Enhancing Kernel Maximum Margin Projection for Face Recognition

Ziqiang Wang
Henan University of Technology, Zhengzhou, 450001, China
Email: wzqagent@126.com

Xia Sun
Henan University of Technology, Zhengzhou, 450001, China
Email: sunxiamail@126.com

***Abstract*—To efficiently deal with the face recognition problem, a novel face recognition algorithm based on enhancing kernel maximum margin projection(MMP) is proposed in this paper. The main contributions of this work are as follows. First, the nonlinear extension of MMP through kernel trick is adopted to capture the nonlinear structure of face images. Second, the kernel deformation technique is proposed to increase the discriminating capability of original input kernel function. Third, the feature vector selection approach is applied to improve computational efficiency of kernel MMP. Finally, the multiplicative update rule is employed to enhance training speed of SVM classifier for face recognition. Experimental results on face recognition demonstrate the effectiveness and efficiency of the proposed algorithm.**

***Index Terms*—face recognition, kernel maximum margin projection, support vector machine(SVM), pattern recognition**

I. INTRODUCTION

The appearance-based face recognition has received extensive attention during the past decades for its huge potentials in many applications, such as human-computer interface, biometric identity authentication and multimedia surveillance. However, a major challenge of face recognition is that the captured face image data often lies in a high-dimensional space. For example, a face image of size $n_1 \times n_2$ is represented as a vector in the face image space $\mathbb{R}^{n_1 \times n_2}$. Due to the consideration of the curse of dimensionality, it is often necessary to conduct dimensionality reduction to acquire an efficient and discriminative lower-dimensionality feature representation before formally conducting classification. To this end, principal component analysis(PCA) and linear discriminant analysis(LDA)[1] are the most well-known techniques.

PCA is an unsupervised dimensionality reduction method which aims at extracting a linear subspace in which the variance of the projected data is maximized. As a supervised dimensionality reduction technique, LDA aims to seek a linear transformation by maximizing the ratio of between-class variance and within-class variance. Some researchers have shown that LDA outperforms PCA for face recognition because discriminant information is utilized in LDA[2]. However, the global linearity of PCA and LDA prohibit their effectiveness for describing the non-linear distributed face images. To deal with this limitation, nonlinear extensions of PCA and LDA through kernel trick have been proposed, such as kernel PCA(KPCA)[3] and kernel LDA(KDA)[4] were used for face recognition and they were found to outperform their linear variants. While the aforementioned methods have attained reasonably good performance in face recognition, they may fail to discover the underlying nonlinear manifold structure as they seek only a compact Euclidean subspace for face recognition. Recent studies show that the face images are sampled from a nonlinear low-dimensional manifold which is embedded in the high-dimensional ambient space[5]. To discover the intrinsic manifold structure of the face image data, manifold learning algorithms for dimensionality reduction such as ISOMAP[6], locally linear embedding (LLE)[7] and Laplacian eigenmap (LE)[8] were recently developed. Although the above manifold learning algorithms can preserve the local or global geometric properties of the nonlinear manifold structure, these algorithms only define an embedding of the training data points and do not present a method for mapping new data points that do not exist in the training set, which is the well-known out of sample problem. A common way to resolve this problem is to use a linearization procedure to construct explicit maps over new testing data, the most representative examples is LPP[9]. Meanwhile, analyses and interpretations about these algorithms are given in view of graph embedding framework[10]. However, these algorithms are designed to best preserve data locality or similarity in the embedding space rather than good discriminating capability. Therefore, they might not be optimal in discriminating different face images which is the ultimate goal of face recognition.

Manuscript received May 14, 2012; revised May 20, 2012; accepted May 25, 2012.
Project number: 70701013.
Corresponding author: Ziqiang Wang.

doi:10.4304/jsw.8.3.724-730

In this paper, we propose a novel face recognition algorithm by using kernel maximum margin projection (KMMP). Different from traditional manifold learning algorithms, which only consider the intraclass geometry structure, while ignoring the interactions of samples from different classes. KMMP simultaneously considers both the intraclass geometry and interclass discrimination information for dimensionality reduction. In addition, unlike the linearity of the original MMP, KMMP can capture the nonlinear structure of face images with kernel trick. Therefore, KMMP can have more discriminating power. Moreover, we also adopted the kernel deformation technique to further increase the discriminating capability and applied the feature vector selection approach to reduce the computational cost of KMMP. Once the high-dimensional face image data are projected into a lower-dimensional feature space, we can apply traditional classification algorithms to classify and recognize different face images.

