A NOVEL FEATURE EXTRACTION METHOD USING PYRAMID HISTOGRAM OF ORIENTATION GRADIENTS FOR SMILE RECOGNITION*

Yang Bai, Lihua Guo, Lianwen Jin+, Qinghua Huang

School of Electronic and Information
South China University of Technology, Guangzhou 510660, P.R. China
+Email: Lianwen.jin@gmail.com

ABSTRACT

Recognizing smiles is of much importance for detecting happy moods. Gabor features are conventionally widely applied to facial expression recognition, but the number of Gabor features is usually too large. We proposed to use Pyramid Histogram of Oriented Gradients (PHOG) as the features extracted for smile recognition in this paper. The comparisons between the PHOG and Gabor features using a publicly available dataset demonstrated that the PHOG with a significantly shorter vector length could achieve as high a recognition rate as the Gabor features did. Furthermore, the feature selection conducted by an AdaBoost algorithm was not needed when using the PHOG features. To further improve the recognition performance, we combined these two feature extraction methods and achieved the best smile recognition rate, indicating a good value of the PHOG features for smile recognitions.

Index Terms— Smile Recognition, Pyramid Histogram of Oriented Gradients, Gabor Feature, AdaBoost, Support Vector Machine

1. INTRODUCTION

Facial expression recognition (FER) has been one of the most pivotal parts in the field of pattern recognition for a long time. Smile is regarded as an important expression. Therefore, it is increasingly attracting attentions from areas of research and development. For instance, one of the applications for smile recognition is to automatically detect the smiles when taking photos. This technique has been commercialized as a new function in digital cameras. As a result, improving the recognition performance of such digital systems becomes an essential task.

Prior to the recognition, the features should be extracted from human face images. In recent years, many methods have been proposed for feature extraction [1-3]. Of these reported techniques, Gabor filter is regarded one of the most successful feature extraction methods. Littlewort et al. [3] proposed to select a subset of Gabor features using an AdaBoost method and thereafter train a Support Vector Machine algorithm using the selected features. They reported very high accuracies in recognition of facial expressions. However, this recognition algorithm is highly time-consuming because the Gabor filter produces an extremely large amount of features and the AdaBoost method is therefore needed for feature selection. Furthermore, as a frequently used feature extraction method, Gabor filter is performed on the whole region of human face and can be used for recognizing various expressions. However, for smile recognition which is emphasized in this study, it is normally deemed that the mouth region plays a much more important role in comparison with the other regions on a face. Unfortunately, as mentioned above, the Gabor filter based smile recognition methods [3] did not lay the stress on the features extracted from the mouth region.

To make the process of feature extraction and selection more efficient and improve the recognition rate by stressing the mouth region, we proposed to use Pyramid Histogram of Oriented Gradients (PHOG) for feature extraction in smile recognition. PHOG was firstly proposed by Bosch et al. [5] and has been successfully applied to object classification in recent years. As a spatial shape descriptor, it represents an image by its local shape and the spatial layout of the shape. In this study, we extracted PHOG features in the mouth region and the PHOG features would accordingly increase the weight of mouth in recognizing a smile. Because the PHOG method results in a much lower number of features compared with Gabor filter, the procedure of feature selection by the AdaBoost method is not required.

In order to examine the performance of PHOG features for smile recognition, we compared the PHOG with Gabor filter in terms of the recognition rate. For further improving the recognition accuracy, we used both of Gabor and PHOG features in the procedure of feature extraction. With an Adaboost method for feature selection and a Support Vector Machine (SVM) as the classifier, we obtained the best recognition performance on Cohn-Kanade AU-Coded Facial

* This paper is partially sponsored by the following research projects: NSFC (no. U0735004, 60772116), GDNSF (no. 07118074).

Expression Database [9], indicating a good value of the PHOG features.

The next section gives the descriptions for the Gabor filter and PHOG feature extraction methods. Section 3 describes the AdaBoost and SVM methods used for smile recognition. The experimental results are presented in section 4 and the conclusions are finally drawn in section 5.

