Combining Null Space-based Gabor Features for Face Recognition

Wei Fan, Yunhong Wang, Wei Liu, Tieniu Tan
National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, P.O.Box 2728, Beijing, 100080, P.R.China
E-mail: {wfan, wangyh, wliu, tnt}@nlpr.ia.ac.cn

Abstract

We propose a novel face recognition strategy combining various discriminating Gabor features in multi-scales and multi-orientations. A bank of well-chosen Gabor filters is applied on the image to construct a group of feature vectors, and then the Null Space-based LDA (NLDA) is performed simultaneously on each orientation channel and the original image to give 5 component classifier outputs, which are then combined to increase the final recognition rate. Experimental results on the FERET database demonstrate the effectiveness and flexibility of our proposed method.

1. Introduction

Despite its inferior accuracy to most other biometric systems, automatic face recognition (AFR) has always been a major focus of research interest for its non-invasive nature and tremendous potentials in commercial and law enforcement applications. Exhaustive survey of AFR techniques [10] indicate the primary difficulty in recent dominating approaches comes from the immense variability of facial imagery due to several confounded factors such as illumination, viewpoint, body movement, and facial expression.

Evidence from psychological studies suggests that, instead of working exclusively with a global, holistic face representation, human vision system (HVS) seems to use both holistic and dominant feature information for the perception and recognition of faces. Whereas a precise search for the main facial features, such as eyes, noses and mouths, still remains very challenging in real applications, hybrid systems using both geometric and photometric information lack in a solid groundwork as they are over-sensitive to those detected feature points/landmarks.

We therefore focused our research towards extracting some significant space and spatial frequency features contained in the 2D image, which jointly characterize a face pattern in multi-scales and multi-orientations. Gabor filters, or other similar wavelets, which achieve such optimal joint localization in both space and frequency domains, have shown to be highly useful in many AFR applications such as facial feature extraction and pose estimation. Lades et al [5] pioneered the use of Gabor filters for face recognition using the Dynamic Link Architecture framework, which was later expanded by Wiskott et al [9] to the "Elastic Bunch Graph Matching" method. Liu et al [6] employed an enhanced Fisher discrimination model (EMD) on an augmented Gabor feature vector, which was derived from the Gabor transformation of face images. The underline basis for these effective approaches is the observation that the textual pattern of human face is fairly constant despite some slight variations introduced by the factors discussed at the beginning.

Unlike the aforementioned approaches, we make an experimental analysis of all the Gabor features from the data fusion standpoint. Given a normalized facial image, a bank of well-chosen Gabor filters is applied on it to construct a group of feature vectors. A "channel-based" set of separate subspaces are built by the Null Space-based Linear Discriminate Analysis (NLDA) of the original images and all the feature vectors corresponding to their respective Gabor channels. As different channel is tuned to different texture information, we believe a combination of these complementary representations should lead to improved recognition. Following the theoretical framework presented in [4], we compare some mainstream rules for combining classifier outputs. Experimental results conducted on the FERET database strongly support our assumption and show high superiority of the newly developed method to those outputs from any individual classifiers.

2. Gabor Feature Extraction

Biological evidence has shown that Gabor filters seem to be a good approximation to the sensitivity profiles of neurons in the visual cortex, and their mathematical properties provide a fair degree of insensitivity to irrelevant variations in image intensity. Motivated by these observations, a number of groups have applied Gabor filters to analyzing textured facial images containing highly specific frequency or orientation characteristics. In this section, the basic properties of the Gabor filters are briefly reviewed, followed by the description of our extracted facial features.

2.1. Gabor Channel Filters

The Gabor filters, or Gaussian wavelets, are complex exponential functions modulated by Gaussian functions. In the particular case of the Gaussian envelope being circularly symmetric, and having the same orientation as the complex sine grating, the expression for a Gabor filter can be defined in the spatial domain by

$$h_{mn}(x,y) = \frac{1}{2\pi\sigma_m^2} e^{-(x^2+y^2)/2\sigma_m^2} \times [e^{i2\pi f_m(x\cos\theta_n + y\sin\theta_n)} - e^{-(2\pi f_m \sigma_m)^2/2}] \quad (1)$$

where m is the index for the scale and n is the index for the orientation. The second term in the square brackets compensates for the DC value. As in [3], the half peak radial and orientation bandwidths are defined by

$$b = \log_2(\frac{\sigma_m f_m + \sqrt{2\ln 2}}{\sigma_m f_m - \sqrt{2\ln 2}}) \qquad \Omega = 2\tan^{-1}(\frac{\sqrt{2\ln 2}}{2\pi\sigma_m f_m}) \quad (2)$$