The rest of paper is organized as follows. In Section II, we give a brief review of MMP algorithm. Section III introduces the enhancing kernel MMP algorithm for face recognition. Extensive experimental results on face recognition are reported in Section IV. Finally, we provide concluding remarks in Section V.

II. Brief Review Of MMP

Maximum margin projection(MMP)[11] is a recently proposed manifold learning algorithm for dimensionality reduction. It is based on locality preserving neighbor relations and explicitly exploits the class information for classification. It is a graph-based approach for learning a linear approximation to the intrinsic data manifold by making use of both labeled and unlabeled data. Its goal is to discover both geometrical and discriminant structures of the data manifold.

Given a set of face images $\{x_1,\ldots,x_n\}\subset \mathbb{R}^m$ and the corresponding class label $c_1,c_2,\ldots,c_n\in\{1,2,\ldots,p\}$, let $X=[x_1,x_2,\ldots,x_n]$. MMP aims to seek a facial feature subspace that preserves the local geometrical and discriminant structures of the high-dimensional face manifold. Let S_b and S_w denote weight matrices of between-class graph G_b and within-class graph G_w respectively. MMP attempts to ensure that the connected points of G_w are as close together as possible while the connected points of G_b are as far apart as possible, it can be obtained by solving the following optimization problem:

$$\begin{aligned}&\arg\min_a \sum_{i=1}^{n}\sum_{j=1}^{n}\left(a^T x_i - a^T x_j\right)^2 S_{w,ij}\\&=\arg\min_a a^T X\left(D_w - S_w\right)X^T a\end{aligned} \tag{1}$$

$$\begin{aligned}&\arg\max_a \sum_{i=1}^{n}\sum_{j=1}^{n}\left(a^T x_i - a^T x_j\right)^2 S_{b,ij}\\&=\arg\max_a a^T X L_b X^T a\end{aligned} \tag{2}$$

with the constraint

$$a^T X D_w X^T a = 1 \tag{3}$$

where $L_b = D_b - S_b$ is the Laplacian matrix of G_b, D_b is a diagonal matrix whose entries on diagonal are column sum of S_b, i.e., $D_{b,ii}=\sum_{j=1}^{n} S_{b,ij}$, D_w is a diagonal matrix whose entries on diagonal are column sum of S_w, i.e., $D_{w,ii}=\sum_{j=1}^{n} S_{w,ij}$.

The definitions of weight matrices S_b and S_w are as follows:

$$S_{w,ij}=\begin{cases}\gamma, & \text{if } x_i \text{ and } x_j \text{ share the same label}\\ 1, & \text{if } x_i \text{ or } x_j \text{ is unlabeled but } x_i\in N_w(x_j) \text{ or } x_j\in N_w(x_i)\\ 0, & \text{otherwise}\end{cases} \tag{4}$$

$$S_{b,ij}=\begin{cases}1, & \text{if } x_i\in N_b(x_j) \text{ or } x_j\in N_b(x_i)\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

where $N(x_i)=\{x_i^1,\ldots,x_i^k\}$ denote the set of its k nearest neighbors, $l(x_i)$ represents the label of x_i, $N_b(x_i)=\{x_i^j \,|\, l(x_i^j)\neq l(x_i), j=1,\ldots,k\}$ contains the neighbors having different labels, and $N_w(x_i)=N(x_i)-N_b(x_i)$ contains the rest of the neighbors.

Then, minimizing (1) and maximizing (2) under the constraint (3) can be reduced to the following optimization problem:

$$\arg\max_a a^T X\left(\beta L_b + (1-\beta) S_w\right) X^T a \tag{6}$$

where $\beta\in[0,1]$ is a suitable constant which controls the weight between the within-class graph and between-class graph.

Finally, by using simple algebraic transformation, the projection vectors of MMP are the eigenvectors associated with the largest eigenvalues of the following generalized eigenvalue problem:

$$X\left(\beta L_b + (1-\beta) S_w\right) X^T a = \lambda X D_w X^T a \tag{7}$$

Since XD_wX^T is nonsingular after applying PCA to throw away the components corresponding to zero eigenvalues, the projection vector of MMP can be regarded as the eigenvectors of the matrix $\left(XD_wX^T\right)^{-1} X\left(\beta L_b + (1-\beta)S_w\right)X^T$ associated with the largest eigenvalues.

