Multiple Fisher Classifiers Combination for Face Recognition based on Grouping AdaBoosted Gabor Features

Yu Su[1] Shiguang Shan[2] Bo Cao[2] Xilin Chen[1] Wen Gao[1,2]
[1]School of Computer Science and Technology,
Harbin Institute of Technology, Harbin, 150001, China
[2]Institute of Computing Technology,
Chinese Academy of Sciences, Beijing, 100080, China
{ysu, sgshan, bcao, xlchen, wgao}@jdl.ac.cn

Abstract

Gabor features has been recognized as one of the most successful representation methods, such as Elastic Graph Matching, Gabor Fisher Classifier, and AdaBoost Gabor Fisher Classifier. One of the key issues in using Gabor features is how to efficiently reduce its high dimensionality. This paper proposes a multiple Fisher classifiers combination approach based on re-grouping Gabor features selected by using re-sampling and AdaBoost. At least two advantages can be observed with the proposed method: (1) more discriminative Gabor features are exploited in distributed Fisher classifiers by re-grouping the selected Gabor features; (2) combination of multiple Fisher classifiers improves the final performance of the classification compared with the traditional Fisher classifier. Our extensive experiments on two large face databases, FERET and CAS-PEAL, have impressively shown the effectiveness of the proposed method.

1. Introduction

In recent years, due to wide potential applications in security, finance, law enforcement and military, such as mug-shot database matching, identity authentication for credit card or driver license, access control and video surveillance, there has been much interest in automatically recognizing faces in still and video images. Related research activities have significantly increased over the past few years [1].

Many methods have been proposed for face recognition such as Eigenface [2], Fisherface [3], Bayesian inference [4], and Elastic Bunch Graph Matching (EBGM) [5]. Especially, in recent years, face descriptors based on Gabor filtering have been recognized as one of the most successful representation methods, such as Elastic Graph Matching [5], Gabor Fisher Classifier [6], and AdaBoost Gabor Fisher Classifier [7,8]. In EBGM, Gabor wavelets were firstly exploited to model faces based on the multi-resolution and multi-orientation local features. Until now, face representation based on Gabor features have achieved great success in face recognition area for the variety of advantages of the Gabor filters. In [6], Liu proposed Gabor-Fisher Classifier

BMVC 2005 doi:10.5244/C.19.60

method and achieved good performance. In this method, fisher linear discriminant analysis was applied on the Gabor features extracted from the grid vertices of the face images. However, Gabor features are too high dimensional and need to be reduced. For instance, in GFC, the high dimensional Gabor features were down-sampled and further reduced dimension by using PCA. In [7], Yang proposed a feature selection method based on AdaBoost to select the most discriminating Gabor features as a more compact face representation for following classification, e.g., Fisher Discriminant Analysis (FDA) [8]. Nevertheless, in this method, in order to preserve enough information for classification, the dimension of selected Gabor features is still as high as several thousands. From the other hand, the maximum dimension of FDA subspace is not larger than the number of subjects in the training set, which is generally only several hundreds. Therefore, only hundreds of dimension can be modeled in the FDA subspace, which means plenty of discriminative information has to be discarded as the dimension of FDA space is far less than the number of features used for training.

This paper proposed a method based on multiple FDAs to increase the total dimension of FDA subspaces by grouping the AdaBoosted Gabor features. Our system of training multiple FDA classifiers comprises three stages. Firstly, we use AdaBoost to select thousands of most discriminating Gabor features as in [7]. Face recognition is a multi-class problem; therefore, in order to using AdaBoost, we use the concept of the intra-personal and extra-personal difference [4] to convert the multi-class problem to a two-class problem. Then we divide the training set consists of intra-personal and extra-personal differences into several subsets by random sampling. On each subset, we can obtain a subset of Gabor feature selected by AdaBoost. Secondly, we re-group these features to form larger subsets. Finally, multiple classifiers are obtained by applying FDA on each feature subset. In the testing process, a face image is classified by each of the FDA classifier. Extensive experiments on FERET [9] and CAS-PEAL [10] databases have shown that the proposed method achieves better performance compared with using the single FDA method.

The remaining part of the paper is organized as follows: in section 2, FDA based on AdaBoosted Gabor features is introduced. In section 3, we present the construction of multiple FDA classifiers. Experiments and analysis are conducted in section 4, followed by discussion and conclusion in section 5.

