Face Recognition with Local Gabor Textons

Zhen Lei, Stan Z. Li, Rufeng Chu, and Xiangxin Zhu

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100080, China
{zlei,szli,rfchu,xxzhu}@nlpr.ia.ac.cn
http://www.cbsr.ia.ac.cn

Abstract. This paper proposes a novel face representation and recognition method based on local Gabor textons. Textons, defined as a vocabulary of local characteristic features, are a good description of the perceptually distinguishable micro-structures on objects. In this paper, we incorporate the advantages of Gabor feature and textons strategy together to form Gabor textons. And for the specificity of face images, we propose local Gabor textons (LGT) to portray faces more precisely and efficiently. The local Gabor textons histogram sequence is then utilized for face representation and a weighted histogram sequence matching mechanism is introduced for face recognition. Preliminary experiments on FERET database show promising results of the proposed method.

Keywords: local textons, Gabor filters, histogram sequence, face recognition.

1 Introduction

Face recognition has attracted much attention due to its potential values for applications as well as theoretical challenges. Due to the various changes by expression, illumination and pose etc, face images always change a lot on grey level. How to extract representation robust to these changes becomes an important problem.

Up to now, many representation approaches have been introduced, including Principal Component Analysis (PCA) [11], Linear Discriminant Analysis (LDA) [2], Independent Component Analysis (ICA) [3] etc. PCA provides an optimal linear transformation from the original image space to an orthogonal eigenspace with reduced dimensionality in sense of the least mean square reconstruction error. LDA seeks to find a linear transformation by maximizing the ratio of between-class variance and the within-class variance. ICA is a generalization of PCA, which is sensitive to the high-order relationships among the image pixels.

Recently, the textons based representations have achieved great success in texture analysis and recognition. The term texton was first proposed by Julesz [6] to describe the fundamental micro-structures in natural images and was considered as the atoms of pre-attentive human visual perception. However, it is a vague concept in the literature because of lacking a precise definition for grey level images. Leung and Malik [7] re-invent and operationalize the concept of textons. Textons are defined as a discrete set which is referred to as the vocabulary of local characteristic features of objects. The goal is to build the vocabulary of textons to describe the perceptually distinguishable

micro-structures on the surfaces of objects. Various of this concept have been applied to the problem of 3D texture recognition successfully [4,7,12].

In this work, we propose a novel local Gabor textons histogram sequence for face representation and recognition. Gabor wavelets, which capture the local structure corresponding to spatial frequency, spatial localization, and orientation selectivity, have achieved great success in face recognition [13]. Therefore, we incorporate the advantages of Gabor feature and textons strategy to form Gabor textons. Moreover, In order to depict the face images precisely and efficiently, we propose local Gabor textons (LGT) and then utilize LGT histogram sequence for face representation and recognition.

The rest of this paper is organized as follows. Section 2 details the construction of local Gabor textons. Section 3 describes LGT histogram sequence for face representation and recognition. The experimental results on FERET database are demonstrated in Section 4 and in Section 5, we conclude this paper .

2   Local Gabor Textons (LGT) Construction

Texture is often characterized by its responses to a set of orientation and spatial-frequency selective linear filters which is inspired by various evidences of similar processing in human vision system. Here, we use Gabor filters of multi-scale and multi-orientation which have been extensively and successfully used in face recognition[13,8] to encode the local structure attributes embedded in face images. The Gabor kernels are defined as follows:

$$\psi_{\mu,\nu} = \frac{k_{\mu,\nu}^2}{\sigma^2} \exp(-\frac{k_{\mu,\nu}^2 z^2}{2\sigma^2})[\exp(ik_{\mu,\nu}z) - \exp(-\frac{\sigma^2}{2})] \tag{1}$$

where μ and ν define the orientation and scale of the Gabor kernels respectively, $z = (x, y)$, and the wave vector $k_{\mu,\nu}$ is defined as follows:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{2}$$

where $k_\nu = k_{max}/f^\nu$, $k_{max} = \pi/2$, $f = \sqrt{2}$, $\phi_\mu = \pi\mu/8$. The Gabor kernels in (1) are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotating via the wave vector $k_{\mu,\nu}$. Each kernel is a product of a Gaussian envelope and a complex plane wave, and can be separated into real and imaginary parts. Hence, a band of Gabor filters is generated by a set of various scales and rotations.