III. The Enhancing Kernel MMP Algorithm

A. Kernel MMP

As a linear dimensionality reduction algorithm, MMP is easy to understand and is very simple to implement, but it often fails to deliver good performance when face images are subject to complex nonlinear changes due to large pose, expression or illumination variations, for it is a linear method in nature. To deal with this limitation, the nonlinear extensions of MMP through "kernel trick" are proposed in the following. Such a generalization is of great importance since the kernel MMP (KMMP) would generally achieve better recognition accuracy, and relax the restriction of MMP being only a linear manifold learning algorithm.

The idea of KMMP is to solve the problem of MMP in a implicit feature space F which is constructed by the kernel trick[12]. The intuition of kernel trick is map the input data x from the original feature space into a higher dimensional Hilbert space F constructed by the nonlinear mapping

$$\varphi : x \in \mathbb{R}^m \to \varphi(x) \in F \tag{8}$$

in which the face image data may be linearly separable. Then building linear MMP algorithms in the feature space implement nonlinear counterparts in the input data space. The map, rather than being given in an explicit form, is presented implicitly by specifying a kernel function $K(,)$ as the inner product between each pair of points in the feature space.

$$K(x_i, x_j) = (\varphi(x_i) \cdot \varphi(x_j)) \tag{9}$$

Performing KMMP in the feature space F means the connected points of G_w are as close together as possible while the connected points of G_b are as far apart as possible. This is equivalent to solving the below optimization problem:

$$\begin{aligned} &\arg\min_a \sum_{i=1}^n \sum_{j=1}^n \left(a^T\varphi(x_i) - a^T\varphi(x_j)\right)^2 S_{w,ij} \\ &= \arg\min_a a^T\varphi(X)(D_w - S_w)\varphi(X)^T a \end{aligned} \tag{10}$$

$$\begin{aligned} &\arg\max_a \sum_{i=1}^n \sum_{j=1}^n \left(a^T\varphi(x_i) - a^T\varphi(x_j)\right)^2 S_{b,ij} \\ &= \arg\max_a a^T\varphi(X) L_b \varphi(X)^T a \end{aligned} \tag{11}$$

with the constraint

$$a^T\varphi(X) D_w \varphi(X)^T a = 1 \tag{12}$$

Then, under the constraint of (12), the optimal objective function of (10) can be rewritten as follows:

$$\begin{aligned} &\arg\min_a \sum_{i=1}^n \sum_{j=1}^n \left(a^T\varphi(x_i) - a^T\varphi(x_j)\right)^2 S_{w,ij} \\ &= \arg\min_a \left(1 - a^T\varphi(X) S_w \varphi(X)^T a\right) \\ &= \arg\max_a a^T\varphi(X) S_w \varphi(X)^T a \end{aligned} \tag{13}$$

Finally, the optimal objective function of KMMP can be formulated as follows by simultaneously considering (10) and (11) with the constraint of (12).

$$\arg\max_a a^T\varphi(X)\left(\beta L_b + (1-\beta) S_w\right)\varphi(X)^T a \tag{14}$$

Then, the transformation vector a that maximizes the above objective function is given by the maximum eigenvalue solution to the generalized eigenvalue problem:

$$\begin{aligned} &\varphi(X)\left(\beta L_b + (1-\beta) S_w\right)\varphi(X)^T a \\ &= \lambda \varphi(X) D_w \varphi(X)^T a \end{aligned} \tag{15}$$

Since the eigenvectors of (15) must lie in the span of all the samples in the feature space F , there exist coefficients $\omega_i, i = 1, 2, \ldots, n$ such that

$$a = \sum_{i=1}^n \omega_i \varphi(x_i) = \varphi(X)\omega \tag{16}$$

where $\omega = [\omega_1, \omega_2, \ldots, \omega_n]^T$.