2. FDA based on AdaBoosted Gabor Features

2.1 Gabor Feature Extraction

Gabor filter can capture salient visual properties such as spatial localization, orientation selectivity, and spatial frequency characteristics. We choose Gabor features to represent the face image considering these excellent capacities. Gabor filters are defined as follows:

$$\psi_{u,v}(z) = \frac{\| k_{u,v} \|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2 \|z\|^2 / 2\sigma^2)} \left[e^{i\vec{k}_{u,v} z} - e^{-\sigma^2/2} \right] \tag{1}$$

where u and v define the orientation and scale of the Gabor filters. Wave vector is defined as follows:

$$k_{u,v} = k_v e^{i\phi_u} \tag{2}$$

where $k_v = \frac{k_{\max}}{f^v}$ and $\phi_u = \frac{u\pi}{8}, \phi_u \in [0, \pi)$.

In our experiments we use the Gabor filters with the following parameters: five scales $v \in \{0,1,2,3,4\}$ and eight orientations $u \in \{0,1,2,3,4,5,6,7\}$ with $\sigma = 2\pi$, $k_{\max} = \pi/2$, and $f = \sqrt{2}$. The same parameters are also taken in [5].

2.2 Feature Selection by AdaBoost

Convolving the image with these 40 Gabor kernels can then generate the Gabor features. Thus, for each pixel position in the face image, 40 complex values can be calculated. Note that, because the phase information of the transform is time-varying, generally, only its magnitudes are used to form the final face representation. Evidently, this will result in a feature with a dimension of 40 times of the original face images size, which is too high dimensional in pixel-wise dense sampling case. Therefore, it is necessary to reduce the dimension of the original Gabor features. In this paper, learning method based on AdaBoost is exploited to select the most discriminating Gabor features.

Face recognition is a multi-class problem; therefore, in order to use AdaBoost, we adopt the concept of Intra-personal difference and Extra-personal difference as in [4] to convert the multi-class problem to a two-class problem. As we know, boosting learning is a strong tool to solve the two-class classification problem and it is exploited in our method to distinguish the intra-personal differences (hereinafter called positive samples) from the extra-personal differences (hereinafter called negative samples).

We use the AdaBoost learning algorithm presented in [11], to select a small set of Gabor features (in this variation of AdaBoost, each feature formulates a weak classifier) from the original Gabor feature space of extremely high dimension. The framework of the selecting process is illustrated in Figure 1.

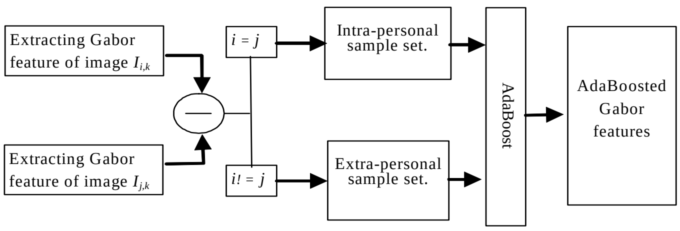

Figure 1: The process of selecting the most discriminating Gabor features

---

Given labeled examples Set S and their initial weights ω_1

Do for *t*=1, … , T:

1. Normalize the weight ω_t
2. For each feature, *k*, train a classifier h_k with respect to the weighted samples;
3. Calculate the classification error rate for each h_k, choose the classifier h_t with the least error rate, get α_t, the weight of h_t.
4. Update the weight of each sample: ω_{t+1},

Get the strong classifier $S(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$

Table1: Feature Selection by AdaBoost

We can obtain a stronger classifier by AdaBoost, which linearly combines a number of weak classifiers, each of which is constructed by using a single Gabor feature. Therefore, AdaBoost can be considered as a feature selection algorithm. The brief process of AdaBoost is described in Table1, in which *T* Gabor features corresponding to the *T* weak classifiers are selected as the most discriminating features.

2.3 AdaBoosted Gabor Fisher Classifier

To classify faces using the AdaBoosted Gabor features, researchers generally explore Fisher Discriminant Analysis (FDA) [3, 8], which tries to seek for an optimal transform W_{opt} by maximizing the following ratio criterion:

$$J_{FLD}(W_{opt}) = \arg\max_{W} \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|} \tag{3}$$

where S_B is the between-class scatter matrix, and S_W is the within-class scatter matrix. Thus, by applying this method, we find the projection directions that on one hand maximize the Euclidean distance between the face images of different classes, and on the other hand, minimize the distance between the face images of the same classes. The ratio is maximized when the column vectors of the projection matrix *W* are the eigenvectors of $S_W^{-1}S_B$ whose rank is at most equal to the number of persons in the training set. So, the number of eigenvectors of $S_W^{-1}S_B$ is at most equal to the number of persons in the training set.

