Recognizing Facial Expressions Based on Gabor Filter Selection

Ziyang Zhang, Xiaomin Mu, Lei Gao
School of Information Engineering
Zhengzhou University
Zhengzhou, China

***Abstract*—Recognition of human emotional state is an important component for efficient human-computer interaction. In this paper a method of Gabor filter selection for facial expression recognition is investigated. We first preprocess facial images based on affine transform to normalize the faces. Then the using of a separability judgment is proposed to evaluate the separability of different Gabor filters, and only use those filters that can better separate different expressions. In the recognition process a PCA and FLDA multiclassifier scheme is used. The experiment result shows that the introducing of Gabor filter selection can not only reduce the dimension of feature space but also reduce the computation complexity significantly, while retaining high recognition rate of above 93%.**

Keywords-Facial expression recognition; Gabor filter selection; FLDA

I. INTRODUCTION

Facial expression recognition plays an important role in effective human-computer interaction, and can be applied to various situations such as patient monitoring and human behavior interpretation. Therefore it attracted a lot of research interests during recent years. Based on the work by Darwin [1], who proposed the existing of basic prototypic facial expressions, most researches classify 6 basic facial expressions: happiness, sadness, surprise, anger, disgust, fear, or 7 expressions that include neutral state.

There have been various kinds of methods to extract expression features from facial images. Among these methods, Gabor features have been used extensively after Lyons et al. [2] first proposed that the Gabor representation shows a significant degree of psychological plausibility. Rose [3] used Gabor wavelet to extract features from facial images, and then used PCA+LDA for dimension reduction and classification, finally got an 86% recognition accuracy on JAFFE database. Zhang et al. [4] adopted Gabor wavelet coefficients and geometric positions to construct the feature vector for each image and applied two-layer perceptron to distinguish seven different facial expressions. Their recognition rate is 90.1%. Shih et al. [5] compared the recognition accuracy of different recognition systems on JAFFE database, and found that the highest recognition rate achieved using Gabor features is 92.00%.

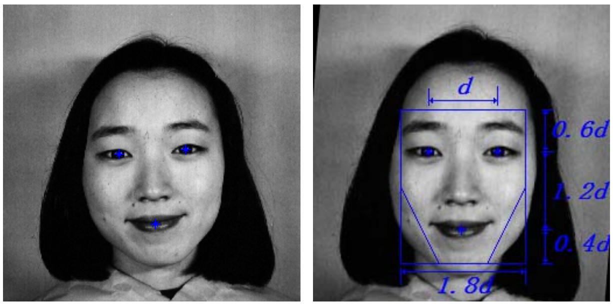
Although Gabor features do not require precise registration of images, Gabor filters suffer from the disadvantage of high dimensional feature space. This results from the fact that for each single filter in the Gabor filter bank, the output is equal in size with the original image, and the features extracted have great redundancy. Therefore, dimension reduction is needed, which was often achieved by principle component analysis (PCA). However, just using PCA is indeed limited, because it causes extra computation and does not reduce the time of feature extraction. In this paper, we propose a method for reducing the number of Gabor filters, based on separability analysis, and build a recognition system using a 2-class FLDA classifier scheme to evaluate the performance of filter selection.

The rest of this paper is organized as follows: Section II presents the preprocessing method of images. Section III introduces the Gabor filter selection method and feature extraction process. Design of classifier is discussed in Section IV. Section V gives the experiment results and conclusions are drawn in Section VI.

