Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 724978, 10 pages
http://dx.doi.org/10.1155/2014/724978

Research Article

An Improved Metric Learning Approach for Degraded Face Recognition

Guofeng Zou,[1] Yuanyuan Zhang,[1] Kejun Wang,[2] Shuming Jiang,[1] Huisong Wan,[1] and Guixia Fu[2]

[1] *Information Research Institute of Shandong Academy of Sciences, Jinan 250014, China*
[2] *College of Automation, Harbin Engineering University, Harbin 150001, China*

Correspondence should be addressed to Guofeng Zou; zgf841122@163.com and Kejun Wang; wangkejun@hrbeu.edu.cn

Received 30 July 2013; Revised 20 December 2013; Accepted 20 December 2013; Published 11 February 2014

Academic Editor: Chung-Hao Chen

To solve the matching problem of the elements in different data collections, an improved coupled metric learning approach is proposed. First, we improved the supervised locality preserving projection algorithm and added the within-class and between-class information of the improved algorithm to coupled metric learning, so a novel coupled metric learning method is proposed. Furthermore, we extended this algorithm to nonlinear space, and the kernel coupled metric learning method based on supervised locality preserving projection is proposed. In kernel coupled metric learning approach, two elements of different collections are mapped to the unified high dimensional feature space by kernel function, and then generalized metric learning is performed in this space. Experiments based on Yale and CAS-PEAL-R1 face databases demonstrate that the proposed kernel coupled approach performs better in low-resolution and fuzzy face recognition and can reduce the computing time; it is an effective metric method.

1. Introduction

The metric is a function which gives the scalar distance between two patterns. Distance metric is an important basis for similarity measure between samples, and it is one of the core issues in pattern recognition. The aim of distance metric learning is to find a distance metric matrix; its essence is to obtain another representation method with better class separability by linear or nonlinear transformation.

In recent years, some researches about distance metric have been done by researchers [1–7]. They learn a distance metric by introducing sample similarity constraint or category information; the distance metric is used to improve the data clustering or classification. These researches can be concluded to two categories: linear distance metric learning and nonlinear distance metric learning. The linear distance metric learning is equivalent to learning a linear transformation in sample space, including a variety of common linear dimensionality reduction methods, such as principal component analysis [8], linear discriminant analysis [9], and independent component analysis method [10]. The nonlinear distance metric learning is equivalent to learning a nonlinear transformation in sample space; the locally linear embedding [11], isometric mapping [12], and Laplace mapping [13] are the traditional nonlinear methods. Recently, some new nonlinear distance metric methods have been proposed. Baghshah and Shouraki [14] proposed the nonlinear metric learning method based on pair-wise similarity and dissimilarity constraints and the geometrical structure of data. Babagholami-Mohamadabadi et al. [15] proposed the probabilistic nonlinear distance metric learning. The deep nonlinear metric learning method [16] based on neural networks is a new nonlinear metric learning method. In addition, there are some more flexible distance metric learning algorithms, which are based on kernel matrix [7, 17, 18].

These traditional distance metric learning methods are defined on the set of single attribute. If the elements belong to different sets with different attribute, these distance measurement methods are incapable for the distance metric. For example, for two images with different resolution, which can be considered to belong to different sets, obviously, the traditional distance metric method will not be able to directly

calculate the distance. The general approach is normalized operation before recognition by using interpolation algorithm or sampling algorithm. But the interpolation inevitably introduced false information, and sampling may miss some useful information, so it is difficult to get high recognition rate.

Aiming at the shortage of traditional distance metric, Li et al. proposed the coupled metric learning (CML) [19–21]. The goal of coupled metric learning is to find a coupled distance function to meet the specific requirement in given task. The essential idea is that, firstly, the data in different collections are projected to the unified coupled space, and the data should be as close as possible in this new space. Then, the generalized metric learning is performed in this unified coupled space. Obliviously, this new metric method will have broader application scope and better recognition effect. However, the proposed coupled metric learning methods are based on linear transformation, which can be called linear coupled metric learning (LCML). These methods have two shortages in dealing with practical problems. First, the practical problems usually are nonlinear, and the linear transformation does not represent the features effectively. Secondly, the image needs to be converted into one-dimensional vector in the LCML algorithm; it is easy to cause the increasing of dimensions of autocorrelation matrix.

Based on the idea of coupled metric learning, we improved the supervised locality preserving projection algorithm and added supervised locality preserving information to coupled metric learning. So the improved coupled metric learning approach based on supervised locality preserving projection (SLPP-CML) was proposed. Introducing kernel technology into the coupled metric learning, we proposed the kernel coupled metric learning approach based on supervised locality preserving projection (SLPP-KCML). The SLPP-KCML realized the extension from linear coupled metric to nonlinear coupled metric. To verify the effectiveness of the proposed method, the experiments based on two face databases were performed. The experimental results show that a higher recognition rate can be achieved in the SLPP-KCML algorithm, and the operation time is reduced greatly.

2. Related Works

The traditional distance metric learning algorithm is to learn a distance function $d(x_i, x_j)$ between the data points expressed as follows:

$$d\left(x_i, x_j\right) = \left\|x_i - x_j\right\|_A = \sqrt{\left(x_i - x_j\right)^T A \left(x_i - x_j\right)}. \quad (1)$$

Distance metric learning aims to find a distance metric matrix A; it is required that A is a real symmetric and positive semidefinite matrix; namely, $A = P^T P$, where P is a transformation matrix

$$d_A\left(x_i, x_j\right) = d_P\left(x_i, x_j\right) = \sqrt{\left(Px_i - Px_j\right)^T \left(Px_i - Px_j\right)}. \quad (2)$$

Obviously, the distance metric learning is realized by learning a transformation matrix P, so the process of distance metric learning is equivalent to the process of obtaining other representation forms with better separability through linear or nonlinear transformation of the samples.