In order to preserve enough information for classification, the dimension of selected Gabor features using AdaBoost is still as high as several thousands. From the other hand, the maximum dimension of FDA subspace is not larger than the number of subjects in the training set, which are generally only several hundreds. Therefore, generally PCA is conducted to further reduce the dimensionality of the selected Gabor features to less than the number of subjects in the training set. This means that only hundreds of dimension can be modeled in the final FDA subspace, which implies plenty of discriminative information may have to be discarded. To better exploit the selected discriminating Gabor features, we propose to combine multiple Fisher classifiers by grouping these selected features.

3. Multiple Fisher Classifiers Combination

A straightforward way to avoid the PCA dimensionality reduction is grouping the selected high-dimensional Gabor features and providing each FDA classifier limited number of Gabor features, that is, less than the number of subjects in the training set. Thus, multiple Fisher classifiers are constructed, which is then further combined to obtain the final classification.

3.1 Random Re-sampling of Extra-personal Differences

Given a training set that includes *N* images for each of the *K* individuals, the total number of image pairs is $\binom{KN}{2}$. A small minority, $K\binom{N}{2}$, of these pairs are from the same individual. In order to deal with the grossly imbalance of the positive samples and negative samples, we have to divide the negative samples into a number of smaller subsets.

A simple and effective method to divide negative sample is random sampling. By random sampling, negative samples can be divided into several random subsets, each of which can be united with all positive samples to form a training set for AdaBoost. Suppose that we divide the negative samples into M random subsets, then we have M training sets. So, M feature subsets can be selected from the original Gabor feature set by apply AdaBoost on each training set. The process was shown in Figure 2.

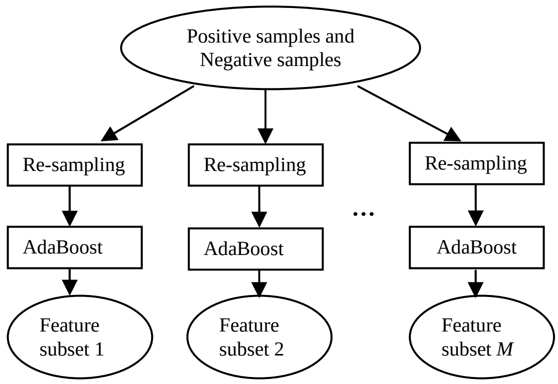

Figure 2: Feature selection by Re-sampling and AdaBoost

3.2 Regrouping Feature Subsets

By applying FDA to each feature subsets, we can obtain M classifiers which constitute a multiple classifiers system. In addition to this simple method, we can re-organize the feature subsets by combine some of them to form a larger one as we can see in Figure 3. For convenience, we call the feature subsets selected by AdaBoost the original feature subsets, and the feature subsets consist of some original feature subsets the re-organized feature subsets. For example, in Figure 2, 'Feature subset *i*' is the original feature subset. In Figure 3, f_i represents the original feature subset and F_i represents the re-organized feature subset. Each re-organized feature subset contains the same number of original feature subsets in order that the number of features in each re-organized feature subset is approximate equal. Suppose that the number of features in each re-organized feature subsets is n, the number of original feature subsets in each re-organized feature subset is N, the number of the re-organized feature subsets is L and the number of person in training set is c, then the maximum dimensions of FDA subspace D can be formulated in equation:

$$D = L * \min(n, c-1) \tag{4}$$

where $L * n$ is equal to the number of all features selected by AdaBoost which is a fixed value. Obviously, D becomes larger as n decrease when n is larger than c-1, and the maximum value of D is about $L * n$. In the following experiments, we can see that the system performance get better as the dimensions of FDA subspace increase. But when $n<c-1$, D does not change as n decrease and its value is fixedly at about $L * n$.

Our purpose for designing multiple classifiers is to increase the dimension of FDA subspace and consequently save more information for classification. However, the phenomenon of over-fitting may appear in our system as in most of the pattern recognition system. According to the theory of Occam's razor, one should not use classifier that is more complicated than necessary. The complexity of the multiple classifiers system is decided by the number of classifiers and the complexity of each classifier. In our system, all the classifiers are formed by FDA, so their complexities are the same. Then, the number of classifiers is the key to the complexity of our system. So, we should find an optimum size of re-organized feature sets in order to keep proper balance between the dimension of FDA subspace and number of classifiers. As the following experiment will show, the optimum size of re-organized feature sets is approximately equal to *c*.