In this paper, we use Gabor kernels at five scales $\nu \in \{0, 1, 2, 3, 4\}$ and four orientations $\mu \in \{0, 2, 4, 6\}$ with the parameter $\sigma = 2\pi$ [8] to derive 40 Gabor filters including 20 real filters and 20 imaginary fitlers which are shown in Fig. 1. By convoluting face images with corresponding Gabor kernels, for every image pixel we have totally 40 Gabor coefficients which are successively clustered to form Gabor textons.

Textons express the micro-structures in natural images. Compared to the periodic changes of texture images, for face images, different regions of faces usually reflect different structures. For example, the eye and nose areas have distinct differences. If we cluster the textons over the whole image used in texture analysis, the size of textons vocabulary could be very large if one wants to keep describing the face precisely, which will increase the computational cost dramatically. In order to depict the face in
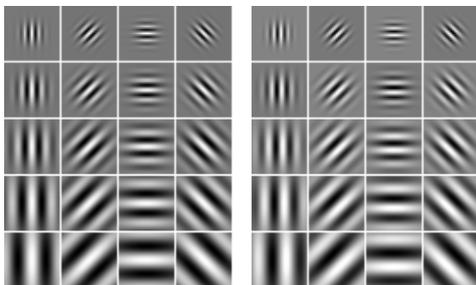
Fig. 1. Gabor filters with 5 scales and 4 orientations. Left are the real parts and right are the imaginary ones.

Fig. 2. An example of face image divided into 7×8 regions with size of 20×15

a more efficient way, we propose local Gabor textons to represent the face image. We first divide the face image into several regions with the size of $h \times w$ (Fig. 2), then for every region, the Gabor response vectors are clustered to form a local vocabulary of textons which are called local Gabor textons (LGT). Therefore, a series of LGT can be constructed corresponding to different regions. Specifically, we invoke K-means clustering algorithm [5] to determine the LGT among the feature vectors. K-means method is based on the first order statistics of data, and finds a predefined number of centers in the data space, while guaranteing that the sum of squared distances between the initial data points and the centers is minimized. In order to improve the efficiency of the computation, the LGT vocabulary L_i corresponding to the region R_i is computed by the following hierarchical K-mean clustering alogrithm:

1. Suppose we have 500 face images in all. These images are divided into 100 groups randomly, each of which contains 5 images. All of them are encoded by Gabor filters.
2. For each of the group, the Gabor response vectors at every pixel in region R_i with the size of $h \times w$ are concatenated to form a set of $5 \times h \times w$ feature vectors. The K-mean clustering algorithm is then applied to these feature vectors to form $k^{'}$ centers.
3. The centers for all the groups are merge together and the K-mean clustering algorithm is applied again to these $100 \times k^{'}$ centers to form k centers.
4. With the initialization of the k centers derived in step 3, apply K-means clustering algorithm on all images in region R_i to achieve a local minimum.

These k centers finally constitute the LGT vocabulary L_i corresponding to region R_i and after doing these operations in all regions, the local Gabor textons vocabularies are constructed.

3 LGT Histogram Sequence for Face Representation and Recognition

With the LGT vocabularies , every pixel in region R_i of an image is mapped to the closest texton element in L_i according to the Euclidean distance in the Gabor feature space. After this operation in all regions, a texton labeled image $f_l(x, y)$ is finally formed with the values between 1 and k .

The LGT histogram, denoted by $H_i(\ell)$, is used to describe the distribution of the local structure attributes over the region R_i. A LGT histogram with the region R_i of the labeled image $f_l(x, y)$ induced by local Gabor textons can be defined as

$$H_i(\ell) = \sum_{(x,y) \in R_i} I\{f_l(x, y) = \ell\}, \ell = 0, \ldots, k - 1, \tag{3}$$

in which k is the number of different labels and

$$I\{A\} = \begin{cases} 1, & \text{A is true} \\ 0, & \text{A is false} \end{cases}$$

The global description of a face image is built by concatenating the LGT histograms $\mathscr{H} = (H_1, H_2, \ldots, H_n)$ which is called local Gabor textons histogram sequence. The collection of LGT histograms is an efficient face representation which contains information about the distribution of the local structure attributes, such as edges, spots and flat areas over the whole image. Fig. 3 shows the process of face representation using LGT histogram sequence. The similarity of different LGT histogram sequences extracted from different images is computed as follows:

$$S(\mathscr{H}, \mathscr{H}') = \sum_{i=1}^{n} S_{\chi^2}(H_i, H_i') \tag{4}$$

where

$$S_{\chi^2}(H_i, H_i') = \sum_{\ell=1}^{k} \frac{(H_i(\ell) - H_i'(\ell))^2}{(H_i(\ell) + H_i'(\ell))} \tag{5}$$

is the Chi square distance commonly used to match two histograms, in which k is the number of bins for each histogram.

