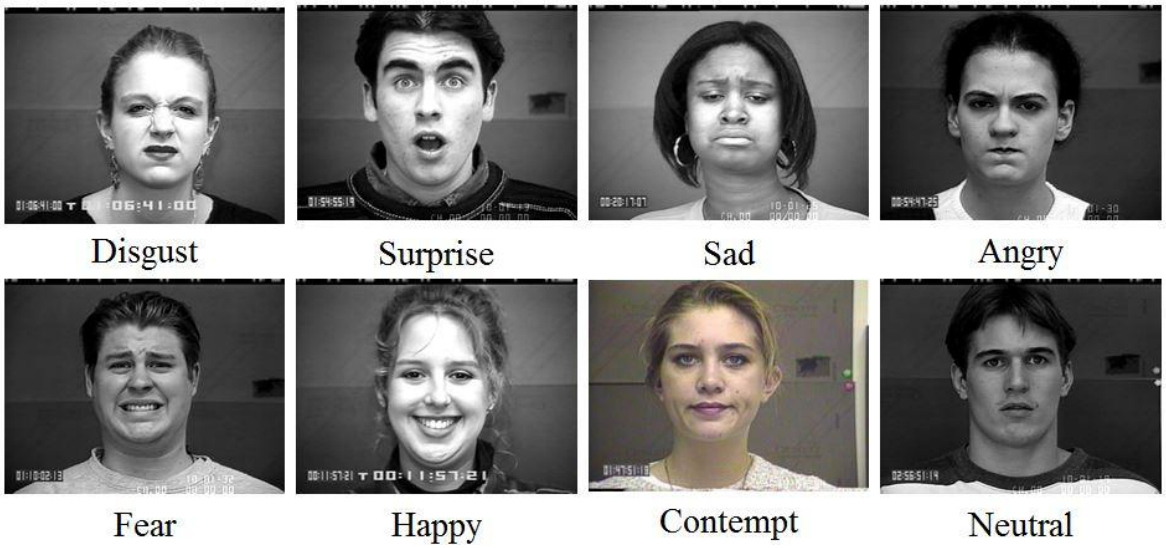

Fig. 3. The eight expressions with different subjects.

TABLE IV. THE CLASSIFICATION RATES OF EACH EXPRESSION WITH DIFFERENT METHODS

	Our method	SPTS [18]	CAPP [18]	SPTS+CAPP [18]
AN	0.84	0.35	0.70	0.75
CO	0.61	0.25	0.22	0.84
DI	0.95	0.68	0.95	0.95
FE	0.72	0.22	0.22	0.65
HA	0.96	0.98	1.0	1.0
SA	0.80	0.28	0.60	0.68
SU	0.99	1.0	0.99	0.96

IV. CONCLUSION

In this paper, we propose an effective method to handle the facial expression recognition problem. Instead of using the whole face, we detect and extract the facial components from the face image. Facial expressions are caused by facial muscle movements and these movements or subtle changes can be described by the HOG features, which are sensitive to the object shapes. The encoded features are used to train a linear SVM. Experiment results on two databases, JAFFE and the extended Cohn-Kanade dataset, show that our proposed method can achieve a good performance. The classification rates of our method on the two datasets are 94.3% and 88.7±2.3%, respectively. Facial expression recognition is a very challenging problem. More efforts should be made to improve the classification performance for important applications. Our future work will focus on improving the performance of the method in the wild environment and on the more subtle expressions such as “contempt”.

ACKNOWLEDGMENT

This work was partially supported by a research grant from The Hong Kong Polytechnic University (Project Code: G-YJ87).

REFERENCES

[1] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A Survey of Affect Recognition Methods Audio, Visual, and Spontaneous Expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 31, pp. 39-58, 2009.

[2] P. Ekman and W. V. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," *Consulting Psychologists Press,* 1978.

[3] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial Action Coding System: The Manual on CD ROM. A Human Face," 2002.

[4] S. Z. Li and A. K. Jain, *Handbook of face recognition*: springer, 2011.

[5] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Automatic Face and Gesture Recognition Proceedings. Third IEEE International Conference on*, 1998, pp. 454-459.

[6] X. Feng, A. Hadid, and M. Pietikäinen, "A coarse-to-fine classification scheme for facial expression recognition," in *Image Analysis and Recognition*, ed: Springer, 2004, pp. 668-675.

[7] S. Khandait, R. C. Thool, and P. Khandait, "Automatic facial feature extraction and expression recognition based on neural network," *International Journal of Advanced Computer Science and Applications,* vol. 2, pp. 113-118, 2012.

[8] L. Zhang and D. Tjondronegoro, "Facial expression recognition using facial movement features," *Affective Computing, IEEE Transactions on,* vol. 2, pp. 219-229, 2011.

[9] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *Image and Vision Computing,* vol. 27, pp. 803-816, 2009.

[10] T. Gritti, C. Shan, V. Jeanne, and R. Braspenning, "Local features based facial expression recognition with face registration errors," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, 2008, pp. 1-8.

[11] P. Viola and M. Jones, "Robust Real-Time Face Detection," *International journal of computer vision,* vol. 57, pp. 137-154, 2004.

[12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International journal of computer vision,* vol. 60, pp. 91-110, 2004.

[13] H. G. Feichtinger and T. Strohmer, *Gabor analysis and algorithms: Theory and applications*: Springer, 1998.

[14] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, pp. 971-987, 2002.

[15] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Computer Vision and Pattern Recognition, 2005. IEEE Conference on*, 2005, pp. 886-893.

[16] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 2, 2011.

[17] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," in *Automatic Face and Gesture Recognition, Proceedings. Third IEEE International Conference on*, 1998, pp. 200-205.

[18] 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*, 2010, pp. 94-101.

[19] G. Guo and C. R. Dyer, "Learning from examples in the small sample case: face expression recognition," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 35, pp. 477-488, 2005.