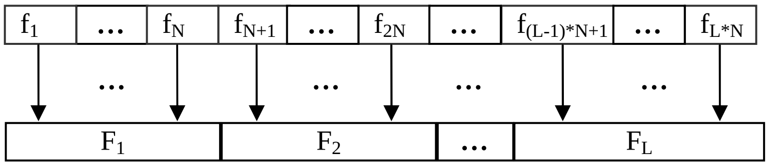

Figure 3: Re-grouping feature subsets

3.3 Combining Multiple FDA Classifiers

By applying FDA to each re-organized feature subsets, we can obtain *L* classifiers which we can combine further to form a stronger classifier. There are many methods of combining multiple classifiers as in [12]. In this paper, we use the most typical sum rule to combine multiple FDA classifiers. More complex combination algorithms may further improve the system performance.

In the process of face recognition, the classifier is essentially required to output the similarity of any two face images. In [3], faces images are firstly projected to the FDA subspace, then the similarity of two faces is defined as the Euclidean distance of its projection vector in FDA subspace. In our system, for any two faces we can obtain *L* similarities given by *L* FDA classifiers. The average of these *L* similarities is defined as the final similarity of the corresponding pair of faces. Then, the nearest neighbor classifier is used for classification. Figure 4 shows how we obtain the similarity of two faces.

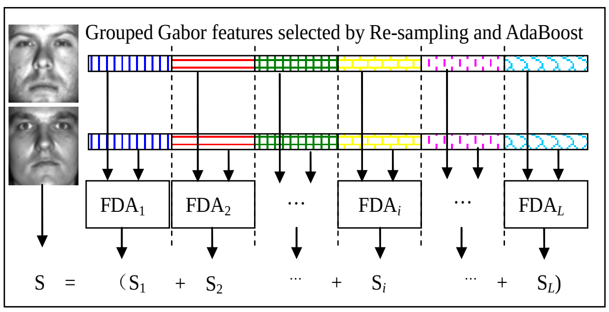

Figure 4: Similarity Computation based on Multiple FDA Classifiers Combination

4. Experiments

We tested the proposed method on two large face databases: FERET [9] and

CAS-PEAL [10]. The training set of FERET includes 1002 images of 429 persons and the training set of CAS-PEAL includes 1200 images of 300 persons. All images are cropped to the size of 64 by 64 and rectified according to the manually located eye positions. Therefore, the number of original Gabor features of each images is $64\times64\times5\times8=163840$, from which we select thousands of the most discriminant ones.

In the following part, we compare the system performances in the condition of different sizes of re-organized feature subsets and different numbers of selected features.

4.1 Sizes of Re-grouped Feature Subsets

In FERET, we run re-sampling and AdaBoost for 18 times from which we get 18 original feature subsets. The number of all the features in the 18 original feature subsets is 991 (about 50 features for each feature subset). We then re-organize the 18 original feature subsets to form some larger re-grouped feature subsets. The relationship of the number of re-organized feature subsets L, the approximate size of each re-organized feature subset n and the Maximum dimension of FDA subspace D is given in Table 2 (refer to Equ.4).

L	n	D
1	991	428
2	450	856
3	330	990
6	165	990
9	110	990
18	55	990

Table 2: The relationship of L, n and D in FERET (c = 429)

As we illuminated in section 3.2, when $n > c$-1, the maximum dimension of FDA subspace D becomes larger as n decrease; when $n < c$-1, D is a fixedly at about $L * n$. Obviously, in order to use more dimension of FDA subspace and less classifiers, we adopt the division which contains 3 classifiers (L=3). The same conclusion is drawn by the experiment that the division which contains 3 classifiers gives the best performance as shown in Figure 5.