By using (16) and (9), we can rewrite (15) as follows:

$$K\left(\beta L_b + (1-\beta) S_w\right) K^T \omega = \lambda K D_w K^T \omega \tag{17}$$

Then, the problem of KMMP is converted into finding the leading eigenvectors of the matrix $(KD_wK^T)^{-1} K(\beta L_b + (1-\beta)S_w)K^T$. Let $\omega_1, \omega_2, \ldots, \omega_l$ be the solution of (17) ordered according to their eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_l$. Thus, for a new face image data x , its projection onto a in the feature space F can be calculated as follows:

$$x \to y = (a \cdot \varphi(x)) = \sum_{i=1}^n \omega_i K(x_i, x) \tag{17}$$

For face recognition, a problem arises that the matrix KD_wK^T can be singular, which stems from the fact that the dimension of the kernel feature space is usually much higher than that of the empirical feature space, a deficiency that is generally known as small sample size (SSS) problem. One possible way to address the SSS problem is by performing kernel PCA (KPCA) projection to reduce the dimension of the feature space and make the matrix KD_wK^T nonsingular.

B. Kernel Deformation Technique

Since the choice of kernel functions can significantly affect the performance of kernel methods, a good choice of the kernel is imperative to the success of any kernel method. Therefore, substantial efforts have been made to design appropriate kernels for the problems at hand. Until now, how to select a suitable kernel function for a given application is still an open issue. Traditionally, the kernel function has been chosen to be either linear, polynomial, or Gaussian kernel. But these kernel functions do not take full advantage of the specific characteristics of face image data. Here a new kernel function $K(x, y)$, which is based on kernel deformation technique[13], is proposed to improve the performance of KMMP as follows:

$$K(x, y) = k(x, y)\left(I + Lk(x, y)\right)^{-1} \tag{19}$$

where $k(x, y)$ is the original input kernel function, we chose the Gaussian kernel $k(x, y) = \exp\left(-\gamma \|x - y\|^2\right)$ since it achieves the superior performance in many pattern classification applications, I is the identity matrix, L is the graph Laplacian matrix that models the underlying geometry structure of face image data, and the graph Laplacian is defined as $L = D - W$, where D is a diagonal matrix given by $D_{ii} = \sum_i W_{ij}$ and

$$W = \begin{cases} e^{-\left(\left\|x_i - x_j\right\|^2 / 2\sigma^2\right)} & \text{,if } x_i \text{ and } x_j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

The main idea of the above kernel deformation technique is to estimate the intrinsic manifold structure via a nearest neighbor graph which preserves the local geometry structure of the face image space, then incorporate them into the kernel deformation procedure. Thus, the resulting new kernel can take advantage of both local geometry structure and discrimination information. In fact, recent researches reveal that the locality features and intrinsic geometric structures in the input space may take on additional discriminating power for classification. Therefore, when an input kernel $k(x, y)$ is deformed according to the local geometry structure, the resulting kernel function $K(x, y)$ may be able to perform much better than the original input kernel. Detailed performance evaluation of the deformation kernel is reported in Section III.

C. Feature Vector Selection Approach

From (16), we can observe that the kernel trick-based KMMP algorithm is computationally expensive in the training phase since its computational complexity is proportional to the number of training points needed to represent the transformation vectors. In fact, the dimensionality of the data subspace spanned by $\varphi(x_i)$ is given by the rank of kernel matrix K, and the $rank(K) \ll n$ for massive training data set. If we replace n with $rank(K)$ and select a corresponding subset of feature vectors in the feature space F, which will greatly improve the computational efficiency of KMMP. Based on the above consideration, we adopt the feature vector selection approach[14] to accelerate the running speed of KMMP.

The essential idea of the feature vector selection is to find a subset which is sufficient to express all the data as a linear combination of the selected subset in the feature space F. Let the selected feature vector subset $S = \{\varphi(x_{s1}), \varphi(x_{s2}), \dots, \varphi(x_{sr})\}$ in the feature space F is known, where r denotes the number of selected feature vector, then we can estimate the mapping $\widehat{\varphi}(x_i)$ of any input data x_i as a linear combination of φ_S in feature space F. The formal description is as follows:

$$\widehat{\varphi}(x_i) = \varphi_S \cdot \beta_i \tag{21}$$

where $\beta_i = \left(\beta_i^1, \beta_i^2, \dots, \beta_i^r\right)$ is the coefficient vector.

Then, the goal of feature vector selection is to find the coefficients β_i so that the estimated mapping $\widehat{\varphi}(x_i)$ approaches to the real mapping $\varphi(x_i)$ as far as possible, which can be attained by minimizing the following objective function:

$$\delta_i = \frac{\left\|\varphi(x_i) - \widehat{\varphi}(x_i)\right\|^2}{\left\|\varphi(x_i)\right\|^2} \tag{22}$$

The above optimization problem is performed by setting the partial derivative of δ_i with respect to β_i to zero. By using matrix form, the optimal objective function of (22) can be rewritten as follow:

$$\min \delta_i = 1 - \frac{K_{Si}^T K_{SS}^{-1} K_{Si}}{K_{ii}} \tag{23}$$

where K_{SS} is a square matrix of dot products of the selected vectors, and K_{Si} is the vector of dot product between x_i and the selected vector set S.