Pervious work has shown different regions of face make different contributions for face recognition [14,1], e.g., the areas nearby eyes are more important than others. Therefore, different weights can be set to different LGT histograms when measure the similarity of two images. Thus, (4) can be rewritten as:

$$S'(\mathscr{H}, \mathscr{H}') = \sum_{i=1}^{n} W_i S_{\chi^2}(H_i, H_i') \tag{6}$$
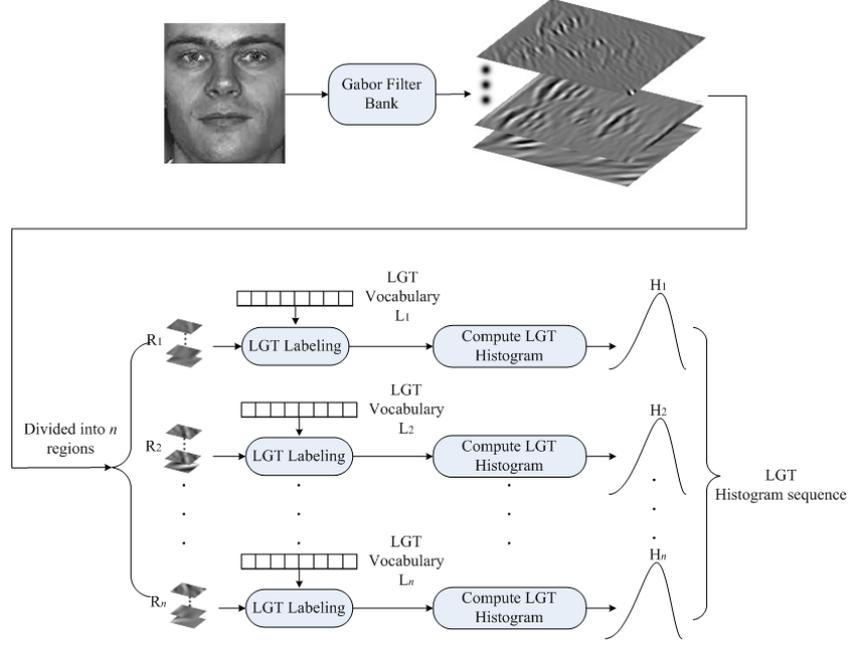
Fig. 3. Face representation using local Gabor textons histogram sequence

where W_i is the weight for i-th LGT histogram and is learned based on Fisher separation criterion [5] as follows.

For a C class problem, the similarities of different images from the same person compose the intra-personal similarity class and those of images from different persons compose the extra-personal similarity as introduced in [9].

For the i-th LGT histogram H_i, the mean and the variance of the intra-personal similarities can be computed by

$$m_{i\_intra} = \frac{1}{N_{intra}} \sum_{j=1}^{C} \sum_{p=1}^{N_j-1} \sum_{q=p+1}^{N_j} S_{\chi^2}(H_i^{p,j}, H_i^{q,j}) \tag{7}$$

$$\sigma_{i\_intra}^2 = \frac{1}{N_{intra}} \sum_{j=1}^{C} \sum_{p=1}^{N_j-1} \sum_{q=p+1}^{N_j} (S_{\chi^2}(H_i^{p,j}, H_i^{q,j}) - m_{i\_intra})^2 \tag{8}$$

where $H_i^{p,j}$ denotes the i-th LGT histogram extracted from the p-th image from the j-th class and $N_{intra} = \sum_{j=1}^{C} \frac{N_j(N_j-1)}{2}$ is the number of intra-personal sample pairs.

Similarly, the mean and the variance of the extra-personal similarities of the i-th LGT histograms can be computed by

$$m_{i\_extra} = \frac{1}{N_{extra}} \sum_{j=1}^{C-1} \sum_{s=j+1}^{C} \sum_{p=1}^{N_j} \sum_{q=1}^{N_s} S_{\chi^2}(H_i^{p,j}, H_i^{q,s}) \tag{9}$$

$$\sigma^2_{i\_extra} = \frac{1}{N_{extra}} \sum_{j=1}^{C-1} \sum_{s=j+1}^{C} \sum_{p=1}^{N_j} \sum_{q=1}^{N_s} (S_{\chi^2}(H_i^{p,j}, H_i^{q,s}) - m_{i\_extra})^2 \qquad (10)$$

where $N_{extra} = \sum_{j=1}^{C-1} \sum_{s=j+1}^{C} N_j N_s$, is the number of extra-personal sample pairs.

