Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FACE RECOGNITION BASED ON OPTIMAL KERNEL MINIMAX PROBABILITY MACHINE

[1]ZHIQIANG ZHOU, [2]ZIQIANG WANG, [3]XIA SUN
[1]School of Information Science and Engineering, Henan University of Technology, Zhengzhou China
E-mail: china126box@126.com

ABSTRACT

Face recognition has received extensive attention due to its potential applications in many fields. To effectively deal with this problem, a novel face recognition algorithm is proposed by using the optimal kernel minimax probability machine. The key idea of the algorithm is as follows: First, the discriminative facial features are extracted with local fisher discriminant analysis (LFDA). Then, the minimax probability machine (MPM) is extended to its nonlinear counterpart by using optimal data-adaptive kernel function. Finally, the face image is recognized by using the optimal kernel MPM classifier in the discriminative feature space. Experimental results on three face databases show that the proposed algorithm performs much better than traditional face recognition algorithms.

Keywords: *Face Recognition, Minimax Probability Machine, Feature Extraction, Kernel Function*

1. INTRODUCTION

In recent years, face recognition has been receiving more and more attention in pattern recognition and computer vision fields. The motivation behind face recognition is to employ it to implement video surveillance, identity authentication, and human-computer interaction. As a result, a large number of face recognition algorithms have been proposed, and surveys in this area can be found in [1].Two issues are central to all these algorithms: how to extract discriminative facial features and how to classify a new face image based on the extracted facial features. Therefore, this work also focuses on the two issues of feature extraction and classifier selection.

Principal component analysis (PCA) and linear discriminant analysis (LDA) are two classic feature extraction algorithms for face recognition [2]. The major idea of PCA is to decompose a data space into a linear combination of a small collection of bases, which are pairwise orthogonal and capture the directions of maximum variance in the training set. As an unsupervised feature extraction algorithm, PCA is optimal in terms of representation and reconstruction, but not for discriminating one face class from others. LDA is a supervised feature extraction algorithm which aims to find a projection subspace on which the data from the same class will be pushed close while the data from different classes will be pulled far away. The projection vectors are commonly obtained by maximizing the between-class scatter and simultaneously minimizing the within-class scatter. Due to the utilization of class label information, LDA is experimentally reported to outperform PCA for face recognition when sufficient labeled face images are provided [3]. However, the performances of LDA are often degraded by the limited available dimensional space and the singularity problem. In addition, independent component analysis (ICA) is another linear feature extraction algorithm [4], which separates the high-order moments of the input data besides the second-order moments in PCA. However, the objective of ICA is to make the components of projected vectors as independent as possible, which may not necessarily be the best for classification problem such as face recognition. Although PCA and LDA have widely been applied to image retrieval, face recognition, and information retrieval, they may fail to discover the underlying manifold structure as they seek only a compact Euclidean subspace for face representation and recognition [5]. Following the above analysis, it is desired to propose an efficient algorithm for feature extraction by explicitly considering the possibly local manifold structure of face image space. Local fisher discriminant analysis (LFDA) [6] is a recently proposed manifold learning algorithm which aims to maximize between-class separability and preserve within-class local manifold structure at the same time. Thus LFDA is helpful for feature extraction of face image data.

Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

As for face recognition, classifier selection is another key issue after feature extraction. At present, the nearest neighbor (NN) classifier is widely used in the face recognition algorithm, which works by finding the neighbor with the minimum distance between the query instance and all labeled data instances. Although the NN classifier is the simplest one for pattern classification, its performance deteriorates dramatically when the input data set has a relatively low local relevance. Support vector machine (SVM) [7] is a popular pattern classification method used in recent years. It obtains top-level performance in different applications because of its good generalization ability in minimizing the VC dimension and achieving a minimal structural risk. The basic idea behind SVM is to find an optimal hyperplane in a high dimensional feature space that maximizes the margin of separation between the closest training examples from different classes. Although SVM classifier has achieved great success in many pattern classification tasks, one major drawback of SVM is that it can not obtain an explicit upper bound on the probability of misclassification of future data. Recently, minimax probability machine (MPM) [8, 9] classifier has been of wide concern since it can obtain an explicit worst-case bound on the probability of misclassification of future data. However, MPM fails to consider how to select an optimal kernel function that adapts well to the input data and the learning. In this paper, we propose an optimal adaptive kernel function to maximize the class separability in the kernel feature space. The final optimized kernel MPM shows that it is more adaptive to the face image data and leads to a substantial improvements in the performance of face recognition.