2. FEATURE EXTRACTION

2.1. Gabor Feature

The Gabor wavelets (kernels, filters) [4] can be defined by

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-(\|k_{\mu,\nu}\|^2 \|z\|^2 / 2\sigma^2)} [e^{ik_{\mu,\nu}z} - e^{-(\sigma^2/2)}] \quad (1)$$

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu}$$

where μ and ν denote the orientation and scale of the Gabor filters, $z = (x, y)$, $\|\cdot\|$ denotes the norm operator, $k_\nu = k_{\max} / f^\nu$, and $\phi_\mu = \pi\mu/8$. f is the spacing factor between kernels in the frequency domain.

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined by (1). Five spatial and eight orientations were usually selected when extracting Gabor features from face images. For details, please refer to [4].

2.2. PHOG features

PHOG is a spatial shape descriptor applied to image classification recently [5]. It represents the spatial distribution of edges and is formulated as a vector representation. This descriptor is mainly inspired by two sources: (1) the use of the pyramid representation [6], and (2) the Histogram of Orientation Gradients (HOG) [7].

In this paper, the PHOG features are extracted from the region of mouth to stress the contribution of features related to the mouth motions. Smile and non-smile samples were represented by its local shape and the spatial layout of the shape around mouth. Here local feature was captured by the distribution over edge orientation within a region, and spatial layout by tiling the image into regions at multiple resolutions. As illustrated in Fig.1, the PHOG descriptor consists of a histogram of orientation gradients over each image sub-region at each resolution. The details of extracting PHOG features are as follows.

Step 1: Extracting edge contours. Giving a sample image, edge contours of the image are extracted first for further processing. As shown in Fig. 1(b), the information of edge contours is a description of shape. In this study, the edge contours were extracted using the Canny edge detector.

Step 2: The image is divided into cells at several pyramid level. As shown in Fig. 1(a), the grid at level l has 2^l cells along each dimension.

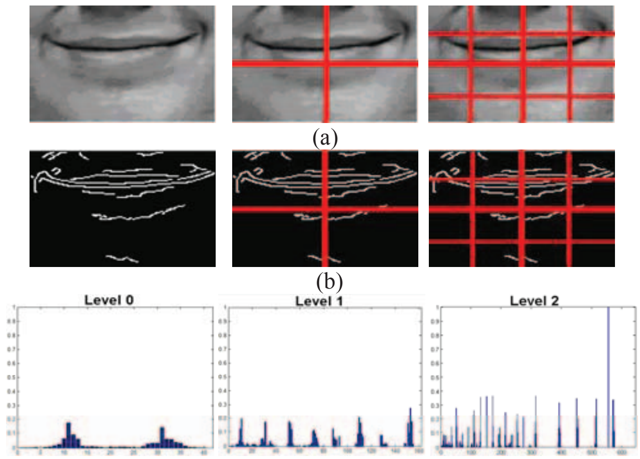

Fig. 1 Extraction of PHOG features from a smile image sample. (a) Grids at three pyramid resolution in the origin image, (b) edge contours are extracted using an edge detector, and (c) concatenation of all the HOG vectors in three pyramid resolutions to obtain the PHOG features of a sub-image.

Step 3: The HOG for each grid at each pyramid resolution level was computed. Local shape is represented by a histogram of edge orientations within an image sub-region quantized into *K* bins. Edge contours were located in step 1, and the orientation gradients were computed at edge contours in the original image. The orientation gradients were computed using 3×3 Sobel mask without Gaussian smoothing. The contribution of each edge was weighted according to its magnitude with a soft assignment to neighboring bins in a manner similar to SIFT [8]. Each bin in the histogram represents the number of edges that have orientations within a certain angular range.

Step 4: The final PHOG descriptor for an image is a concatenation of all the HOG vectors at each pyramid resolution. The concatenation of all the HOG vectors introduces the spatial information of the image. Each HOG is normalized to sum to unity taking into account all the pyramid levels. Consequently, level 0 is represented by a *K*-vector corresponding to the *K* bins of the histogram, level 1 by a 4*K*-vector, and the PHOG descriptor of the entire image is a vector with dimensionality $K\sum_{l\in L} 4^l$. For levels up to $L=1$ and $K=20$ bins it is a 100-vector.