If $X \subset R^{D_x}$, $Y \subset R^{D_y}$ represent two different collections, respectively, the function $d(x, y)$ is the distance metric between data $x \in X$ and data $y \in Y$. If $D_x \neq D_y$, the traditional method does not work for distance metric. Even if $D_x = D_y$, because the data $x \in X$ and data $y \in Y$, which belong to different attribute collections, the distance metric has no physical meaning.

The coupled distance metric is a distance function for the data elements of different kinds of collections. The elements of collections X and Y are mapped from the original space to a common coupled space R^{D_C} by using the mapping functions f_x and f_y. Then, the distance metric is performed in the coupled space. The measured distance can be represented mathematically as

$$\begin{aligned} d^c(x, y) &= d_A(\tilde{x}, \tilde{y}) = d_A\left(f_x(x), f_y(y)\right) \\ &= \left\|f_x(x) - f_y(y)\right\|_A \\ &= \sqrt{\left(f_x(x) - f_y(y)\right)^T A \left(f_x(x) - f_y(y)\right)}, \end{aligned} \quad (3)$$

where matrix A is a real symmetric and positive semidefinite matrix. Letting $A = W_a W_a^T$, we can get

$$\begin{aligned} &d^c(x, y) \\ &= \sqrt{\left(f_x(x) - f_y(y)\right)^T W_a W_a^T \left(f_x(x) - f_y(y)\right)} \\ &= \sqrt{\left(W_a^T f_x(x) - W_a^T f_y(y)\right)^T \left(W_a^T f_x(x) - W_a^T f_y(y)\right)} \\ &= \left\|W_a^T \left(f_x(x) - f_y(y)\right)\right\|. \end{aligned} \quad (4)$$

The goal of coupled metric learning can be achieved by minimizing the distance function; the objective function is as follows:

$$\min J = \min \sum_{(i,j)\in C} \left\|W_a^T \left(f_x(x_i) - f_y(y_j)\right)\right\|^2, \quad (5)$$

where C is a correlation matrix of elements in collections X and Y. According to different supervised information, we can obtain different matrix C, so as to realize the different coupled metric learning.

3. The Coupled Metric Learning Based on Supervised Locality Preserving Projection (SLPP-CML)

The coupled distance metric learning must be used under the constraints of supervised information. In this paper, we improved the supervised locality preserving projection (SLPP) algorithm [22]. Based on the improved SLPP algorithm, we proposed the coupled metric learning method based on supervised locality preserving projection.

In order to better illustrate the coupled metric learning algorithm based on supervised locality preserving projection, we provide a theorem about the matrix norm.

Theorem 1. *Letting $A \in R^{n\times n}$, then the Frobenius norm has the following properties:*

(1) $\|A\|^2 = \mathrm{Tr}(A^TA) = \sum_{i=1}^{n} \lambda_i(A^TA)$, *where λ_i is the eigenvalue of the matrix A^TA;*

(2) $\mathrm{Tr}(A^TA) = \mathrm{Tr}(AA^T)$, $\mathrm{Tr}(\cdot)$ *represents the trace operation.*

The coupled metric learning based on supervised locality preserving projection includes the following steps.

Step 1 (building the neighborhood relation in the same collection). We use the k nearest neighbor method. First, building within-class adjacency graph in the same collection, if the data point $x_i(y_i)$ is one of the k within-class nearest neighbors of data point $x_j(y_j)$, we connect these two data points; and then, building between-class adjacency graph in same collection; if the data point $x_i(y_i)$ is one of the k between-class nearest neighbors of data point $x_j(y_j)$, these two data points are connected.

Step 2 (building the connected relation between two collections). If the data points x_i and y_j in two different collections belong to the same class, then these two points are connected, otherwise not connected.

Step 3 (constructing the relation matrix in the same collection). According to the neighborhood relations, the relation matrixes (similarity matrixes) of within-class and between-class are constructed, respectively.

Within-class similarity matrix is W corresponding to within-class adjacency graph and the within-class similarity value is W_{ij}. The definition is as follows:

$$W_{ij} = \begin{cases} \exp\left(-\dfrac{\|x_i - x_j\|^2}{t}\right) & \text{if } x_i \text{ connected } x_j, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Between-class similarity matrix is B corresponding to between-class adjacency graph and the between-class similarity value is B_{ij}. It can be defined as follows:

$$B_{ij} = \begin{cases} \exp\left(-\dfrac{\|x_i - x_j\|^2}{t}\right) & \text{if } x_i \text{ connected } x_j, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where parameter t is the average distance between all sample points.

Step 4 (constructing the relation matrix S between two collections). The similarity value is as follows:

$$S_{ij} = \begin{cases} 1 & \text{if } x_i \text{ connected } y_j, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Step 5 (calculating the final similarity matrix C between two collections). As shown in Figure 1, the similarity relations between element $x_1 \in X$ and elements of collection Y include the following several situations.

(a) The similarity between x_1 and y_1: these two data points in different collections belong to the same class and they are connected to each other, so the similarity of which is $C_{11} = S_{11} = 1$.

(b) The similarity between x_1 and y_5: these two data points belong to different class, but the relationship between y_5 and y_3 is the between-class neighborhood relation in same collection, and the similarity B_{35} is the maximum similarity value, so similarity between x_1 and y_5 is $C_{15} = B_{35}$.

(c) The similarity between x_1 and y_6: these two data points belong to different class and the y_6 does not have between-class neighborhood relation with any element of collection Y of class 1. But there is a between-class neighborhood relation of same collection between y_5 and y_3, and within-class neighborhood relation between y_5 and y_6. So the similarity between x_1 and y_6 is defined as the product of between-class similarity B_{35} and within-class similarity W_{56}, which is the maximum similarity between y_6 and y_3; that is, $C_{16} = B_{35} \cdot W_{56}$.

(d) The similarity between x_1 and y_9: these two data points belong to different class; there are not any between-class neighborhood relations between the elements of class 1 and class 3 in collection Y; namely, $C_{19} = 0$.