II. IMAGE PREPROCESSING

It is known that most of the facial expression information is concentrated around areas such as the eyes and the mouth. Including irrelevant areas like hair or the background might produce within-class dissimilarity and cause incorrect decisions, so it would be better to minimize any other possible variations except for the patterns we are classifying. Here we introduce a method similar with [6], based on affine transform, to extract and normalize the face region:

- Firstly, manually locate 3 key points (center of eyes and center of mouth) on the facial image. Then use an affine transform (which is a kind of coordinate transform that consists of translation, scaling and rotation) to locate the 3 points at fixed positions in the output image, so that faces in all images are approximately of the same size and in the same place, as shown in Fig. 1.
- Secondly, we use the geometric face model [7] illustrated in Fig. 1(b) to crop out the face rectangle, the size of which is 109×133 pixels in our case. We then crop the left and right lateral parts of faces to only consider their internal area.
- At last we perform histogram equalization to eliminate the variations in illumination and skin colors.



(a) (b)

Figure 1. Examples of image preprocessing.

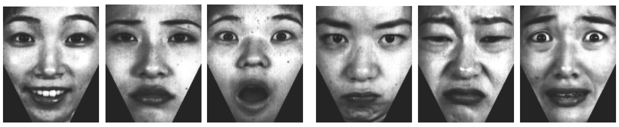


Figure 2. Example images after preprocessing.

Fig. 2 shows some example images after preprocessing. The original images are from JAFFE [2] (Japanese Female Facial Expression) database.

III. Gabor Filter Selection and Feature Extraction

A. Gabor functions and wavelets

A complex-valued two dimensional Gabor function can be given as follows [tutorial on Gabor filters]:

$$\mathrm{g}(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x_r^2}{\sigma_x^2}+\frac{y_r^2}{\sigma_y^2}\right)+j2\pi f_0 x_r\right], \quad (1)$$

where

$$x_r = x\cos\theta + y\sin\theta\,, \quad (2)$$

and

$$y_r = -x\sin\theta + y\cos\theta \quad (3)$$

represent a rotation. θ specifies the orientation of the parallel stripes of the filters. σ_x and σ_y are the standard deviation of the Gaussian factor and determine the size of its receptive field. The half-peak magnitude (3dB) support region of the Gabor filter is an ellipse centered at $u_0 = f_0\cos\theta$ and $v_0 = f_0\sin\theta$ in the spatial frequency domain, with its long and short axes equal to $2\sqrt{2\ln 2}/2\pi\sigma_x$ and $2\sqrt{2\ln 2}/2\pi\sigma_y$, respectively.

The use of scale factor a^{-m}, so that $\sigma_x{}'=\sigma_x\cdot a^m$, $\sigma_y{}'=\sigma_y\cdot a^m$ and $f_0{}'=f_0\cdot a^{-m}$, together with the orientation parameter θ, generate a bank of Gabor filters called Gabor wavelets. However, these filters are not orthogonal, resulting in the redundancy of information.

As the some redundancy comes from the overlapping of different filters in the spatial frequency domain, it is reasonable to consider minimizing the area of overlap, by placing the 3dB profiles of the filter responses so that these ellipses are approximately tangent with each other [8], as shown in Fig. 3. Let f_h denotes the center frequency of the first scale. Let K be the number of orientations and S be the number of scales. Then the design strategy results in the following formulas:

$$\sigma_u=\frac{(a-1)f_h}{(a+1)\sqrt{2\ln 2}},\ \sigma_v=\tan\left(\frac{\pi}{2K}\right)\cdot\frac{f_h}{\sqrt{2\ln 2}}\,. \quad (4)$$

This paper uses $a=2$, $f_h=0.4$, $S=4$, and $K=6$. These parameter values have been proved to generate good representation of image texture feature [8]. The filters' frequency responses are shown in Fig. 3.