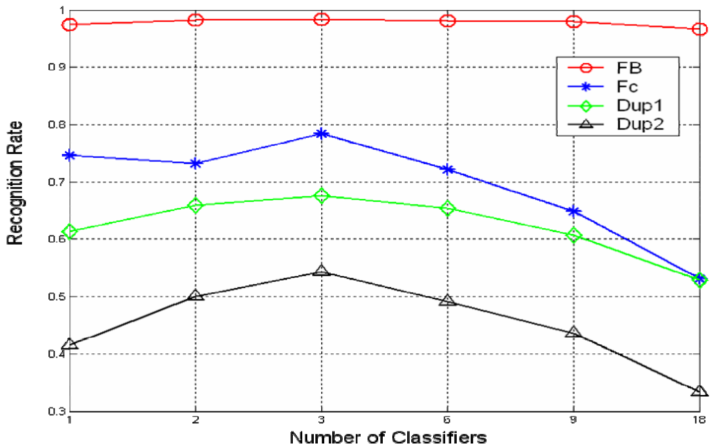

Figure 5: The comparison of different number of classifiers in FERET

Similar experiments are conducted on CAS-PEAL face database and we also get 18 original feature subsets. The number of all the features in the 18 original feature subsets is 2667 and each original feature subset has about 150 features. The relationship of *L*, *n* and *D* is given in Table 3. According to the above analysis, we adopt the division which contains 9 classifiers (*L*=9). As shown in Figure 6, the best performance is given by the division which contains 9 classifiers.

L	*n*	*D*
1	2667	299
2	1330	598
3	880	897
6	440	1794
9	290	2610
18	145	2610

Table 3: The relationship of *L*, *n* and *D* in FERET (c = 300)

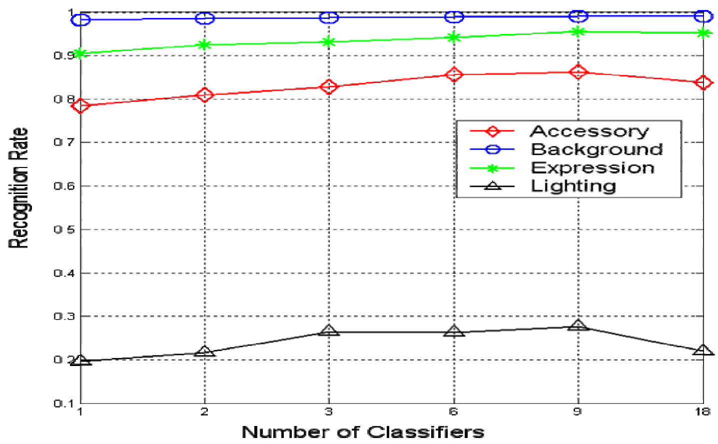

Figure 6: The comparison of different number of classifiers in CAS-PEAL

By the experiments in FERET and CAS-PEAL, we have validated our standpoint in section 3.2 that the optimum size of re-organized feature subset is approximately equal to *c*.

4.2 Number of total features

In figure 5 and figure 6, we can see that the best performance given by multiple classifiers is much better than the performance given by one classifier using all the features (i.e., *L*=1). Meanwhile, we find the optimum size of re-organized feature subset which is fixed in the following experiments. Figure 7 and Figure 8 compare the performance of using multiple classifiers and one classifier in the condition of different numbers of features in both FERET and CAS-PEAL. Obviously, the performances of multiple classifiers system based on different feature set are better than single classifier based on all of the features. From Figure 7 and Figure 8, we can also see that the system performance in most tests set get better as the number of features increases. So, the number of Gabor features in our system is more than the number of persons in training set in order to achieve better performance.

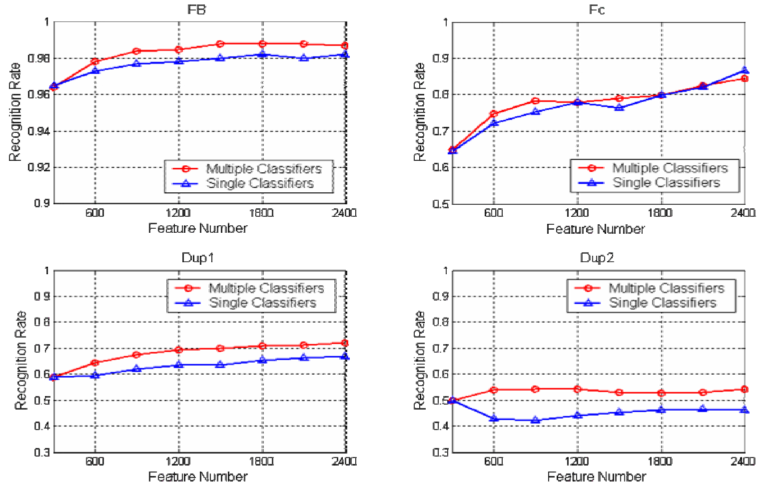

Figure 7: The comparison of Multiple Classifiers System and Single Classifier System in FERET