Then, the ultimate goal of feature vector selection method is make (23) apply to all the sample data, which can be summarized as the following form:

$$\max_S J_S = \frac{1}{n} \sum_{x_i \in X} \left(\frac{K_{Si}^T K_{SS}^{-1} K_{Si}}{K_{ii}} \right) \tag{24}$$

The solution of the above optimal problem can be obtained with an iterative algorithm described in [14], and the algorithm stops when K_{SS} is no longer invertible or the predefined number of selected vectors is reached.

D. Classification Method

After the transformation by KMMP, the facial feature matrix is obtained for each face image. Then, face recognition becomes a pattern classification task, the SVM classifier is used for classification because of its good generalization ability in minimizing the VC dimension and achieving a minimal structural risk[12]. The optimal objective function of SVM is as follows:

$$Q(\alpha) = \max_\alpha \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{25}$$

with the constraint

$$\sum_{i=1}^n \alpha_i y_i = 0 \text{ and } 0 \le \alpha_i \le C, i = 1, 2, \dots, n. \tag{26}$$

Once the optimal α is obtained by solving the quadratic programming (QP) problem of (25), the classification decision function of SVM classifier is given as follows:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right) \tag{27}$$

Although the quadratic programming (QP) problem of (25) has the important computational advantage of not suffering from of local minima, given n training samples, the naive implementation of QP solver is of

$O(n^3)$ computational complexity, which is computationally infeasible on very large face image data sets. Hence, a replacement of the naive method for solving QP solutions posed by the SVM classifier is highly desirable. To this end, we applied the following multiplicative update rule-based method[15] to improve the training speed of SVM classifier.

$$\alpha_i \leftarrow \alpha_i \left[\frac{1+\sqrt{1+4(A^+\alpha)_i(A^-\alpha)_i}}{2(A^+\alpha)_i} \right] \tag{28}$$

where

$$A_{ij} = y_i y_j K(x_i, x_j) \tag{29}$$

$$A_{ij}^+ = \begin{cases} A_{ij}, if\ A_{ij} \geq 0 \\ 0\ \ , otherwise \end{cases} \tag{30}$$

$$A_{ij}^- = \begin{cases} |A_{ij}|, if\ A_{ij} < 0 \\ 0\ \ , otherwise \end{cases} \tag{31}$$

The remarkable advantages of the multiplicative iterative updates in (28) is that it can be parallel implemented and never violate the nonnegativity condition constraints. Furthermore, it has been proved that the multiplicative updates rule can monotonically improve the optimal objective function of (25).

E. The Whole Algorithm Description

Based on the above statements, we summarize our proposed enhancing KMMP algorithm for face recognition as follows:

1) PCA Preprocessing. We project the face images x_i into the PCA subspace by throwing away the components corresponding to zero eigenvalue.
2) Computing the weight matrices S_b and S_w of the between-class graph G_b and the within-class graph G_w according to (5) and (4), respectively.
3) Constructing the optimal objective function of kernel MMP according to (14) and (12).
4) Computing the adaptive kernel function according to the kernel deformation technique defined in (19).
5) Obtaining the subset of feature vectors in the feature space by solving the optimal problem defined in (24).
6) Obtaining the transformation vectors of KMMP by solving the generalized eigen-problem defined in (17).
7) Projecting the face images into the lower-dimensional feature space via (17).
8) Constructing the optimal objective function of SVM in the lower-dimensional feature space via (25) and (26).
9) Obtaining the optimal support vectors via the multiplicative update rule defined in (28).
10) The face images can be classified and recognized in terms of (27).

In summary, the face recognition process has three steps. We first calculate the face subspace by dimensionality reduction algorithm KMMP. Then, facial images are projected into the face subspaces. Finally, the SVM classifier which is trained via multiplicative update rule is adopted to recognize new facial images.