In our experiment, the HOG descriptor was quantized into 40 orientation bins in the range of [0, 360]. We selected $L = 2$ and $K = 40$, then the descriptor is an 840-vector.

3. CLASSIFIERS

Having obtained features extracted using the Gabor filters and/or the PHOG descriptor, we performed three classifiers including a fast SVM algorithm [10], AdaBoost [11], and AdaBoost+SVM. For the first classifier, the SVM algorithm could make use of the extracted features for training and group face images into smile or non-smile class. For the 2nd classifier, AdaBoost was applied to both selecting features and training a classifier. The trained classifier was then used for smile recognition. For the 3rd classifier, the AdaBoost method was firstly conducted for only selecting features and the SVM classifier was trained using the selected features.

4. EXPERIMENTS AND RESULTS

In this section, we compared the two feature extraction methods experimentally. The smile recognition system was trained and tested using Cohn-Kanade AU-Coded Facial Expression Database [9]. This database includes approximately 2000 image sequences captured from over 200 subjects, and consists of 100 university students ranging in age from 18 to 30 years. A set of sample images are showed in Fig. 2.

4.1. Experimental Data

We selected 1315 frames from the dataset for our experiment. For each expression of a subject, the first 5 frames were selected as neutral expression images, and the last 5 frames were selected as peak expression images. Then these data were divided into two classes: smile and non-smile facial expressions. We finally had 431 positive and 884 negative samples. For the training, 339 positive samples and 692 negative samples were selected. The remaining images were used for testing. The images of the same class for training and testing were selected from different persons' sequences.

4.2. Comparison between PHOG and the Gabor Features

Before extracting the Gabor features, each of the face images was rotated to make the two eyes at the same height. We could obtain the face area according to the eyes' coordinates and the distance between the two eyes. The re-located face was then scaled to the size of 64×48 pixels. Finally, all images were converted into a Gabor magnitude representation using a bank of Gabor filters with eight orientations and five spatial frequencies. We extracted the Gabor features with a spacing step of 2. Accordingly, the Gabor filters generated 30,720 features. The best recognition rates obtained using the three classifiers were 94.334%, 93.309%, and 95.652%, respectively.

Fig. 2 A set of sample images obtained from Cohn-Kanade Database.

The PHOG features were obtained in the mouth area. The mouth area could be found after locating the coordinates of eyes and nose. We selected $L = 2$ and $K = 40$ so the descriptor was an 840-vector. The three classifiers mentioned in section 3 were used to train and test the extracted features. As demontrated in Table 1, the recognition rates achieved by the PHOG method were comparable with those produced by the Gabor features. In particular, the PHOG outperformed the Gabor filters when using only the SVM classifier. Although the Gabor features outperformed the PHOG features when using the classifiers of AdaBoost and AdaBoost+SVM, the feature selection could not be avoided and the training process was computationally expensive.

Table 1 Results of different feature extraction methods

Features	SVM	AdaBoost	Adaboost+SVM
Gabor	93.478%	93.661%	95.652%
PHOG	95.652%	92.958%	93.478%
PHOG+Gabor	93.478%	95.422%	96.739%

4.3. Combination of the Gabor and PHOG features

We simply combined the Gabor and PHOG features to form a new set of features, in which each value of PHOG was multiplied by a constant of 100. Let $X = \{x_1, x_2 ... x_m\}$ and $Y = \{y_1, y_2 ... y_n\}$ be the feature vectors of the Gabor and PHOG features, respectively, where m and n denoted the vector dimensions for X and Y, the new set (C) of features was defined as:

$$C = \{X \cup (100Y)\} = \{x_1, x_2 ... x_m, 100y_1, 100y_2 ... 100y_n\} \quad (2)$$

We used the Gabor and PHOG features obtained in section 4.2 to construct C. With the new feature set C, the recognition rates were obtained using the three classifiers as given in Table 1 respectively. Fig. 3 shows the learning curve of recognition rate against the training round. The optimal recognition rate was achieved when the training round was around 500.