Step 6 (constructing the optimal objective function). Cosider the following:

$$\begin{aligned} J &= \sum_{(i,j)\in C} \left\|W_a^T\left(f_x(x_i) - f_y(y_j)\right)\right\|^2 \\ &= \sum_i \sum_j \left\|W_a^T\left(f_x(x_i) - f_y(y_j)\right)\right\|^2 \cdot C_{ij}, \end{aligned} \tag{9}$$

where the functions f_x and f_y are considered to be linear; that is, $f_x(x) = W_x^T x$, $f_y(y) = W_y^T y$. The optimal objective function can be rewritten as follow:

$$\begin{aligned} J &= \sum_i \sum_j \left\|W_a^T\left(f_x(x_i) - f_y(y_j)\right)\right\|^2 \cdot C_{ij} \\ &= \sum_i \sum_j \left\|W_a^T W_x^T x_i - W_a^T W_y^T y_j\right\|^2 \cdot C_{ij}. \end{aligned} \tag{10}$$

Letting $P_x = W_x W_a$, $P_y = W_y W_a$, we can get

$$\begin{aligned} J &= \sum_i \sum_j \left\|W_a^T W_x^T x_i - W_a^T W_y^T y_j\right\|^2 \cdot C_{ij} \\ &= \sum_i \sum_j \left\|P_x^T x_i - P_y^T y_j\right\|^2 \cdot C_{ij}. \end{aligned} \tag{11}$$

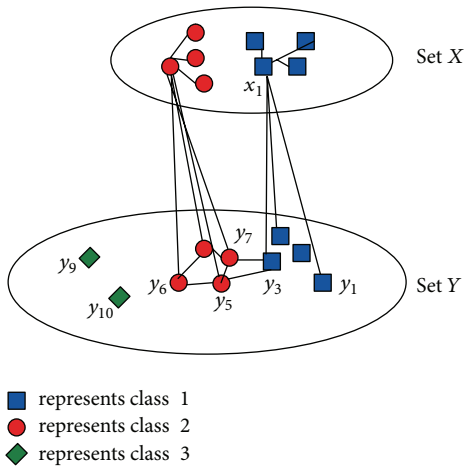


Figure 1: The relationship between the elements of different collections.

Therefore, our method aims to learn two linear transformations P_x and P_y.

According to Theorem 1, (12) is an alternate matrix expression of (11):

$$\begin{aligned} J = {} & \operatorname{Tr}\left[P_x^T X F_1(C) X^T P_x + P_y^T Y F_2(C) Y^T P_y \right. \\ & \left. - P_x^T X C Y^T P_y - P_y^T Y C^T X^T P_x\right] \\ = {} & \operatorname{Tr}\left(\begin{bmatrix} P_x \\ P_y \end{bmatrix}^T \begin{bmatrix} X & \\ & Y \end{bmatrix} \begin{bmatrix} F_1(C) & -C \\ -C^T & F_2(C) \end{bmatrix} \begin{bmatrix} X & \\ & Y \end{bmatrix}^T \begin{bmatrix} P_x \\ P_y \end{bmatrix}\right), \end{aligned} \tag{12}$$

where $\operatorname{Tr}(X)$ represent computing the trace of matrix X, $F_1(C)$ and $F_2(C)$ are diagonal matrixes, and their diagonal elements are the row or column sums of similarity matrix C, respectively.

Assuming that $P = \left[\begin{smallmatrix} P_x \\ P_y \end{smallmatrix}\right]$, $Z = \left[\begin{smallmatrix} X & Y \end{smallmatrix}\right]$, $\Gamma = \left[\begin{smallmatrix} F_1(C) & -C \\ -C^T & F_2(C) \end{smallmatrix}\right]$, (12) can be rewritten as follow:

$$J(P) = \operatorname{Tr}\left[P^T Z \Gamma Z^T P\right]. \tag{13}$$

To make the equation have a unique solution, $P^T Z Z^T P = I$ and $P^T Z e = 0$ as constraints are added, where $e = [1, 1, \ldots 1]^T$ is a vector with dimensions of $(N_x + N_y) \times 1$, N_x and N_y are the numbers of samples in collections X and Y. The solution to make (13) minimized is obtained by generalized eigendecomposition of $(Z\Gamma Z^T)p = \lambda(ZZ^T)p$ and taking eigenvectors $p_2, p_3, \ldots p_{m+1}$ corresponding to the second to $(m+1)$th smallest eigenvalues $\lambda_2, \lambda_3, \ldots \lambda_{m+1}$. Assuming that $P = [p_2, p_3, \ldots p_{m+1}]$, its dimension is $(D_x + D_y) \times m$, D_x and D_y are the dimensions of samples in collections X and Y, so the transformation matrix P_x corresponds to the 1st to D_xth rows of P and P_y corresponds to the $(D_x + 1)$th to D_yth rows of P.

Step 7. Bringing the matrix P_x and P_y to (11), the distance metric of the elements belonging to different collections can be realized.

4. The Kernel Coupled Metric Learning Based on Supervised Locality Preserving Projection (SLPP-KCML)

In practical dimension reduction and measurement process, the linear model is not well to represent the features, and it is difficult to map two complex collections to the same space using the linear transformation. So combining the kernel method, we extend the SLPP-CML algorithm; a nonlinear coupled metric learning methods based on the supervised locality preserving projection is proposed.