IV. Experimental Results

To demonstrate the performance of our proposed enhancing KMMP algorithm for face recognition, the extensive experiments are carried out on a hybrid face image database of 123 persons and 1925 images, which is a collection of the following three databases:

1) The Yale database (http://cvc.yale.edu/projects/ yalefaces/yalefaces.html) contains 165 images of 15 individuals, each person has 11 different images under various facial expressions and lighting conditions.
2) The ORL database (http://www.uk.research.att. com/ facedatabase.html) contains images from 40 individuals, each providing 10 different face images.
3) The subset of the CMU PIE database[16]. There are 68 peoples and each person has 20 different face images.

All the face images are manually aligned, cropped, and then resized to 32×32 pixels. Histogram equalization was used for the normalization of the facial image luminance. Some sample face images after preprocessing of the three databases are shown in Figure 1-Figure 3. The mixture face image database is splitted into two non-overlapping set for training and testing. Each database partition is performed with one-half images per person for training, and the rest of these databases for testing. To reduce statistical variability, final results are based on averages over ten random repetitions.

In our experiments, the proposed KMMP algorithm is compared with kernel PCA(KPCA)[3], kernel LDA(KDA)[4], kernel LPP(KLPP)[9], and MMP[11]. The original input Gaussian kernel function $k(x, y)$ with parameters set to $\gamma = 2^{(n-20)/2.5}\gamma_0$, $n = 0,1,\ldots,20$ are used, where γ_0 is the stand deviation of the training set. We report the best result of each algorithm from among the 21 experiments. The parameter C in SVM classifier is tuned by 10-cross validation for data sets. The experimental results are shown in Figure 4. From these results, we can make the following observations.

1) Our proposed KMMP algorithm consistently outperforms KPCA, KDA, KLPP, and MMP algorithms, which demonstrate that KMMP can effectively utilize local manifold structure as well as the discriminant information for face recognition. Meanwhile, KMMP can capture the nonlinear structure of face images with kernel trick. Therefore, KMMP can have more discriminating power. Besides the factor such as choosing KMMP for dimensionality reduction, the selection of kernel deformation technique,

feature vector selection approach, and the training SVM classifier based on multiplicative update rule also play an important role.

2) KPCA performs the worst, this is probably because KPCA is an unsupervised algorithm that ignores the valuable label information for classification.

3) The performance of KDA and KLPP very close to each other. A possible explanation is as follows: Although KDA is a supervised method, it does not consider the manifold structure; For KLPP, it aims to discover the local manifold structure rather than discriminating information. Thus, it is debatable whether the label information or local manifold structure is more important.

4) Since MMP considers both local manifold structure and label information, it achieves higher performance than KPCA, KDA and KLPP. However, as a linear algorithm, MMP may fail to capture the nonlinear structure due to the high variability of the image content and style. Therefore, the performance of MMP is inferior to KMMP.

In addition, the construction of kernel is one of the key techniques in our enhancing KMMP algorithm, we use the kernel deformation technique in constructing the kernel function. We can also use other kinds of kernel function such as linear, polynomial, or Gaussian kernel. In this experiment, we test the KMMP recognition performance under different kernel functions. The recognition results are show in Figure 5. As can be seen, our proposed kernel deformation technique performs much better than linear, polynomial, and Gaussian kernel. This is mainly because the kernel deformation technique can takes advantage of both local geometry structure and discrimination information, and the intrinsic geometric structures in the input space may take on additional discriminating power for classification. Therefore, the kernel deformation technique achieves better recognition rates than traditional kernel functions.

Finally, to verify the efficiency of our proposed enhancing KMMP algorithms, we only record the computational times of the above three kernel-based algorithms in this experiment. The running time comparisons are reported in Table I. The experimental results show that the proposed enhancing KMMP algorithms are much more efficient than the other kernel-based algorithms(such as KPCA, KDA and KLPP). The main reason could be attributed to the fact that the feature vector selection strategy accelerates the running speed of KMMP, and the multiplicative update rule-based method further improve the training speed of SVM classifier. Therefore, our proposed KMMP algorithm could dramatically reduce the computational time when compared to other kernel-based algorithms on large scale face recognition problem.

V. Conclusions

In face recognition, dimensionality reduction techniques are widely employed to reduce the dimensionality of face image data and enhance the discriminatory information. In this paper, an enhancing KMMP algorithm for face recognition is proposed. The experimental results demonstrate the effectiveness and efficiency of the proposed algorithm.

Appendix A Figure(1-5) and Table I

Figure 1. Face image examples of the ORL database.