B. Feature Extraction at Sample Points

Suppose the gray scaled image ready for feature extraction is expressed by $\mathrm{i}(x,y)$. Given a Gabor filter $\mathrm{g}_{mn}(x,y)$ which is of m th scale and n th orientation, the Gabor transform of the image is then defined as:

$$\mathrm{W}_{mn}(x,y)=\int \mathrm{i}(x_1,y_1)\cdot \mathrm{g}_{mn}(x-x_1,y-y_1)dx_1dy_1\,. \quad (5)$$

This convolution process produces an equal sized output image for each single Gabor filter. Even if we use only one filter, we get $109\times133=14497$ Gabor coefficients in total. The feature dimension is already too high for further processing. Considering that passing through the low-pass Gabor filter blurs the image, we may not use all the coefficients for further classification. Instead, this paper samples the filtered images at given points (10 rows, 8 columns in this case, with lateral points discarded, resulting in 62 points), whose locations in the original image is illustrated in Fig. 4. Then combine these sampled coefficients into a row vector $\mathbf{x}$. In consideration of convenience, we denote the whole procedure by:

$$\mathbf{x}=\mathrm{FeatureExtract}\left[\mathbf{i},\mathbf{g}_{mn}\right]. \quad (6)$$

C. Selecting Gabor Filters based on Separability Analysis

Though a set of filters has been build, whose frequency responses are approximately tangent to each other, they still have some redundancy, as long as they are not orthogonal.

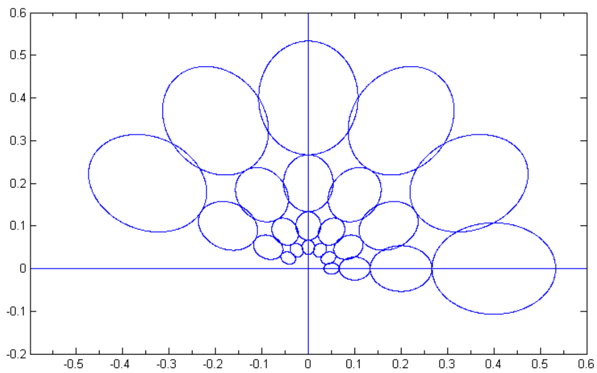


Figure 3. Half-peak (3dB) frequency responses of the Gabor filter bank

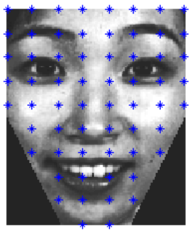


Figure 4. Locations of sample points

Besides, using so many filters (24 in this case) requires large amount of computation resource and generates high dimensional feature space. However, if we want to select a subset of filters from the original filter bank, which principle should we follow? This paper proposes the using of a separability judgment to evaluate the separability of a given filter, and then select the appropriate ones [9].

Suppose the sample image set is $\mathbf{I}$, which contains N images, and there are totally c classes, which are denoted as $w_1, w_2, \cdots w_c$. For each single filter $g_{mn}(x, y)$ in the filter bank we derived former, apply the following steps to compute its judgment function:

- Use the procedure defined in (7) to get the feature vector set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]$:

$$\mathbf{x}_k = \text{FeatureExtract}\left[\mathbf{i}_k, \mathbf{g}_{mn}\right],\ k = 1, 2, \cdots, N\ . \tag{7}$$

  Then the feature set $\mathbf{X}$ has the same class label with the original image set $\mathbf{I}$.

- Let $\mathbf{m}_i\ (i = 1, 2, \ldots c)$ denote the mean value of the feature vectors in class w_i, n_i represents the number of sample images in class w_i, and $\mathbf{m}$ denotes the mean value of feature vectors in all different classes. Then the within-class scatter matrix and between-class scatter matrix is defined as:

$$\mathbf{S}_w = \sum_{i=1}^{c} \sum_{\mathbf{x} \in \mathbf{w}_i} \left(\mathbf{x} - \mathbf{m}_i\right)\left(\mathbf{x} - \mathbf{m}_i\right)^T, \tag{8}$$

$$\mathbf{S}_b = \sum_{i=1}^{c} n_i \left(\mathbf{m}_i - \mathbf{m}\right)\left(\mathbf{m}_i - \mathbf{m}\right)^T \tag{9}$$

- We evaluate the separability of the feature set $\mathbf{X}$ using the following judgment function [9]:

$$\mathrm{J}(\mathbf{X}) = \frac{\mathrm{tr}(\mathbf{S}_b)}{\mathrm{tr}(\mathbf{S}_w)}, \tag{10}$$

  where tr() represents the trace of matrix.