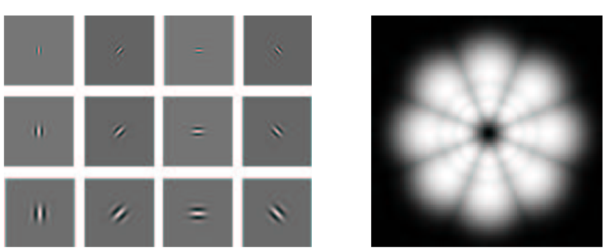
Here b is chosen to be 1 for all filters in the bank and corresponds to a half peak radial bandwidth of one octave. The complete bank of filters used for sampling the joint space-frequency domain is obtained by rotation, with a 45° step, to get 4 orientation channels ($n \in \{0,1,2,3\}$), and by limiting the number of central frequencies to 3 ($f_{m-1} = f_m/\sqrt{2}, m \in \{0,1,2\}$), starting with the Nyquist frequency ($f_2 = \pi/2$). Fig.1 shows the even part of our designed 3×4 Gabor filters and their coverage of the Fourier domain.

2.2. Gabor Feature Representation

Given a facial image $I(x,y)$, we convolve it with a filter $h_{mn}(x,y)$ in the bank, and denote the filtered image by

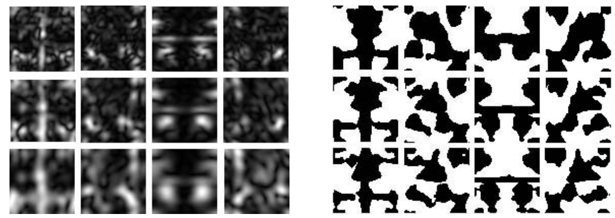
$$W_{mn}(x,y) = I(x,y) * h_{mn}(x,y) \quad (3)$$

The results of an even (cosine-type) and an odd (sine-type) part can be combined in a single magnitude image corresponding to the energy of $W_{mn}(x,y)$. The multi-hierarchical Gabor representations (normalized to zero mean and unit variance) of a human face image $I(x,y)$ are illustrated in Fig.2.(a). Applying the convolution theorem, $W_{mn}(x,y)$ can be computed efficiently via Fast Fourier Transform (FFT).



(a) (b)

Figure 1. (a) Even part of the 3×4 Gabor filters; (b) Coverage of the Fourier domain by their corresponding frequency channels.



(a) (b)

Figure 2. (a) Magnitude of the filtered images for a human face; (b) Feature masks ($\rho = 0.5$) with the selected points in black.

As the dimensionality of the total resulted Gabor feature vectors is very high, the memory and computational requirements for performing subsequent subspace analysis are inevitably large. We therefore adopt a simple "compressing" scheme to determine the dominating regions of different Gabor channels. Given N training images, we perform (3) to obtain 3×4 set of N feature images for each channel. For a particular feature set, its point-wise variance image can be computed as

$$V_{mn} = \{v_{mn}(i,j) | v_{mn}(i,j) = var(W_{mn}(i,j))\} \quad (4)$$

where i, j are image coordinate index and var is the variance operator. The ρ percentage of the points in V_{mn} with the largest value is selected to form a feature mask (Fig.2.(b)). It's clear to see that most key points in the masks appear at the important facial features, thus can be used to generate lower dimensional representative feature vectors O_{mn}^{ρ} by concatenating the masked Gabor filtered image W_{mn}^* .

Fig.2 also illustrates that both the magnitude images and the feature masks are relatively insensitive to the Gabor kernel scale, while vary significantly across different orientations. We therefore divide all Gabor channels to 4 groups according to their respective orientations and conduct 4 augmented feature vectors as follows:

$$X_n^{\rho} = ({O_{1n}^{\rho}}^T {O_{2n}^{\rho}}^T {O_{3n}^{\rho}}^T)^T \quad (5)$$

where T is the transpose operator. These feature vectors thus encompasses important discriminating information from each orientation channel.

3. Null Space-based LDA (NLDA)

The Gabor feature vectors introduced above reside in a high dimensional space even after the compressing stage.

It's necessary for us to further reduce the feature space to a lower dimensional representation.