Figure 2. Face image examples of the Yale database.

Figure 3. Face image examples of the PIE database.

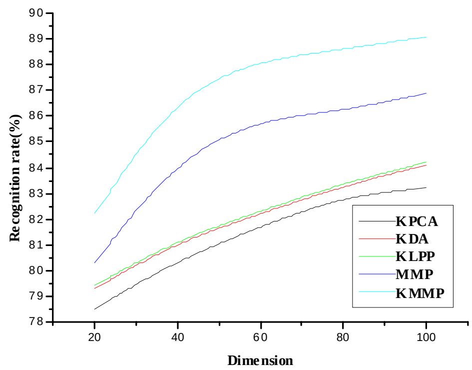

Figure 4. Comparison results of the five algorithms.

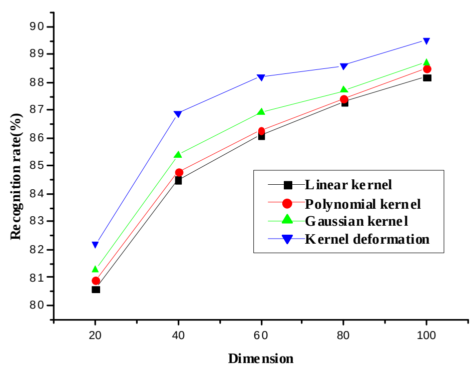

Figure 5. Comparison results of different kernel functions.

TABLE I.
RUNNING TIME COMPARISON ON THE FACE IMAGE DATABASE

Algorithms	Running time(s)
KPCA	67.2
KDA	83.5
KLPP	59.3
KMMP	34.8

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No.70701013, the Natural Science Foundation of Henan Province under Grant No.102300410020, the National Science Foundation for Post-doctoral Scientists of China under Grant No. 2011M500035, and the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No.20110023110002.

REFERENCES

[1] R.O.Duda, P.E.Hart, and D. G. Stork, *Pattern Classification*, 2nd edition. Hoboken, NJ: Wiley-Interscience, 2000.
[2] B.Mandal and H.-L. Eng, "Regularized discriminant analysis for holistic human activity recognition," *IEEE Intelligent Systems*, vol.27, pp.21-31, January 2012.
[3] B.Scholkopf, A.Smola, and K.R.Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*,vol.10,pp.1299-1319,July 1998.
[4] G.Baudat and F.Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol.12, pp.2385-2404, October 2000.
[5] Y.Huang, D.Xu, and F.Nie, "Patch distribution compatible semisupervised dimension reduction for face and human gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.22, pp.479-488, March 2012.
[6] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp.2319-2323, December 2000.
[7] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp.2323-2326, December 2000.
[8] M.Belkin and P.Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol.15, pp.1373-1396,June 2003.
[9] X. He, P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems*, Cambridge, MA:MIT Press,2003,pp.385-392.
[10] S.Yan, D.Xu, B.Zhang, H.-J.Zhang, Q.Yang, and S.Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29, pp.40-51, January 2007.
[11] X.He, D.Cai, and J.Han, "Learning a maximum margin subspace for image retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol.20, pp.189-201, February 2008.
[12] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[13] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: from transductive to semi-supervised learning," in *Proceedings of 2005 International Conference of Machine learning*, New York: ACM Press, 2005, pp.824-831.
[14] G.Baudat and F.Anouar, "Kernel-based methods and function approximation," in *Proceedings of 2001 International Conference on Neural Networks*, Piscataway: IEEE Press, 2001, pp.1244-1249.
[15] F.Sha, Y.Lin, L.K. Saul, and D.D.Lee, "Multiplicative updates for nonnegative quadratic programming," *Neural Computation*, vol.19, pp.2004-2031,August 2007.
[16] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illuminlation, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.25, pp.1615-1618, December 2003.

Ziqiang Wang was born Zhoukou, Henan, China in 1973. He received the PhD degree from China University of Mining &Technology (Beijing) and the Master of Science degree from Xi'an Petroleum University, both in computer application, in 2011 and 2002, respectively. His research interests include data mining and pattern recognition.

He is currently an associate professor in the Henan University of Technology.

Xia Sun was born Xi'an, Shanxi, China in 1978. She received the Master of Science degree in computer application from Huazhong University of Science and Technology in 2008. Her research interests include data mining and pattern recognition.

She is currently a lecturer in the Henan University of Technology.