Then, the weight for i-th LGT histogram W_i is derived by the following formulation

$$W_i = \frac{(m_{i\_intra} - m_{i\_extra})^2}{\sigma^2_{i\_intra} + \sigma^2_{i\_extra}} \qquad (11)$$

Finally, the weighted similarity of LGT histogram sequences (6) is used for face recognition with a nearest neighbor classifier.

4   Experimental Results and Analysis

In this section, we analyze the performance of the proposed method using the FERET database. The FERET [10] database is a standard testbed for face recognition technologies. In our experiments, we use a subset of the training set containing 540 images from 270 subjects for training. And four probes against the gallery set containing 1196 images are used for testing according to the standard test protocols. The images in fb and fc probe sets are with expression and lighting variation respectively, and dup I, dup II probe sets include aging images. All images are rotated, scaled and cropped to 140×120 pixels according to the eye positions and then preprocessed by histogram equalization. Fig. 4 shows some examples of the cropped FERET images.

Fig. 4. Example FERET images used in our experiments

There are some parameters which influence the performance of the proposed algorithm. The first one is the size of the local region. If the size is too big (e.g. the whole image), it could lose the local spatial information and may not reveal the advantage of the local analysis. On the other hand, if the region is too small, it would be sensitive to the mis-alignment. Another parameter needed to be optimized is the size of LGT vocabulary (the number of local Gabor textons in every vocabulary). To determine the value
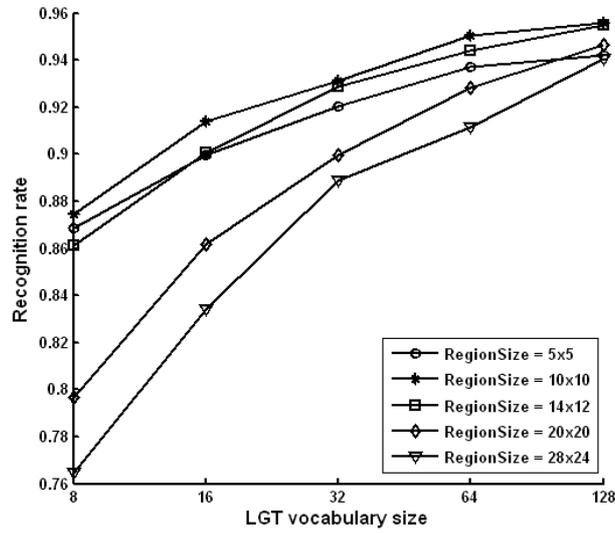
Fig. 5. The recognition rates of varying the values of parameters: the size of LGT vocabulary and the region size
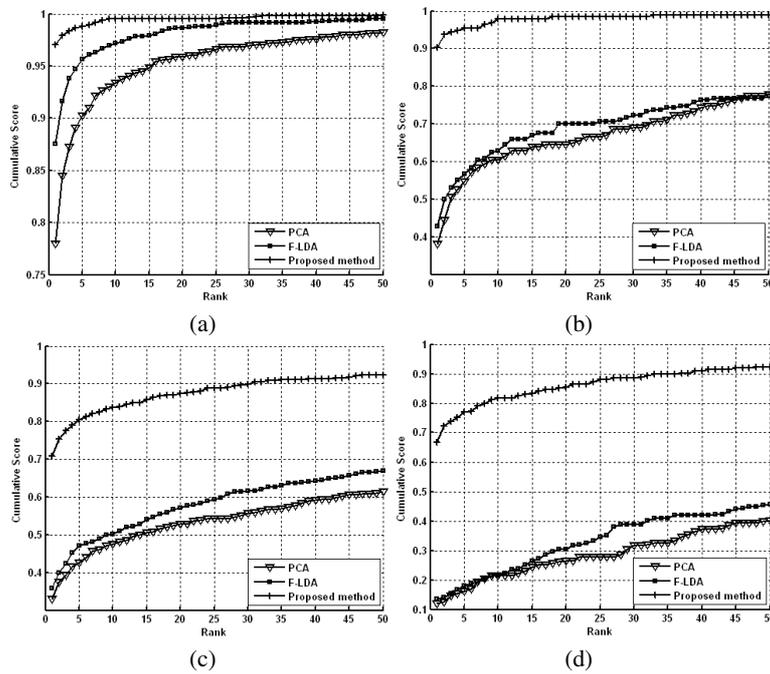