The rest of the paper is organized as follows. In section 2, we introduce how to extract discriminative facial feature with LFDA. An effective face recognition algorithm based on optimal kernel MPM is proposed in Section 3. Experimental results are shown in Section 4. Conclusions are reported in Section 5.

2. DISCRIMINATIVE FACIAL FEATURE EXTRACTION WITH LFDA

Local fisher discriminant analysis (LFDA) [6] is a recently proposed manifold learning algorithm for discriminative feature extraction, which combines the advantages of LDA and locality preserving projection. It selects features through maximizing between-class separability and preserving the within-class local manifold structure at the same time, thus achieving maximum discrimination.

Given a set of face images $x_1, x_2, \ldots, x_n \in \mathbb{R}^D$, Let $X = [x_1, x_2, \ldots, x_n]$. Let S^w and S^b denote the local within-class scatter matrix and the local between-class scatter matrix, respectively. Their definitions are as follows:

$$S^w = \frac{1}{2}\sum_{i,j=1}^{n} W_{ij}^w \left(x_i - x_j\right)\left(x_i - x_j\right)^T \quad (1)$$

$$S^b = \frac{1}{2}\sum_{i,j=1}^{n} W_{ij}^b \left(x_i - x_j\right)\left(x_i - x_j\right)^T \quad (2)$$

Where W^w and W^b are two local weight matrices defined on the data points, their components can be obtained through the following computation:

$$W_{ij}^w = \begin{cases} A_{ij} / n_l, & \text{if } c_i = c_j = l \\ 0, & \text{if } c_i \neq c_j \end{cases} \quad (3)$$

$$W_{ij}^b = \begin{cases} A_{ij}\left(1/n - 1/n_l\right), & \text{if } c_i = c_j = l \\ 1/n, & \text{if } c_i \neq c_j \end{cases} \quad (4)$$

Where c_i denotes the class label of image x_i, $l \in \{1, 2, \cdots, c\}$ represent the class label of image, c denotes the total number of class label, and n_l denotes the number of images in the lth class. A_{ij} denotes the following affinity matrix:

$$A_{ij} = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{\sigma_i \sigma_j}\right) \quad (5)$$

where σ_i represents the local scaling of the images around x_i, which is determined by

$$\sigma_i = \left\|x_i - x_i^k\right\| \quad (6)$$

where x_i^k is the k-nearest neighbor of image x_i, and the parameter k is empirically set to 7 in the following experiments.

LFDA aims to find an optimal transformation matrix $U \in \mathbb{R}^{D \times d}$ by solving the following maximization problem:

$$J(U) = \arg\max_U \frac{Tr\left(U^T S^b U\right)}{Tr\left(U^T S^w U\right)} \quad (7)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

As can be seen from the above optimal objective function, LFDA looks for an optimal transformation matrix such that nearby data pairs in the same class are made close and the data pairs in different classes are separated from each other; far apart data pairs in the same class are not imposed to be close.

Thus, the optimal U are the eigenvectors associated with the largest eigenvalues of the following generalized eigen-problem:

$$S^b U = \lambda S^w U \tag{8}$$

Since S^w is nonsingular after some preprocessing (such as PCA projection) steps on X, the column vectors of U can also be regarded as the eigenvectors of the matrix $\left(S^w\right)^{-1} S^b$ associated with the largest eigenvalues. Let the column vectors $U_1, U_2, \cdots, U_d$ be the solution of (8) ordered according to their eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_d$. Thus, the discriminative feature y of image x can be computed as follows:

$$\begin{aligned} x \to y &= U^T x \\ U &= \left(U_1, U_2, \cdots, U_d\right) \end{aligned} \tag{9}$$

where y is the lower-dimensional discriminative feature representation of the face image x, and U is the transformation matrix.

Now, we get the lower-dimensional feature representation of the original face images. In the reduced semantic space, those face images belonging to the same class are close to one another and belonging to different classes are far away each other. Therefore, we can apply effective classifier algorithm to implement face recognition in the reduced feature space. In the next section, we will introduce how to classify different face images with optimal kernel MPM classifier.