Assuming that the mapping functions f_x and f_y are nonlinear functions, namely, $f_x = \phi_x(x)$, $f_y = \phi_y(y)$, using the nonlinear mapping $\phi : R^n \to F$, $x \to \phi_x(x)$, $y \to \phi_y(y)$, the sample data can be mapped to the high dimensional Hilbert space. The criterion can be defined by

$$J = \sum_{(i,j)\in C} \left\| W_a^T \phi_x(x_i) - W_a^T \phi_y(y_j) \right\|^2. \tag{14}$$

An alternate matrix expression is as follow:

$$\begin{aligned} J = {} & \operatorname{Tr}\left(W_a^T \Phi_x(x) G_x \Phi_x(x)^T W_a + W_a^T \Phi_y(y) G_y \Phi_y(y)^T W_a \right. \\ & - W_a^T \Phi_x(x) C \Phi_y(y)^T W_a \\ & \left. - W_a^T \Phi_y(y) C^T \Phi_x(x)^T W_a\right) \\ = {} & \operatorname{Tr}\left(\begin{bmatrix} W_a \\ W_a \end{bmatrix}^T \begin{bmatrix} \Phi_x(x) & \\ & \Phi_y(y) \end{bmatrix} \begin{bmatrix} G_x & -C \\ -C^T & G_y \end{bmatrix} \right. \\ & \left. \times \begin{bmatrix} \Phi_x(x) & \\ & \Phi_y(y) \end{bmatrix}^T \begin{bmatrix} W_a \\ W_a \end{bmatrix}\right), \end{aligned} \tag{15}$$

where Tr represent computing the trace of matrix, G_x and G_y are diagonal matrixes, and their diagonal elements are the row or column sums of similarity matrix C, respectively.

Letting $W_a = \Phi_x(x) A_x$, $W_a = \Phi_y(y) A_y$, we can get

$$\begin{aligned} J = \operatorname{Tr}\left(\begin{bmatrix} \Phi_x(x) A_x \\ \Phi_y(y) A_y \end{bmatrix}^T \begin{bmatrix} \Phi_x(x) & \\ & \Phi_y(y) \end{bmatrix} \begin{bmatrix} G_x & -C \\ -C^T & G_y \end{bmatrix} \right. \\ \left. \times \begin{bmatrix} \Phi_x(x) & \\ & \Phi_y(y) \end{bmatrix}^T \begin{bmatrix} \Phi_x(x) A_x \\ \Phi_y(y) A_y \end{bmatrix}\right). \end{aligned} \tag{16}$$

The kernel function $K_{i,j}^x = (\phi(x_i) \cdot \phi(x_j))$, $K_{i,j}^y = (\phi(y_i) \cdot \phi(y_j))$; the kernel matrixes K^x and K^y are real symmetric matrices. Equation (17) is an alternative expression of (16):

$$\begin{aligned} J(A_x, A_y) = \operatorname{Tr}\left(\begin{bmatrix} A_x \\ A_y \end{bmatrix}^T \begin{bmatrix} K^x & \\ & K^y \end{bmatrix} \begin{bmatrix} G_x & -C \\ -C^T & G_y \end{bmatrix} \right. \\ \left. \times \begin{bmatrix} K^x & \\ & K^y \end{bmatrix}^T \begin{bmatrix} A_x \\ A_y \end{bmatrix}\right). \end{aligned} \tag{17}$$

Obviously, the coupled metrics learning in kernel space is a process of calculating the transformation matrix A_x and A_y.

Assuming that $A = \begin{bmatrix} A_x \\ A_y \end{bmatrix}$, $K = \begin{bmatrix} K^x & \\ & K^y \end{bmatrix}$, $\Gamma = \begin{bmatrix} G_x & -C \\ -C^T & G_y \end{bmatrix}$, similar to the SLPP-CML algorithm, solving the optimal solution can be transformed into the generalized eigenvalue problem. The generalized characteristic equation is $Ea = \lambda Fa$, and $E = K\Gamma K^T$, $F = KK^T$; a is the eigenvector corresponding to eigenvaule λ. The eigenvectors corresponding to the minimum to the D_cth smallest eigenvalues construct the feature matrix A; the size of matrix A is $(N_x + N_y) \times D_c$, where N_x and N_y are the numbers of training samples of collections X and Y. Finally, we can get the feature matrix A_x corresponding to the data matrix X and the feature matrix A_y corresponding to the data matrix Y.

In addition, the samples mapped to the high dimensional space need centering processing. In the linear coupled metric learning, the centering can be realized by abandoning the eigenvector corresponding to eigenvalue of "zero." However, the centering of nonlinear coupled metric learning in kernel space can be realized by centering the kernel matrix K^x and K^y

$$\widetilde{K} = K - \frac{1}{n}K1_{nn} - \frac{1}{n}1_{nn}K + \frac{1}{n^2}1_{nn}K1_{nn}, \tag{18}$$

where n is the dimension of kernel matrix K. 1_{nn} is a matrix with size of $n \times n$ and all elements are one.

5. Experiment and Analysis

5.1. Introduction of the Face Database. The proposed coupled metric learning approach is used for face recognition. It is tested on Yale face database [23] and CAS-PEAL-R1 face database [24]. The Yale face database contains 165 pictures of 15 people with the size of 100×100 and 256 gray levels. These images were taken in different expression and illumination conditions. In experiment, we used the former 6 images per person as training samples, a total of 90, and the other images were used as test samples.

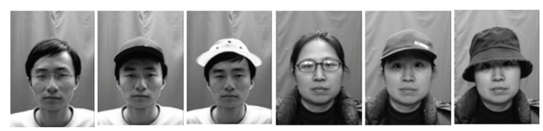
The CAS-PEAL-R1 face database contains 30863 face images, which was divided into two parts: (1) the frontal face image subset; (2) the nonfrontal face image subset. In the experiment, we used the accessory data set of the frontal face image subsets (CAS-PEAL-R1-FRONTAL-Accessory). The face images per person in CAS-PEAL-R1-FRONTAL-Accessory contain 6 different appendages; there are 3 images with different glasses and 3 images with different hats. We selected 300 images of 50 people with the size of 360×480 and 256 gray levels in the experiment; the odd-numbered images were used as training samples and even-numbered images were used as test samples, respectively. Some training images are shown in Figure 2 and some test images are shown in Figure 3.

(a) Some images in Yale face database

(b) Some images in CAS-PEAL-R1-FRONTAL-Accessory

Figure 2: Some original training face images.

(a) Some images in Yale face database



(b) Some images in CAS-PEAL-R1-FRONTAL-Accessory

Figure 3: Some original test face images.