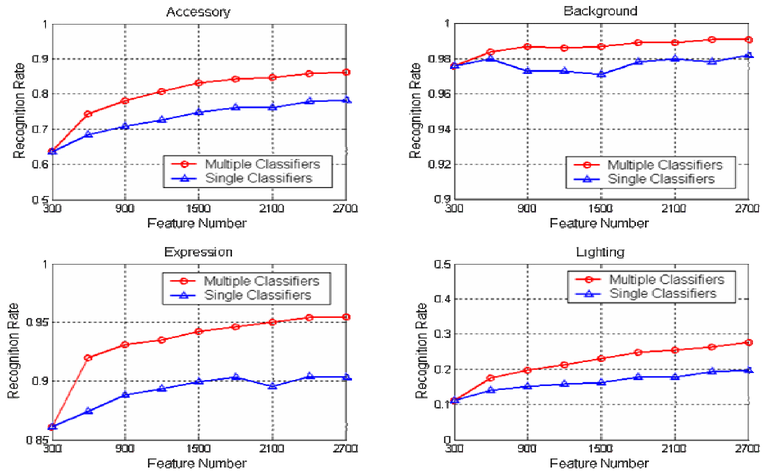

Figure 8: The comparison of Multiple Classifiers System and Single Classifier System in CAS-PEAL

5. Conclusion

Gabor features has been recognized as one of the most successful representation methods. However, one of the key issues in using Gabor features is how to efficiently reduce its high dimensionality: Gabor-based representation is too high dimensional even after selected by some feature selection methods, compared with the number of persons c in the training set. When we apply Fisher Discriminant Analysis on Gabor features, we argue that some discriminative information may have to be discarded as the maximum dimension of FDA space (c-1) is, far less than the number of Gabor features.

In order to increase the total dimension of FDA subspace, we propose to re-group the AdaBoosted Gabor features into some smaller feature subsets. Then multiple Fisher classifiers can be obtained by training one FDA on each of the smaller feature subset, thus the total dimension of FDA subspace can be more than *c*-1. The number of features *n* used in each Fisher classifier is the main parameter in our approach. We have theoretically and empirically shown that the optimum value of *n* is approximately equal to *c*, which seems a balance point between the generalizability and the complexity. Our extensive experiments on two large face databases, FERET and CAS-PEAL, have impressively shown the effectiveness of the proposed method compared with the traditional Fisher classifier.

Acknowledge

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, "100 Talents Program" of CAS, ShangHai Municipal Sciences and Technology Committee (No.03DZ15013), and ISVISION Technologies Co., Ltd.

Reference

[1] R.Chellappa, C.L.Wilson, S.Sirohey, Human and Machine Recognition of faces: A survey, Proceedings of the IEEE, vol.83, no.5, 1995

[2] M. Turk, and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, vol.3, no. 1, pp. 71-86, 1991.

[3] P.N.Belhumeur, J.P.Hespanha etc. Eigenfaces vs Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence* vol.20, no.7, pp.711-720, 1997

[4] Moghaddam, Pentland. Beyond Eigenfaces: Probabilistic Matching for Face Recognition. *IEEE International Conference on Automatic Face and Gesture Recognition, AFGR'98*, pp. 30-35, 1998.

[5] L. Wiskott, J. Fellous, N. Kruger and C. von der Malsburg, Face recognition by elastic bunch graph matching", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-669, 1997.

[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, no. 4, pp. 467-476, 2002

[7] P.Yang, S. Shan, W.Gao, S.Z.Li, Face recognition using Ada-Boosted Gabor features, *IEEE international conference on Automatic Face and Gesture Recognition, AFGR'04*, pp.356–361, 2004.

[8] S.Shan, Study on some key issues in face recognition, Ph.D thesis, CAS, 2004

[9] PJ Phillips, H. Moon, P. Rauss, and SA Rizvi. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE international conference on Computer Vision and Pattern Recognition, CVPR'97*, pp.137-143, 1997.

[10] B.Cao, S.Shan, etc, Baseline Evaluations On The CAS-PEAL-R1 Face Database, Lecture Notes in Computer Science (LNCS3338), Advances in Biometric Person Authentication, pp370-378, 2004.12

[11] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *IEEE international conference on Computer Vision and Pattern Recognition, CVPR'01*, vol 1, pp. 511–518, 2001.

[12] J. Kittler and F. Roli. (Eds): Multiple Classifier Systems. Springer, 2004.