3. FACE RECOGNITION WITH OPTIMAL KERNEL MPM CLASSIFIER

Minimax probability machine (MPM) [8, 9] classifier is a recently proposed classifier algorithm. The main advantage of MPM is that it can minimize the worst-case probability of misclassification of future data points under all possible choices of class-conditional densities with given mean and covariance matrix. In addition, it can cope with the nonlinear decision boundaries by exploiting kernel trick. In the following, we first introduce MPM and its kernel extension, and then we discuss how to select an optimal kernel function that adapts well to the input face image data and the face recognition task, so that the recognition accuracy can be guaranteed.

Let x and y denote two facial feature vectors in the reduced feature space, and their means vectors and covariance matrices are represented as $\{\bar{x}, \Sigma_x\}$ and $\{\bar{y}, \Sigma_y\}$, respectively. MPM aims to find the hyperplane $a^T z = b$ ($a \in \mathbb{R}^m \setminus \{0\}$, $b \in \mathbb{R}$ and $z \in \mathbb{R}^n$) which can separate the two classes of points with maximal probability with respect to their means and covariance matrices. It can be formally described as follows:

$$\begin{aligned} \max_{\alpha,a,b} \alpha \quad s.t. \quad & 1-\alpha \geq \sup \Pr\left\{a^T x \leq b\right\} \\ & 1-\alpha \geq \sup \Pr\left\{a^T y \geq b\right\} \end{aligned} \tag{10}$$

where α represents the lower bound of the accuracy for the classification of future data points.

By using the following theorem introduced in [8]:

$$\begin{aligned} &\sup \Pr\left\{a^T y \geq b\right\} = \frac{1}{1+d^2} \\ &\text{with} \quad d^2 = \inf_{a^T y \geq b} \left(y-\bar{y}\right)^T \Sigma_y^{-1} \left(y-\bar{y}\right) \end{aligned} \tag{11}$$

The problem (10) can be equivalently transformed into the following problem:

$$\min_a \sqrt{a^T \Sigma_x a} + \sqrt{a^T \Sigma_y a} \quad \text{s.t.} \quad a^T\left(\bar{x}-\bar{y}\right) = 1 \tag{12}$$

This is a second order cone program (SOCP) problem, which can be solved by using interior methods [10]. Once the optimal a_* and b_* are obtained by solving (12) and (10), then classification of new data point z is done by computing:

$$\operatorname{sgn}\left(a_*^T z - b_*\right) = \begin{cases} +1, & z \text{ belongs to class } x \\ otherwise, & z \text{ belongs to class } y \end{cases} \tag{13}$$

The MPM algorithm described above is a linear method, which may fail to deal with highly nonlinear data. To extend MPM to the nonlinear case, we discuss how to perform MPM in Reproducing Kernel Hilbert Space (RKHS) [7], which gives rise to kernel MPM.

Let face image data x and y be mapped into kernel space via nonlinear mapping function $\varphi(\ )$:

$$x \to \varphi(x) \sim \left(\overline{\varphi(x)}, \sum_{\varphi(x)}\right) \tag{14}$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$y \to \varphi(y) \sim \left(\overline{\varphi(y)}, \sum_{\varphi(y)}\right) \quad (15)$$

Kernel MPM aims to find the hyperplane $a^T\varphi(z) = b$ which can separate the two classes of points with maximal probability with respect to their means and covariance matrices in the kernel space. Similar to MPM, the optimal objective function of kernel MPM can be described as follows:

$$\begin{aligned} &\min_a \sqrt{a^T \Sigma_{\varphi(x)} a} + \sqrt{a^T \Sigma_{\varphi(x)} a} \\ &\text{s.t.} \quad a^T\left(\overline{\varphi(x)} - \overline{\varphi(y)}\right) = 1 \end{aligned} \quad (16)$$

Since any solution a must lie in the span of all the samples in the kernel feature space, there exists $i = 1, 2, \cdots, n_x$ and $j = 1, 2, \cdots, n_y$ such that

$$a = \sum_{i=1}^{n_x} \alpha_i \varphi(x_i) + \sum_{j=1}^{n_y} \beta_i \varphi(y_i) \quad (17)$$