5.2. The Low-Resolution Face Recognition. Due to the differences between the different resolution cameras and the uncertainty of distance between camera and face, the resolution of face image that we collected is not uniform. Obviously, traditional measurement method can not be used to calculate the distance between two images with different resolution. The general handling method is interpolation operation, but the interpolation operation is easy to introduce false information. With the increase in false information, the distortion degree increases, as shown in Figure 4. Aiming at the problem of recognition rate declining because of image distortion, the researchers realized the low-resolution image compensation by increasing the image restoration preprocessing. But the image restoration algorithm is more complex, and the quality of image restoration has great impact on final recognition results.

However, the proposed coupled metric learning method can directly realize the feature extraction and measurement of different resolution images. This method not only saves computing time, but also avoids the negative impact of image restoration on recognition performance. To better illustrate the experimental processes, Figure 5 gives the flow of the degraded face recognition.

In experiment, the training samples include clear face and degraded face images. The size of original normal training face image is adjusted to 64×64 pixel, and these adjusted faces are used as clear face images. However, there are not original low-resolution face images in public face database, so we obtained the low-resolution training face image through blurring and sampling original normal training face image, and the size of low-resolution face image is 16×16.

The test samples are the low-resolution face images, which were generated through blurring and sampling original normal test face image introduced in Section 5.1.

5.2.1. Experiment 1: The Low-Resolution Face Recognition Based on SLPP-CML. Through the theoretical analysis, the

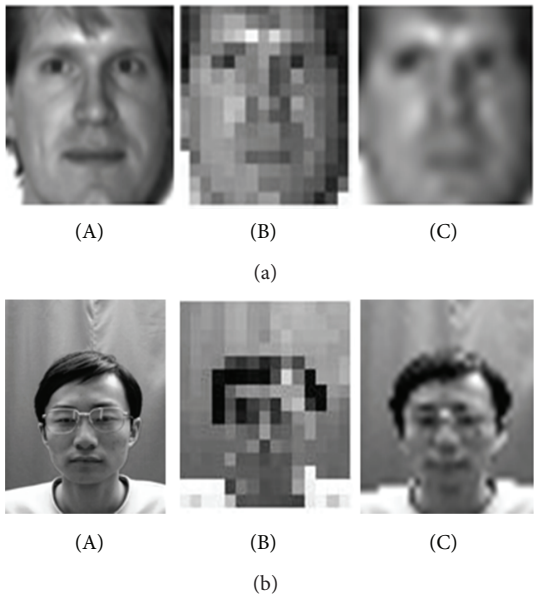


Figure 4: (a) (A), (B), and (C) are the normal face, low-resolution face, and face after restoration, respectively, in Yale face database. (b) (A), (B), and (C) are the normal face, low-resolution face, and face after restoration, respectively, in CAS-PEAL-R1-FRONTAL-Accessory.

SLPP-CML algorithm has two influence factors: (1) the number of neighbors k of supervised locality preserving projection; (2) the reserved dimensions D_c of the feature. Therefore, the recognition results based on different parameters should be discussed and analyzed. Figure 6 shows the change of recognition rate with the change of feature dimensions, when the number of neighbors takes different values.

These recognition rate curves are in two different face databases. These curves have a general change law, with the increase in feature dimensions; the recognition rate kept a decreasing trend after increasing, and the best recognition results can be achieved only in the optimal feature dimensions.

In Yale face database, the recognition rate kept a higher trend when feature dimensions remain 10–20. The optimal recognition rate is 86.67% when feature dimension is 10 and the number of neighbors is 5. In CAS-PEAL-R1 face database, the recognition rate can reach the maximum value 86.67%, when feature dimension is 40 and the number of neighbors is 2.

The experimental data show that the number of training samples of each class is 6 in Yale face database and the recognition effect is optimal when the number of neighbors is 5. In CAS-PEAL-R1 face database, we can obtain optimal recognition rate when the number of training samples of each class is 3 and the number of neighbors is 2. Obviously, the number of neighbors is $T - 1$; T is the number of training samples of each class.

In addition, in order to illustrate the effectiveness of SLPP-CML method, the comparative experiments were carried out. The experiment results are shown in Table 1.

The experimental data illustrated that the recognition results of feature extraction after restoration are not satisfactory. The coupled metric learning in [19] can not overcome the influences of within-class multiple modes, so the identification effect is not good. The coupled metric learning in [21] is conducive to resolving within-class multiple modes; the recognition effects have been greatly improved, but it does not fully consider the between-class relationships of training samples. The proposed SLPP-CML takes advantage of the supervisory of category information, while the within-class and between-class relationship information of training samples have been considered into the metric learning, so we can get better recognition results.

5.2.2. Experiment 2: The Low-Resolution Face Recognition Based on SLPP-KCML. The SLPP-KCML algorithm is a nonlinear coupled metric learning algorithm. Through the analysis, there are three factors which affect this algorithm: (1) the number of neighbors k of supervised locality preserving projection; (2) the reserved dimensions D_c of the feature; (3) the kernel function. Based on the experiment result, the number of the nearest neighbors is the same as that of SLPP-CML. For the kernel function, we choose the Gauss function $k(x, y) = \exp(-\|x - y\|^2/\alpha)$; the value of adjustable factor α affects the function performance. So, in this paper, the experiments were carried out according to the different adjustable factors and the change of feature dimensions on recognition rate; the experimental results are shown in Figure 7.

The curves indicated that, in Yale face database, the optimal recognition rate is 89.33% when the value of α is 0.5 and the feature dimension is 20; compared with the SLPP-CML algorithm, the recognition rate increased by 2.66%. In CAS-PEAL-R1 face database, when $\alpha = 0.7$ and the feature dimensions $D_c = 40$, recognition rate is 91.33%; compared with the SLPP-CML algorithm, the recognition rate increased by 4.66%. Obviously, the nonlinear coupled metric learning method can effectively extract the classification information of face image and obtain a high recognition rate.

Considering the training time, the SLPP-CML algorithm requires calculating the eigenvalue and eigenvector of the image covariance matrix. The resolution of clear face image is 64×64 pixel and low-resolution face image is 16×16 pixel. So the dimensions of image covariance matrix are 4352×4352 and the average training time is about 553.905 seconds.