Linear Discriminate Analysis (LDA), which utilizes the most discriminating features, provides an effective low-dimensional feature space for classification. Its goal is to seek an optimal projection W_{opt}, from the raw data space to a reduced feature space, which maximizes the ratio of the between-class scatter matrix S_b to the within-class scatter matrix S_w, i.e.,

$$W_{opt} = arg \max_{W} \frac{|W^T S_b W|}{|W^T S_w W|} \tag{6}$$

The most widely used method performs Principle Component Analysis (PCA) firstly to make the resulting S_w full-rank, and then standard LDA is used to seek the finial projection. However, it was mentioned in [1] that the optimal discriminate vectors of LDA could be derived from the null space (or kernel in [1]) of S_w. In fact, if a certain vector q belongs to the null space of S_w (i.e.$q^T S_w q = 0$), and also satisfies $q^T S_b q \neq 0$, the radio in (6) will definitely reach the maximum value. In this correspondence, Huang et al [2] introduced an efficient null space approach to solve the small sample size problem. Our recent work [7] gave an in-depth study to this method, which can be described as follows:

- Remove the null space of S_t (the total scatter matrix).

Perform eigen-anlysis on S_t, chose all the eigenvectors corresponding to the nonzero eigenvalues to construct a projection matrix P, and then we get:

$$P^T S_w P = S_w^{'} \qquad P^T S_b P = S_b^{'} \tag{7}$$

- Extract the null space of S_w.

Perform eigen-anlysis on $S_w^{'}$, chose all the eigenvectors corresponding to the zero eigenvalues to construct a projection matrix Y, and then we get:

$$Y^T S_w^{'} Y = 0 \qquad Y^T S_b^{'} Y = S_b^{''} \tag{8}$$

We've proved in [7] that through these two steps, $S_b^{''}$ is full-rank, and there is no need to further diagonalize it lest the selected projections overfit the training samples. Therefore, the overall NLDA projection matrix is $W_{opt} = PY$.

4. Cross-Module Combination

On combining classifiers, it is of great importance to extract independent or uncorrelated feature sets. Empirical results in Section 2.2 illustrated that, the Gabor representation from different channels seems to provide an observer with multiple cues and this in itself facilitates data fusion. Thus we simultaneously apply NLDA on these 4 feature groups as well as the original images, and perform the decision level combination that is fairly appropriate for component classifiers using complementary information.

For a c class problem, let $d_{m,n}$ be the Euclidean distance between the mth class center and the test sample in the nth module, an estimation of posterior probability can be approximated as

$$P(\omega_m|x_n) = \frac{1/d_{m,n}^2}{\sum_{m=1}^{c} 1/d_{m,n}^2} \tag{9}$$

Based on some probability assumptions, [4] gave a theoretical justification to a number of common combination schemes, which we list in Table.1. As mentioned in [4], the combination rule developed under the most restrictive assumptions, the SUM rule, out-performs other schemes, and its sensitivity to estimation errors was also investigated. Our experiments in Section 5 empirically evaluate these mainstream combination rules and obtain results consistent with those analysis in [4].

Table 1. Some combination rules and their formulations.

Rule	Formulation
Product	$h = arg \max_m \prod_n P(\omega_m|x_n)$
Sum	$h = arg \max_m \sum_n P(\omega_m|x_n)$
Max	$h = arg \max_m \max_n P(\omega_m|x_n)$
Min	$h = arg \max_m \min_n P(\omega_m|x_n)$

5. Experiments and Discussion

To demonstrate the effectiveness of our method, extensive experiments were performed on the FERET database [8]. We have selected 70 subjects from this database with 6 up-right, frontal-view images of each subject. The images were selected to bear with moderate differences in illumination, expressions and facial details. Using the manually detected centers of the two eyes, all images were properly rotated, translated and scaled to fit a grid size of 64×64, followed by a histogram equalization step to eliminate lighting effect. Each of these images can then be segmented by means of an predefined mask centered at the middle of the normalized image rectangle, as illustrated in Fig.3.

The first experiment was designed to evaluate the performance of various combination rules and their superiority over any NLDA based single decision making scheme, i.e. using information from only one Gabor channel or the original images. We randomly divided the database into two parts: 3 images of each person were selected as training samples, while the remaining 3 as testing images. The L1 norm similarity measure and the Nearest Neighbor rule (for non-combination method only) were adopted to classify each probe input. We repeated the experiments for all

Figure 3. Normalized subjects from the FERET database

possible combination of the training and testing sets, and obtained the following averaged results. (Table.2)

Table 2. Recognition rate (%) for different combination rules and different single decision making schemes.