This criterion is different with the Fisher criterion in that the latter one uses the determinant of the between-class and within-class scatter matrix, which are often close to singular and make it difficult to compute. Though this criterion is not directly related to the final recognition accuracy, it does render some information about the separability of the feature extracted. Experiment results will show that the separability value of feature from all filters is generally between the maximum and the minimum value of the features from single filter.

We will sort the filters based on the judgment value and use different number of filters that have the best separability. Experiment results will show that the discard of many filters with lower J value can hardly influence the final recognition result.

IV. Design of Classifier

For a c class problem, with a t dimensional feature space, there must be at least $t + c$ training samples to guarantee the feasibility of linear discriminant analysis (LDA) [10]. However, after Gabor filter selection, the feature dimension t is still higher than the total number of training samples. [11] and [12] propose the use of an intermediate space, i.e. before carrying out LDA, use PCA first to project the feature space into an intermediate space of $N - c$ dimensions, where N is the number of training samples. Here we adopt this method, using PCA to reduce the feature dimension to $N - c$:

$$\mathbf{y}_k = \mathbf{W}_{pca}^T \mathbf{x}_k,\ k = 1, 2, \cdots, N\ . \tag{11}$$

Then a 2-class FLDA multiclassifier scheme is build for the recognition process [13]. FLDA assumes the discriminant function to be a linear function of the feature data. In this case the data is the feature vector obtained in (12). Thus for each of the 6 facial expressions, construct a 2-class discriminant function:

$$g_i(\mathbf{y}) = \mathbf{W}_i^T \mathbf{y} + \mathbf{w}_{0i},\ i = 1, 2, \cdots c\ , \tag{12}$$

where $\mathbf{W}_i$ is a vector, and $\mathbf{w}_{0i}$ is a constant. Fisher's LDA finds the vector $\mathbf{W}_i$ that best separates the two classes, by maximizing the ratio of the determinants of between-class and within-class scatter matrix.

This paper builds 6 such 2-class classifiers, each corresponding to one of the 6 basic expressions (HA, SA, SU, AN, DI, FE). In the training process, for each classifier, we label all the training samples that do not belong to the corresponding expression as one class. The output of each classifier is the probability of belonging to the corresponding expression. For example, in the "HA" classifier, all the samples of happiness are labeled as "*happiness*", while all other training samples are labeled as "*non-happiness*". If the output is greater than 50%, the sample may be classified as "*happiness*", and otherwise "*non-happiness*".

The output of each classifier is used to determine the final classification result. The decision rule here is that, if the output is greater than 50% in one or more than one classifiers, we classify the sample to the class that has the greatest output value. Otherwise, if all the classifiers' outputs are below 50%, label the sample as neutral. The classification scheme is shown in Fig. 5.

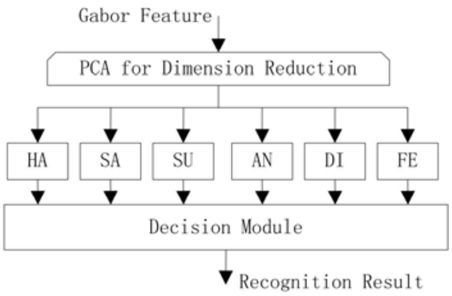


Figure 5. Multi-classifier scheme

V. EXPERIMENT RESULTS AND ANALYSIS

The database used here to examine our recognition system is the JAFFE [2] database. It contains 10 female expressors, each delivering 7 facial expressions (happiness, sadness, surprise, anger, disgust, fear, as well as neutral). There are 213 gray scaled images in total. This paper uses the leave-one-out strategy [5] to test the recognition rate.