However, the matrix of SLPP-KCML is related to the number of classes and the number of training samples of each class, so the dimension of the covariance matrix of Yale face database is 90×90, the size of covariance matrix of CAS-PEAL-R1 face database is 150×150, and the average training time is about 6.687 seconds. Obviously, the training speed of the SLPP-KCML algorithm is faster than SLPP-CML algorithm, and the recognition time of these two algorithms is about 0.0225 seconds. Based on the above analysis, the efficiency of SLPP-KCML algorithm is better than SLPP-CML algorithm.

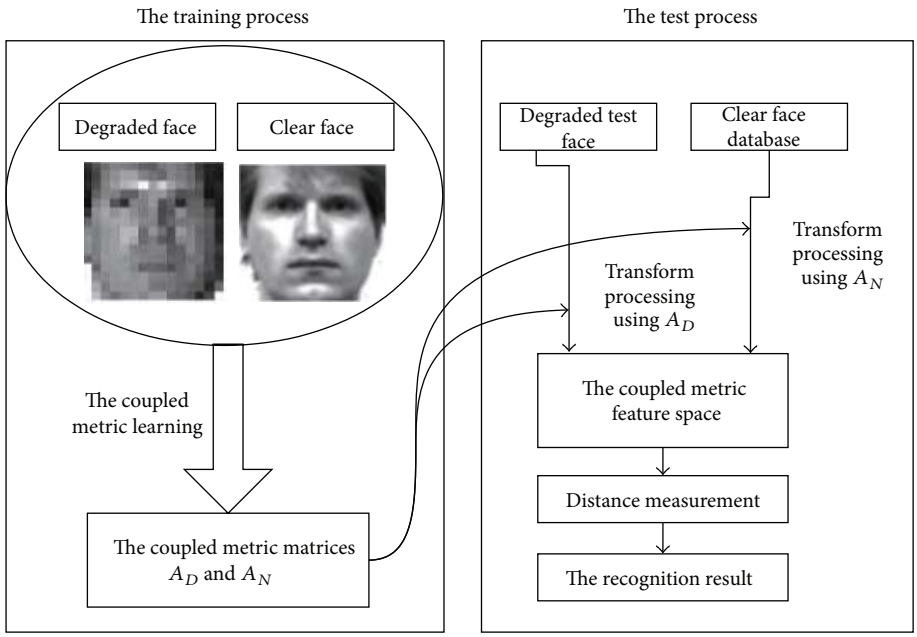
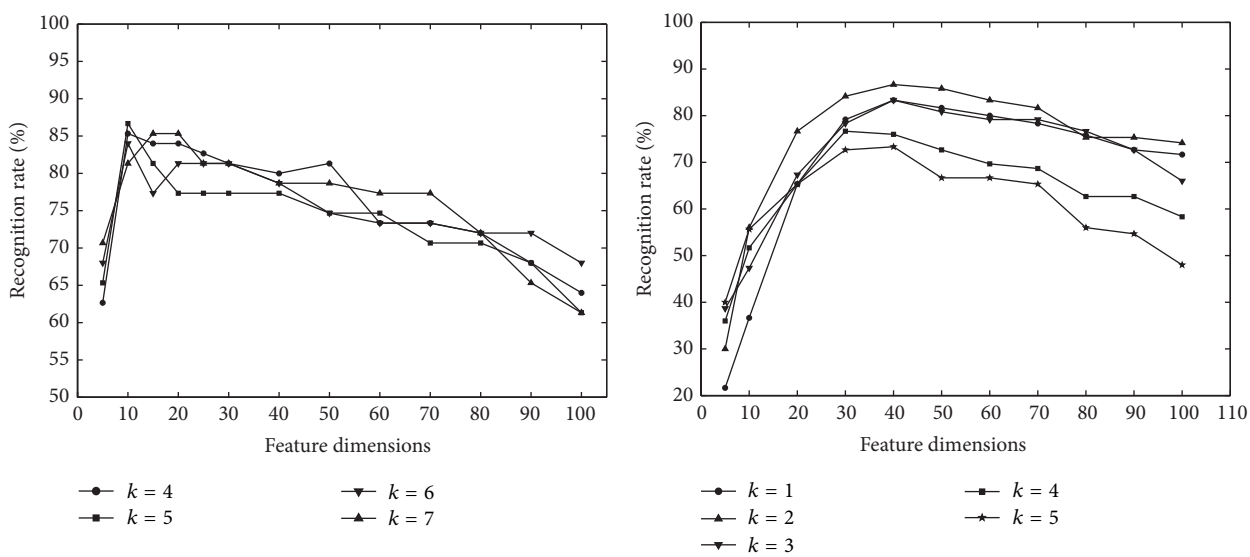


Figure 5: Degraded face recognition process. In training process, we obtained the matrix A_D for degraded face and the matrix A_N for clear face image. In test process, matrices A_D and A_N transformed the degraded test face and clear test face to the coupled metric space, calculating the distance.