	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$	$\rho = 1.00$
Product	95.79	96.67	96.79	96.69
Sum	96.10	96.78	96.86	96.79
Max	94.43	95.21	95.33	95.41
Min	83.88	89.52	90.81	91.02
Original image	92.26			
Channel 1 (0 °)	52.93	69.50	72.55	75.00
Channel 2 (45 °)	78.19	86.79	89.07	89.71
Channel 3 (90 °)	84.55	89.81	90.86	91.21
Channel 4 (135 °)	82.98	88.29	90.21	90.38

It was noted that any single Gabor channel, capturing only part of the overall facial texture information, did not provide superior performance to the original intensity image. But as different channels potentially offered complementary representation about the face pattern, the integration of classifier outputs from them and the original image yielded significantly improved performance. We also found that the MIN and MAX rules, which rely on order statistics, were less robust in our particular circumstance where component classifiers might bear high degree of overlap of the distribution of the posterior probability estimation.

Our next experiment compared some appearance based methods that have become quite popular in the literature, namely LDA/NLDA (based on all Garbor feature or intensity feature), Liu et al [6] with alternative methods developed in this paper. We chose the compression factor ρ to be 0.25 and adopt SUM rule for classifier combination. Test process was almost the same with the first experiment, except that we let the number of training samples per class k increase from 2 to 5, while keeping an empty intersection with the testing set. The averaged results (Table.3) from this experiment led to such intuitive conclusions:

- In respect of small size problem, Null Space-based LDA is more suitable for dimensionality reduction.
- Gabor feature, which captures salient visual properties such as spatial localization, orientation selectivity, is superior to intensity representation for classification.
- Hybrid system based on combination schemes consistently outperforms a single best classifier.

Table 3. Recognition rate (%) for different methods.

	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Intensity+LDA	72.62	84.21	91.62	95.24
Gabor+LDA	77.00	86.02	91.90	95.71
Intensity+NLDA	84.02	92.26	95.29	97.86
Gabor+NLDA	85.33	93.76	97.43	99.52
Liu et al [6]	81.14	89.71	93.91	97.38
Our method	90.74	96.10	98.14	99.52

6. Conclusion

In this paper, a novel face recognition strategy combining various Null Space-based Gabor features is proposed and evaluated on the FERET database. Our excellent performance mainly relies on the discriminating feature selection, improved dimensionality reduction and the final combination strategy. Liu et al [6] claimed that EFM outperforms LDA, but in our experiment it was numerically unstable and only gave limited improvement. Its regular subsampling scheme may lose important discriminant information in the face region. As more Gabor channels were selected and applied on larger size images, high accuracy was also reported in [6]. However, in our experiment no single decision making schemes can exceed the performance achieved by the combination system.

Acknowledgements: Portions of the research in this paper use the FERET database of facial images collected under the FERET program. This work is funded by the research grants from the Natural Sciences Foundation of China (No.60332010, 69825105, 60121302, 60335010), the National 863 Program of China (Grant No. 2001AA11480), and the Chinese Academy of Sciences.

References

[1] P.N.Belhumeur, J.P.Hespanha, and D.J.Kriegman. "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7 , pp. 711-720, 1997.

[2] R.Huang, Q.S.Liu, H.Q.Lu, and S.D.Ma. "Solving the Small Sample Size Problem of LDA", *Proc. of the 16th International Conference on Pattern Recognition*, vol. 3, pp. 29-32, Quebec, Canada, August, 2002.

[3] A.K.Jain, and F.Farrokhnia. "Unsupervised Texture Segmentation Using Gabor Filters", *Pattern Recognition*, vol. 24, pp. 1167-1186, 1991.

[4] J.Kittler, M.Hatef, R.Duin, and J.Matas "On Combining classifiers", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239, 1998.

[5] M.Lades, J.C.Vorbruggen, J.Buhmann, J.Lange, C. von der Malsburg, P.R.Wurtz, and W.Konen. "Distortion Invariant Object Recognition in the Dynamic Link Architecture", *IEEE Trans. Comput*, vol. 42, pp. 300-311, 1993.

[6] C.Liu, and H.Wechsler. "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition", *IEEE Trans. Image Processing*, vol. 11, pp. 467-476, 2002.

[7] W.Liu, S.Z.Li, Y.H.Wang, and T.N.Tan. "Null Space-based Kernel Fisher Discriminant Analysis for Face Recognition", *To appear in the 6th International Conference on Automatic Face and Gesture Recognition*, Seoul, Korea, May, 2004.

[8] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss. "The FERET database and evaluation procedure for face recognition algorithms", *Image and Vision Computing J*, Vol. 16, No. 5, pp 295-306, 1998.

[9] L.Wiskott, J.M.Fellous, N.Kruger, and C. von der Malsburg. "Face Recognition by Elastic Bunch Graph Matching", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775-792, July, 1997.

[10] W.Zhao, R.Chellappa, A.Rosenfeld, and P.J.Philips. "Face Recognition: A Literature Survey", *CVL Technical Report, University of Maryland*, Oct, 2002.