Substituting (17) into (16) and using the kernel function $K(z_1, z_2) = \varphi(z_1)^T \varphi(z_2)$, the optimal problem (16) can be rewritten as

$$\begin{aligned} &\min_\gamma \sqrt{\frac{1}{n_x}\gamma^T \tilde{K}_x^T \tilde{K}_x \gamma} + \sqrt{\frac{1}{n_y}\gamma^T \tilde{K}_y^T \tilde{K}_y \gamma} \\ &\text{s.t.} \quad \gamma^T\left(\tilde{k}_x - \tilde{k}_y\right) = 1 \end{aligned} \quad (18)$$

where

$$\gamma = \left[\alpha_1, \alpha_2, \cdots, \alpha_{n_x}, \beta_1, \beta_2, \cdots, \beta_{n_y}\right]^T \quad (19)$$

$$\tilde{k}_x \in \mathbb{R}^{n_x+n_y} \text{ with } \left[\tilde{k}_x\right]_i = \frac{1}{n_x}\sum_{j=1}^{n_x} K(x_j, z_i) \quad (20)$$

$$\tilde{k}_y \in \mathbb{R}^{n_x+n_y} \text{ with } \left[\tilde{k}_y\right]_i = \frac{1}{n_y}\sum_{j=1}^{n_y} K(y_j, z_i) \quad (21)$$

$$z_i = \begin{cases} x_i, & \text{for } i = 1, 2, \cdots, n_x \\ y_{i-n_x} & \text{for } i = n_x+1, n_x+2, \cdots, n_x+n_y \end{cases} \quad (22)$$

$$\tilde{K} = \begin{pmatrix} K_x - 1_{n_x}\tilde{k}_x^T \\ K_y - 1_{n_y}\tilde{k}_y^T \end{pmatrix} = \begin{pmatrix} \tilde{K}_x \\ \tilde{K}_y \end{pmatrix} \quad (23)$$

1_m is a column vector with ones of dimension m, K_x and K_y denote the first n_x rows and n_y rows of the kernel matrix K.

Since the optimal problem (18) is also a second order cone program, which can be solved by using interior-point methods. Once the optimal γ_* is obtained, the classification decision rule of kernel MPM is given by

$$\begin{aligned} f(z) &= \operatorname{sgn}\left(a_*^T\varphi(z) - b_*\right) \\ &= \operatorname{sgn}\left(\sum_{i=1}^{n_x+n_y}\left[\gamma_*\right]_i K(z_i, z) - b_*\right) \end{aligned} \quad (24)$$

If $f(z) = +1$ then the testing data z is classified as from class x, otherwise the testing data z is classified as from class y.

From the above computer process, we can observe that kernel function K play an important role in kernel MPM. The most commonly used kernels include Gaussian kernel and polynomial kernel. However, the nonlinear structure captured by these data-independent kernels may not be consistent with the intrinsic manifold structure. To improve the classification performance of kernel MPM algorithm, we adopt the following adaptive kernel learning method.

Since the data-dependent kernel can be obtained via pairwise constraints [11], we can construct the following similarity matrix T to represent the pairwise constraints

$$T_{ij} = \begin{cases} +1, & (x_i, x_j) \in SP \\ -1, & (x_i, x_j) \in DP \\ 0, & otherwise \end{cases} \quad (25)$$

Where *SP* denotes similar pairwise constraint (the data pairs share the same class), and *DP* denotes dissimilar pairwise constraint (the data pairs have different classes). Let L be the normalized graph Laplacian matrix defined as follows:

$$L = I - D^{-1/2} W D^{-1/2} \quad (26)$$

where I is the identity matrix, D is a diagonal matrix with its elements are $D_{ii} = \sum_j W_{ij}$, and W is the weight matrix defined on the whole data set, its definition is as follows:

$$W_{ij} = \begin{cases} 1, & \text{if } x_i \text{ is among the } k \text{ nearest neighbor of } x_j \\ & \text{or } x_j \text{ is among the } k \text{ nerest neighbor of } x_i \\ 0, & \text{otherwise} \end{cases} \quad (27)$$

Then the manifold adaptive kernel learning can be formulated the following minimization problem:

$$\min_{K \succ 0} \operatorname{Tr}(LK) + C\sum_{(x_i, x_j) \in S \cup D} l\left(T_{ij} K_{ij}\right) \quad (28)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

where $\mathrm{Tr}(\ )$ is the matrix trace, $l(\ )$ is the square hinge loss function, and C is the positive constant to control the tradeoff between the empirical loss $l(\ )$ and the intrinsic data manifold.

Since the optimal problem in (28) belongs to typical semi-definite programming (SDP) problem, which can be easily computed using the standard SDP solver SeDuMi [12]. By using the obtained optimal kernel function K in the computation process of kernel MPM, we can greatly improve the classification performance of the kernel MPM algorithm. The latter experimental results validate this conclusion.