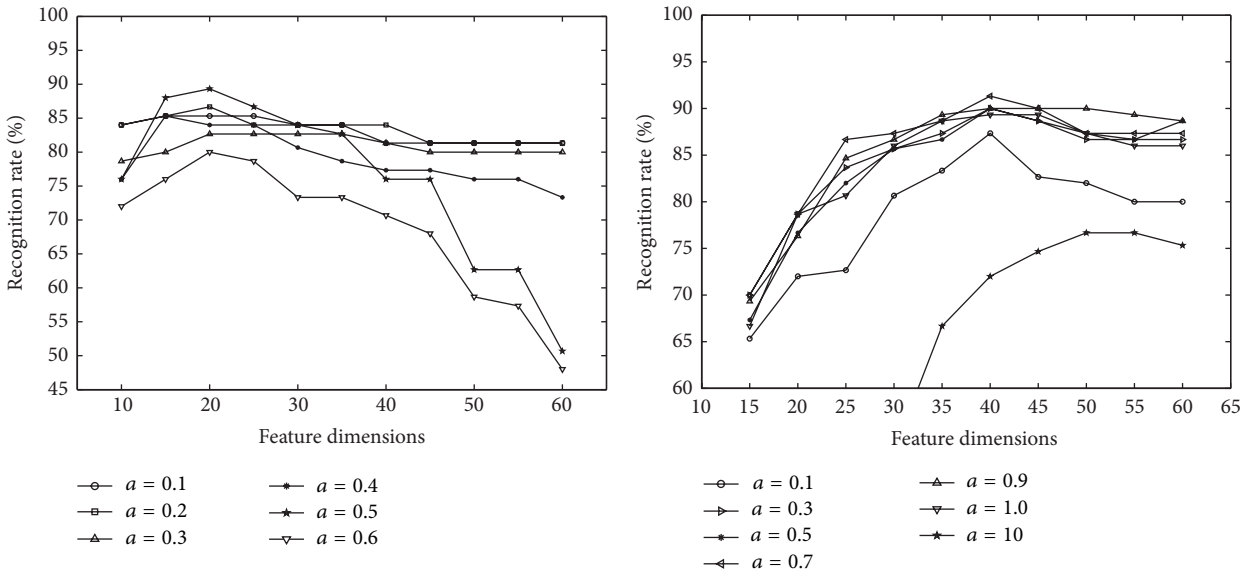
(a) The experiment result in Yale face database

(b) The experiment result in CAS-PEAL-R1 face database

Figure 6: The recognition rate under different dimensions and different numbers of the nearest neighbors.

Table 1: Experimental comparison of this method with other methods.

Method	Yale face database	CAS-PEAL-R1 face database
Image restoration [25] + PCA	61.33	55.33
CML [19]	77.33	74.67
CLPM [21]	82.67	80.67
SLPP-CML	86.67	86.67



(a) The experiment result in Yale face database

(b) The experiment result in Yale CAS-PEAL-R1 face database

Figure 7: The recognition rate under different dimensions and different adjustable factors.

(A) (B) (C) (A) (B) (C)

(a) (b)

Figure 8: (a) (A), (B), and (C) are the normal face, blurred face image, and deblurring results, respectively, in Yale face database. (b) (A), (B), and (C) are the normal face, blurred face image, and deblurring face, respectively, in CAS-PEAL-R1-FRONTAL-Accessory.

5.3. The Fuzzy Face Recognition. Besides the low-resolution face image, the blurring image usually makes the performance of face recognition system decrease. The fuzzy face image is shown in Figure 8. Obviously, it is difficult to identify the fuzzy face image; a part of face details can be restored by deblurring algorithm, but it still cannot provide enough information in identification.

Similar to Section 5.2, we carried out the recognition experiments of SLPP-CML algorithm based on different face databases, and the comparative experiments with other methods were made. In the experiment, the clear images with size of 64×64 are the same as those used in Section 5.1, and the fuzzy face images were generated by convolution of corresponding clear face image. The training samples are composed by clear training face images and generated fuzzy face images. The test samples are the fuzzy face images by convolution of the clear test face images and the size is 64×64 pixels. The number of training and test samples has been introduced in Section 5.1. The experiment results are shown in Table 2.

The experimental data is the best recognition rate of each algorithm; the number of neighbors of SLPP-CML and SLPP-KCML algorithm is $T - 1$, where T is the number of training samples of each class. The optimal value of adjustable factor of Gaussian kernel function in SLPP-KCML algorithm is 0.7. Table 3 gives the feature dimensions of training samples in Yale and CAS-PEAL-R1 face database of SLPP-CML and SLPP-KCML algorithm.

6. Conclusions

Aiming at the problem that the traditional metric methods can not calculate the distance of the elements in different data sets, we proposed the coupled metric learning method based on supervised locality preserving projection. First, the elements of different sets are mapped to the coupled space combined with the within-class and between-class information, and then the metric matrix learning is performed. Furthermore, we extended this algorithm to nonlinear space, and the kernel coupled metric learning

Table 2: Experimental comparison of this method with other methods.

Method	Yale face database	CAS-PEAL-R1 face database
Deblurring [26] + PCA	64.00	68.67
CML [19]	77.33	79.33
CLPM [21]	85.67	83.33
SLPP-CML	86.67	88.67
SLPP-KCML	90.00	92.66

Table 3: The feature dimensions in our proposed algorithm.

Method	Yale face database	CAS-PEAL-R1 face database
SLPP-CML	15	39
SLPP-KCML	18	40

method based on supervised locality preserving projection is proposed. In kernel coupled metric learning approach, two elements of different collections are mapped to the unified high dimensional feature space by kernel function, and then the traditional metric learning is performed in this space. In order to verify the effectiveness of the proposed algorithm, we have done a lot of experiments on two face databases. This algorithm can effectively extract the face nonlinear features, and the operation is simple. Low-resolution and fuzzy face recognition experiments show that the proposed method can obtain a higher recognition rate and has a high computational efficiency.

Appendix

Proof of Theorem 1

(1) Assuming $A \in R^{n\times n}$, the Frobenius norm of matrix A is $\|A\| = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}$. So

$$\|A\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left|a_{ij}\right|^2,$$

$$\begin{aligned} A^TA &= \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^{n} a_{i1}^2 & \sum_{i=1}^{n} a_{i1}a_{i2} & \cdots & \sum_{i=1}^{n} a_{i1}a_{in} \\ \sum_{i=1}^{n} a_{i2}a_{i1} & \sum_{i=1}^{n} a_{i2}^2 & \cdots & \sum_{i=1}^{n} a_{i2}a_{in} \\ \vdots & & \ddots & \vdots \\ \sum_{i=1}^{n} a_{in}a_{i1} & & \cdots & \sum_{i=1}^{n} a_{in}^2 \end{bmatrix}. \end{aligned} \tag{A.1}$$

Obviously, $\mathrm{Tr}(A^TA) = \sum_{i=1}^{n} a_{i1}^2 + \sum_{i=1}^{n} a_{i2}^2 + \cdots + \sum_{i=1}^{n} a_{in}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2$.