A. Separability of Filters

After performing Gabor filter selection based on all the images in JAFFE, the separability judgment value of each single filter in the filter bank is shown in Table 1.

From Table 1 we can see that features from scale 2 and scale 3 have better overall separability than scale 0 and scale 1, and the separability value decreased as scale decreases. It shows that filters of lower frequency can give more separable information about facial expressions. Besides, result also shows that within each scale, filters of direction 2 and 3 generally have larger separability value, showing that filters at the orientation of approximately 60~120 degrees are more suitable in facial expression recognition.

B. Recognition Rate

We sort the 24 filters in terms of their separability judgment value. To evaluate the performance of different number of filters, the first k filters that have the largest separability value are used for feature extraction, where the number k ranges from 4 to 24. The minimum number 4 is set to make sure the feature dimension is larger than the number of training samples, to enable the use of PCA in the classification method. The recognition rates and computation times of different number k are given in Table 2 and illustrated in Fig. 6. The time data is collected in the environment of Intel(R) Core(TM) 2 Duo CPU T6570, 3GB RAM, with Windows 7 and MATLAB 7.4 software installed. The time value is account for the training process and the test of one sample image.

Results show that the use of all 24 filters generates the highest recognition rate of 95.77%, and when filter number decreases, the recognition rate also decreases. This indicates that all the filters are helpful for classification.

When using all the filters, the time of computation is 38.92 seconds, which might be too long for a practical system. When the number of filters decreases, the time of computation gets shorter, approximately in an exponential way. However, the range of decrease of recognition rate is small. When filter number is more than 8, the recognition rates are all above 90%. And when the number of filters goes up to 9, the recognition rate rises to 93.90%, which is not a large lose from the highest 95.77%, while the time of computation is reduced to 1.88 seconds, 20 times less than the highest 38.92 seconds. The recognition rate 93.90% is already higher than many existing systems [3,4,6].

This result demonstrates that the discard of filters with lower separability makes very limited influence on the final recognition rate, while significantly reduces the time of computation and makes the recognition system more efficient.

On the other hand, the high recognition rate of our system may result from the accuracy of the preprocessing of images. Because in the normalized facial images, eyes and mouths are approximately at the same location, which minimizes the irrelevant variations except for the expression differences we are recognizing.

VI. CONCLUSION

This paper proposes a facial expression recognition system based on Gabor filter selection. The goal of selecting Gabor filters is to reduce the computational complexity, while retaining good recognition rates. First a more precise face normalization method is introduced to the Gabor-based system. Then filter selection is performed based on separability analysis. This paper proposes the use of a separability judgment to evaluate each single filter in the bank of Gabor filters, and chose those filters with better separability. At last a PCA and FLDA based multiclassifier scheme is used for recognition.

TABLE I. SEPARABILITY JUDGMENT VALUE OF FILTERS

	Orientation					
	0	***1***	***2***	***3***	***4***	***5***
Scale 0	1.3796	1.4710	1.7896	3.2997	2.4374	1.5516
Scale 1	1.9285	4.0514	4.7343	3.8087	5.5104	3.5472
Scale 2	3.7134	5.6872	7.5386	5.7061	6.8071	5.4168
Scale 3	3.4735	5.0914	8.6913	8.3066	10.8978	5.0698