In short, our proposed face recognition algorithm has two steps. First, we extract the discriminative facial feature with LFDA, and then the face image is recognized by using the optimal kernel MPM classifier in the reduced feature space.

4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of our proposed algorithm and compare it with the state-of-the-art algorithms for face recognition. Our empirical study on the face recognition was conducted based on three real-world face databases.

The facial recognition technology (FERET) face database [13] is a commonly used database for the test of state-of-art face recognition algorithms. In the following, the proposed algorithm is tested on a subset of this database which contains 1400 images of 200 subjects. The subset contains the images whose names are marked with two-character strings: "ba," "bj," "bk," "be," "bf," "bd," and "bg." Each subject has seven images involving variations in illumination, pose, and facial expression. In our experiment, each original image is cropped so that each cropped image only contains the portions of the face and hair. Then, the facial areas were cropped and resized to 32×32 and preprocessed by the histogram equalization. Some sample images after preprocessing of the FERET databases are shown in Figure 1. Six out of seven images of each subject are randomly chosen for training, and the remaining one is used for testing. Thus, the training set size is 1200 and the test set size is 200.The final recognition accuracy is computed by averaging all ten trials.

The UMIST database contains 20 persons with totally 564 face images [14]. There are variations of race, sex, and appearance with different subjects. The size of each image is approximately 220×220 pixels, with 256 gray levels per pixel, which are resized into 32×32 pixels in experiment. In addition, preprocessing to locate the faces was applied. Original images were normalized such that the two eyes were aligned at the same position. Some sample images after preprocessing of the UMIST databases are shown in Figure 2.The training set is a randomly selected subset with ten images per individual, and the remaining images of the database are used as the testing set. Thus, the training set size is 200.The final recognition accuracy is computed by averaging all ten trials.

The Yale face database (http://cvc.yale.edu/projects/yalefaces/yalefaces.html) contains 165 gray scale images of 15 individuals. The images demonstrate variations in lighting condition, facial expression. In this experiment, all the images are aligned by fixing the locations of the two eyes. Histogram equalization is applied as a preprocessing step. Some sample images after preprocessing of the Yale databases are shown in Figure 3. We randomly select five images of each individual to construct the training set and the rest images of the database to form the testing set. Thus, the numbers of the training samples and testing samples are 75 and 90, respectively. The final recognition accuracy is computed by averaging all ten trials.

Figure 1: Face Image Examples from the FERET Database

Figure 2: Face Image Examples from the UMIST Database

Figure 3: Face Image Examples from the Yale Database

To evaluate our proposed optimal kernel MPM (OKMPM) algorithm, we systematically compare it with Eigenface [2], Fisherface [2], Laplacianface [5], SVM [7], and the original MPM [9] algorithms on FERET, UMIST, and Yale databases. For fair comparison, we first apply LFDA to extract facial feature, then SVM (or MPM) classifier is adopted to recognize different face images in the reduced feature space. The classification accuracy for each algorithm on the three databases is reported on the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1-Table 3, respectively. From these results, we can make the follow observations: 1) our proposed OKMPM performs much better than Eigenface, Fisherface, and Laplacianface, SVM, and MPM algorithms on the three databases, which show that simultaneously using the LFDA-based feature extraction method and the optimal kernel MPM classifier can effectively improve the performance of face recognition. 2) Eigenface performs the worst. Laplacianface outperforms Eigenface and Fisherface since Laplacianface considers the manifold structure of face images. 3) Although MPM achieves better performance than SVM by explicitly considering the lower bound of the classification accuracy, it still performs worse than our proposed OKMPM. The main reason could be attributed to the fact that OKMPM can effectively capture the nonlinear manifold structure with the optimal data-adaptive kernel function.

Table 1: Recognition Accuracy Comparisons On The FERET Database

Algorithm	Accuracy
Eigenface	61.2%
Fisherface	68.5%
Laplacianface	79.3%
SVM	84.6%
MPM	85.3%
OKMPM	89.4%

In addition, to test whether our proposed optimal adaptive kernel really improve the performance of kernel MPM, we also test the performance of kernel MPM when the kernel is set different kernel functions, such as Gaussian kernel, polynomial kernel, and Sigmoid kernel. The experimental results on the three databases are shown in Table 4. As can be seen, our proposed optimal kernel function achieves the best performance among the compared kernel functions. The possible explanations are as follows: Gaussian kernel, polynomial kernel, and sigmoid kernel are all the data-independent kernels, which may not be consistent with intrinsic manifold structure. However, our proposed optimal kernel is obtained by using the pairwise constraints and exploiting the local geometry of face images. Therefore, the obtained optimal kernel can effectively capture the nonlinear manifold structure of face images, which leads to better performance of kernel MPM.