Therefore, $\|A\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2 = \mathrm{Tr}(A^TA)$.

According to the properties of the trace of a matrix, if $A \in R^{n\times n}$, the eigenvalue of matrix A is $\lambda_1, \lambda_2, \ldots, \lambda_n$; then; $\mathrm{Tr}(A) = \sum_{i=1}^{n}\lambda_i$.

Obviously, $A \in R^{n\times n}$; then, $A^T \in R^{n\times n}$, $A^TA \in R^{n\times n}$; then, $\mathrm{Tr}(A^TA) = \sum_{i=1}^{n}\lambda_i(A^TA)$, where $\lambda_i(A^TA)$ is the eigenvalue of matrix A^TA.

Based on the above analysis, we can get that $\|A\|^2 = \mathrm{Tr}(A^TA) = \sum_{i=1}^{n}\lambda_i(A^TA)$.

(2) Assuming $A \in R^{n\times m}, B \in R^{m\times n}$, then $AB \in R^{n\times n}$ and $BA \in R^{m\times m}$.

According to the definition of the trace, we get that $\mathrm{Tr}(AB) = \sum_{i=1}^{n}(AB)_{ii}$. And then, $\mathrm{Tr}(AB) = \sum_{i=1}^{n}(AB)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{m}A_{ij}B_{ji} = \sum_{j=1}^{m}\sum_{i=1}^{n}B_{ji}A_{ij} = \sum_{j=1}^{m}(BA)_{jj} = \mathrm{Tr}(BA)$. The above equation shows that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$; then, $\mathrm{Tr}(A^TA) = \mathrm{Tr}(AA^T)$.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research is supported by National Research Foundation for the Doctoral Program of Higher Education of China (20102304110004).

References

[1] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, pp. 209–216, ACM Press, Corvallis, Ore, USA, June 2007.

[2] S. Xiang, F. Nie, and C. Zhang, "Learning a Mahalanobis distance metric for data clustering and classification," *Pattern Recognition*, vol. 41, no. 12, pp. 3600–3612, 2008.

[3] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[4] B. Kulis, P. Jain, and K. Grauman, "Fast similarity search for learned metrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2143–2157, 2009.

[5] H. Zhengping, L. liang, X. Chengqian et al., "Sparse distance metric learning with L1-norm constraint for one-class samples in high-dimensional space and its application," *Mathematics in Practice and Theory*, vol. 41, no. 6, pp. 116–124, 2011.

[6] L. Wang, T. Liu, and H. Jia, "Chunk incremental distance metric learning algorithm based on manifold regularization," *Acta Electronica Sinica*, vol. 39, no. 5, pp. 1131–1135, 2011.

[7] M. S. Baghshah and S. B. Shouraki, "Kernel-based metric learning for semi-supervised clustering," *Neurocomputing*, vol. 73, no. 7-9, pp. 1352–1361, 2010.

[8] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.

[9] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

[10] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[11] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[12] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[13] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing System*, vol. 14, pp. 585–591, 2002.

[14] M. S. Baghshah and S. B. Shouraki, "Non-linear metric learning using pairwise similarity and dissimilarity constraints and the geometrical structure of data," *Pattern Recognition*, vol. 43, no. 8, pp. 2982–2992, 2010.

[15] B. Babagholami-Mohamadabadi, A. Zarghami, H. A. Pourhaghighi, and M. T. Manzuri-Shalmani, "Probabilistic non-linear distance metric learning for constrained clustering," in *Proceedings of the 4th MultiClust Workshop on Multiple Clusterings, Multi-View Data, and Multi-Source Knowledge-Driven Clustering (MultiClust '13)*, no. 4, ACM, New York, NY, USA, 2013.

[16] X. Cai, C. Wang, B. Xiao, X. Chen, and J. Zhou, "Deep nonlinear metric learning with independent subspace analysis for face verification," in *Proceedings of the 20th ACM International Conference on Multimedia*, pp. 749–752, ACM, New York, NY, USA, 2012.

[17] M. S. Baghshah and S. B. Shouraki, "Learning low-rank kernel matrices for constrained clustering," *Neurocomputing*, vol. 74, no. 12-13, pp. 2201–2211, 2011.

[18] S. H. Liu, J. Y. Zhang, J. Xu, and H. E. Jia, "Kernel-kNN: a new kNN algorithm based on informational energy metric," *Acta Automatica Sinica*, vol. 36, no. 12, pp. 1681–1688, 2010.

[19] B. Li, H. Chang, S. Shan, and X. Chen, "Coupled metric learning for face recognition with degraded images," in *Advances in Machine Learning: 1st Asian Conference on Machine Learning (ACML '09)*, vol. 5828 of *Lecture Notes in Computer Science*, pp. 220–233, Springer, Halmstad, Sweden, 2009.

[20] X. Ben, W. Meng, and R. Yan, "Dual-ellipse fitting approach for robust gait periodicity detection," *Neurocomputing*, vol. 79, pp. 173–178, 2012.

[21] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 20–23, 2010.

[22] Z. Shen, Y. Pan, and S. Wang, "A supervised locality preserving projection algorithm for dimensionality reduction," *Pattern Recognition and Artificial Intelligence*, vol. 21, no. 2, pp. 233–239, 2008.

[23] The Yale face Database, http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html.

[24] The CAS-PEAL-R1 Face Database, http://www.ict.ac.cn/jszy/jsxk_zlxk/jsxk/200707/t20070706_2179538.html.

[25] Y. Wang, Z. Sun, Z. Cai, and W. Wang, "Super-resolution image restoration based on maximum-likelihood estimation," *Chinese Journal of Scientific Instrument*, vol. 29, no. 5, pp. 949–953, 2008.

[26] Y. Xiaofei, C. Fujie, and Y. Xuejun, "Template matching by wiener filtering," *Journal of Computer Research and Development*, vol. 37, no. 12, pp. 499–1503, 2000.