Fig. 6. Cumulative match curves on fb (a), fc (b), dup I (c) and dup II (d) probe sets

of the parameters, we use the training set to cluster the local Gabor textons and test on fb probe set to evaluate the performance of varying the values of parameters. The result is shown in Fig. 5. Note the similarities of histogram sequences are not weighted here.

The region size varies from 5×5 to 28×24 and the size of LGT vocabulary varies from 8 to 128. As expected, too larger region size results in a decreased recognition rate because of the loss of spatial information. Considering the trade-off between recognition and computational cost, we choose the region size of 10×10 and the LGT vocabulary size of 64. Therefore, the face image here is divided into 14×12 regions with the size of 10×10 and for every region, 64 local Gabor textons and the weight of the corresponding LGT histogram are learned from the training set as described in Section 2 and 3. Finally, the performance of the proposed method is tested on four probe sets against the gallery. Fig. 6 demonstrates the cumulative match curves (CMC) on four probe sets and Table 1 shows the rank-1 recognition rates of the proposed method compared to some well-known methods.

Table 1. The rank-1 recognition rates of different algorithms on the FERET probe sets

Methods	fb	fc	dup I	dup II
PCA	0.78	0.38	0.33	0.12
LDA	0.88	0.43	0.36	0.14
Best Results of [10]	0.96	0.82	0.59	0.52
The proposed method	**0.97**	**0.90**	**0.71**	**0.67**

From the results, we can observe that the performance of the proposed method is very comparable with the other well-known methods. It significantly outperforms PCA and LDA methods and is better than the best results in [10] in all of the four probe sets. These results indicate the proposed method is accurate and robust to the variation of expression, illumination and aging.

5 Conclusions

We have studied the use of textons as a robust approach for face representation and recognition. In particular, we propose local Gabor textons extracted by using Gabor filters and K-means clustering algorithm in local regions. In this way, we incorporate the effectiveness of Gabor features and robustness of textons strategy simultaneously and depict the face images precisely and efficiently. The local Gabor textons histogram sequence is then proposed for face representation and a weighted histogram sequence matching mechanism is introduced for face recognition. The preliminary results on FERET database show the proposed method is accurate and robust to the variations of expression, illumination and aging etc. However, one drawback of our method is that the length of the feature vector used for face representation slows down the recognition speed indeed. A possible choice is to apply subspace methods such as PCA, LDA etc. to reduce the dimensionality of the feature vectors which will be tested in our future work.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In: Proceedings of the European Conference on Computer Vision, Prague, Czech, pp. 469–481 (2004)
2. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. PAMI 19(7), 711–720 (1997)
3. Comon, P.: Independent component analysis - a new concept? Signal Processing 36, 287–314 (1994)
4. Cula, O., Dana, K.: Compact representation of bidirectional texture functions. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1041–1047. IEEE Computer Society Press, Los Alamitos (2001)
5. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley, Chichester (2000)
6. Julesz, B.: Texton, the elements of texture perception, and their interactions 290(5802), 91–97 (March 1981)
7. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision 43(1), 29–44 (2001)
8. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Transactions on Image Processing 11(4), 467–476 (2002)
9. Moghaddam, B., Jebara, T., Pentland, A.: Bayesian face recognition. Pattern Recognition 33(11), 1771–1782 (2000)
10. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
11. Turk, M.A., Pentland, A.P.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
12. Varma, M., Zisserman, A.: Classifying images of materials: achieving viewpoint and illumination independence. In: Proceedings of the European Conference on Computer Vision, pp. 255–271 (2002)
13. Wiskott, L., Fellous, J., Kruger, N., malsburg, C.V.: Face recognition by elastic bunch graph matching. IEEE Trans. PAMI 19(7), 775–779 (1997)
14. Zhang, W.C., Shan, S.G., Gao, W., Zhang, H.M.: Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In: Proceedings of IEEE International Conference on Computer Vision, pp. 786–791. IEEE Computer Society Press, Los Alamitos (2005)