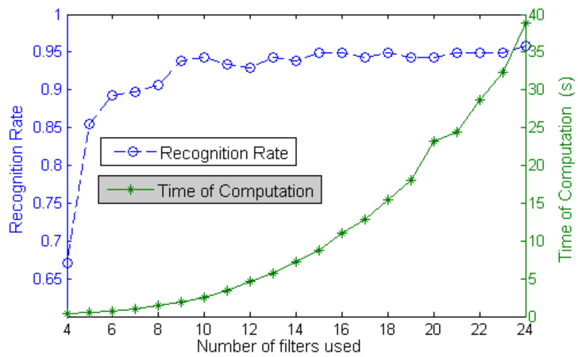


Figure 6. Experiment results with different k values

TABLE II. EXPERIMENT RESULT

Number of Filters	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***
Recognition Rate	0.6714	0.8545	0.8920	0.8967	0.9061	0.9390	0.9437	0.9343	0.9296	0.9437	0.9390
Time	0.3509	0.4909	0.7315	0.9870	1.3864	1.8764	2.5507	3.4484	4.6022	5.6740	7.1835
Number of Filters	***15***	***16***	***17***	***18***	***19***	***20***	***21***	***22***	***23***	***24***	
Recognition Rate	0.9484	0.9484	0.9437	0.9484	0.9437	0.9437	0.9484	0.9484	0.9484	0.9577	
Time	8.7971	11.0923	12.8921	15.4897	17.9800	23.2030	24.4697	28.7090	32.3913	38.9191	

Experiment results show that our proposed method can significantly reduce the time of computation (from 38.92 seconds to 1.88 seconds), while the recognition rate only decreases from 95.77% to 94.00%, which is still higher than most existing systems. Although by no means conclusive, this study sheds some very interesting lights and leads to new directions in facial expression recognition.

However, the good performance of our system may be partly due to the precise preprocessing of facial images, based on affine transform. The positions used to carry out affine transform is located manually, thus may not be suitable for an automatic system. In the future a more accurate automatic method to normalize faces will make great improvement on the recognition system.

REFERENCES

[1] C. Darwin, The Expression of Emotions in Man and Animals, John Murray, reprinted by Univ. of Chicago Press, 1965.

[2] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with Gabor wavelets,” Proceedings of the Third IEEE Conference on Face and Gesture Recognition, Nara, Japan, April 1998, pp.200-205.

[3] N. Rose, “Facial expression classification using Gabor and log-Gabor filters,” Proceedings of the 7th international conference on automatic face and gesture recognition (FGR'06), 2006.

[4] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, “Comparison between geometry-based and Gabor-wavelets-based facial expression recognition usingmulti-layer perceptron,” Proceedings of the Third IEEE Conference on Face and Gesture Recognition, Nara, Japan, April 1998, pp.200-205.

[5] Frank Y. Shih, Chao-Fa Chuang and Patrick S. P. Wang, “Performance comparisons of facial expression recognition in JAFFE database,” International Journal of Pattern Recognition and Artificial Intelligence, Vol. 22, No. 3 (2008) 445–459.

[6] S. Dubuisson, F. Davoine and M. Masson, “A solution for facial expression representation and recognition,” Signal Processing: Image Communication, 17 (2002) 657–673.

[7] Frank Y. Shih and Chao-Fa Chuang, “Automatic extraction of head and face boundaries and facial features,” Information Science, 158 (2004) 117-130.

[8] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data,” IEEE Trans. Pattern Anal. Machine Intell., vol. 18, no. 8, pp. 837-842, Aug. 1996.

[9] Xiao-Yuan Jing, Yuan-Yan Tang and David Zhang, “A Fourier-LDA approach for image recognition,” Pattern Recognition, 38 (2005) 453-457.

[10] R. O. Duda, P. E. Hart and D. G. Stork, Patter Classification, 2nd Edition, John Wiley & Sons, 2001.

[11] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenface vs. Fisherface: recognition using class specific linear projection,” IEEE Trans. Pattern Anal. Machine Intell., vol. 19, no. 7, pp. 711-720, July 1997.

[12] A. M. Martinez and A. C. Kak, “PCA versus LDA,” IEEE Trans. Pattern Anal. Machine Intell., vol. 23, no. 2, pp. 228-233, Feb. 1997.

[13] Yongjin Wang and Ling Guan, “Recognizing human emotional state from audiovisual signals,” IEEE trans. Multimedia, vol. 10, no. 5, pp. 936-946, Aug. 2008.