Table 2: Recognition Accuracy Comparisons On The UMIST Database

Algorithm	Accuracy
Eigenface	91.8%
Fisherface	93.2%
Laplacianface	94.4%
SVM	95.8%
MPM	96.1%
OKMPM	98.9%

Table 3: Recognition Accuracy Comparisons On The Yale Database

Algorithm	Accuracy
Eigenface	56.2%
Fisherface	77.6%
Laplacianface	88.4%
SVM	90.3%
MPM	90.7%
OKMPM	93.5%

Table 4: Recognition Accuracy Of Kernel MPM Comparisons Under Different Kernel Functions

Kernel	FERET	UMIST	Yale
Gaussian kernel	86.4%	94.3%	89.5%
Polynomial kernel	86.1%	93.9%	89.2%
Sigmoid kernel	85.6%	93.7%	89.0%
Optimal kernel	89.4%	98.9%	93.5%

5. CONCLUSION

In this paper, we have proposed a novel face recognition algorithm based on the optimal kernel minimax probability machine (OKMPM). It can effectively capture the nonlinear manifold structure with the optimal data-adaptive kernel function and obtain an explicit upper bound on the probability of misclassification of future data. The experimental results show that the proposed OKMPM algorithm performs much better than traditional face recognition algorithms. In our future work, we will

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

focus on the theoretical analysis and accelerating issues of our OKMPM algorithm.

ACKNOWLEDGMENTS

This work is supported by NSFC (Grant No. 70701013), the National Science Foundation for Post-doctoral Scientists of China (Grant No. 2011M500035), and the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No.20110023110002).

REFRENCES:

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: a literature survey", ACM Computing Surveys, Vol.35, No.4, 2003, pp.399-458.

[2] P.Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, No.7, 1997, pp.711-720.

[3] Z.Fan, Y.Xu, and D.Zhang, "Local linear discriminant analysis framework using sample neighbors", IEEE Transactions on Neural Networks, Vol.22, No.7, 2011, pp.1119-1132.

[4] D.V.Maarten, N.Dimitri, and V.H.Sabine, "A combination of parallel factor and independent component analysis", Signal Processing, Vol.92, No.12, 2012, pp.2990-2999.

[5] X.He, S.Yan, Y.Hu, P.Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.27, No.3, 2005, pp.328-340.

[6] M.Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis", Journal of Machine Learning Research, Vol.8, 2007, pp.1027-1061.

[7] K.-R.Muller, S.Mika, G.Ratsch, K.Tsuda, and B.Scholkopf, "An introduction to kernel-based learning algorithms", IEEE Transactions on Neural Networks, Vol.12, No.2, 2001, pp.181-201.

[8] G.R.G. Lanckriet, L.E.Ghaoui, C.Bhattacharyya, and M.I.Jordan, "A robust minimax approach to classification," The Journal of Machine Learning Research, Vol.3, 2002, pp.555-582.

[9] G.R.G.Lanckriet, L.E.Ghaoui, C.Bhattacharyya, and M.I.Jordan, "Minimax probability machine", Advances in Neural Information Processing Systems, Vol.14, 2001, pp.801-807.

[10] I.W H. Tsang, J.T.-Y. Kwok, "Efficient hyperkernel learning using second-order cone programming", IEEE Transactions on Neural Networks, Vol.17, No.1, 2006, pp.48-58.

[11] J. Zhuang, I.W.Tsang, and S.C.H.Hoi, "SimpleNPKL: simple non-parametric kernel learning", Proceedings of the 26th International Conference on Machine Learning (ICML), 2009, pp.1273-1280.

[12] B.Borchers, "CSDP, a C library for semidefinite programming", Optimization Methods & Software, Vol.11, No.1, 1999, pp. 613- 623.

[13] P.J.Phillips, M.Hyeonjoon, S.A.Rizvi, and P.J.Rauss, "The FERET evaluation methodology for face-recognition algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.10, 2000, pp. 1090-1104.

[14] D.B.Graham and N. M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition", NATOASI Series F, Computer and Systems Sciences, Vol.163, 1998, pp.446-456.