According to Table 1, it can be found that the use of the PHOG features that were extracted from the region of mouth was able to improve the overall accuracy of smile recognition comparing with using only the Gabor features. What's more, the PHOG has a feature vector with a much shorter length, leading to a more rapid training process. In particular, the PHOG features trained by the SVM classifier generated as high a recognition rate as the Gabor features did, implying that the procedure for the AdaBoost training was not needed. Thus, the value of PHOG features in smile recognition can be verified. Furthermore, when combining the PHOG and Gabor features together, we achieve the best recognition accuracy of 96.73% using AdaBoost+SVM classifier.

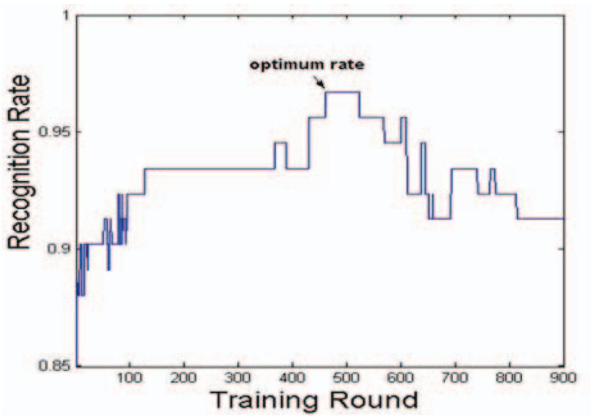

Fig. 3 Learning curve of recognition rate against of training round ith the combination of Gabor and PHOG features using Adaboost+SVM classifier.

5. CONCLUSIONS

In this paper, we proposed a novel feature extraction method based on PHOG features for smile recognition. The PHOG features with a much lower number of dimensions were extracted from the mouth area in face images and increased the contributions of mouth motions to the smile recognition. In the comparison with the Gabor features using a publicly available dataset, it achieved as high a recognition rate as the Gabor features did. In addition, the procedure of feature selection by AdaBoost was not required to achieve a good enough performance when using the PHOG features and the SVM classifier, which indicates less training efforts will be involved. Furthermore, we combined the Gabor and PHOG feature extraction methods before the classification, and obtained improved recognition rates. According to the experimental results, it has shown that the hybrid PHOG and Gabor features approach has achieved promising performance for smile recognition, and we expect that it will also has great potential for other facial expression recognition task, which merits our future study.

6. REFERENCES

[1] A.J.Calder, A.M.Buton, P.Miller, A.W.Young and S.Akamatsu, "A principal component analysis of facial expression", Vision Research, vol. 41, pp. 1179-1208, April 2001.

[2] P.N. Belhumeur, J.P.Hespanha and D.J.Kriegman, "Eigenfaces vs. fisherface: recognition using class specific linear projection", IEEE Transactions on Pattern Recognition and Machine Intelligence, vol.19, pp. 228-233, 1997.

[3] Littlewort.G, Bartlett. M. S, Fasel.I, Susskind.J, Movellan.J, "Dynamics of Facial Expression Extracted Automatically from Video", 2004 Conference on Computer and Pattern Recognition, vol.5, pp. 80, June 2004.

[4] C.Liu, H.Wechsles, "Independent Component Analysis of Gabor Features for Face Recognition", IEEE transactions on Neural Networks, vol. 14, pp. 919-928, July 2003.

[5] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel", In Proceedings of the International Conference on Image and Video Retrieval, pp. 401-408, 2007.

[6] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2169-2178, 2006.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886-893, June 2005.

[8] D. Lowe, "Distinctive image features from scale invariant keypoints", International Journal of Computer Vision, vol. 60, pp. 91-110, 2004.

[9] T.Kanade, J.Cohn, Y.Tian, "Comprehensive database for facial expression analysis", IEEE 4th International Conference on Face and Gesture Recognition, pp. 46-53, Mar 2000.

[10] Zhi-jie He, Lian-Wen Jin, "A new fast training algorithm for SVM", Machine Learning and Cybernetics, vol. 6, pp. 3451-3456, July 2008.

[11] Yoav Freund and Robert E. Schapire, "A decision-theoretic generalization of on-line learningand an application to boosting", Journal of Computer and System Sciences, vol. 55, pp. 119-139, August 1997.