

## Research Article

# A Novel Multisupervised Coupled Metric Learning for Low-Resolution Face Matching

Guofeng Zou <sup>1,2</sup>, Guixia Fu <sup>1</sup> and Xiang Peng<sup>2</sup>

<sup>1</sup>College of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255049, China

<sup>2</sup>School of Engineering, The University of British Columbia's Okanagan Campus, Kelowna V1V 1V7, Canada

Correspondence should be addressed to Guofeng Zou; [zgf841122@163.com](mailto:zgf841122@163.com)

Received 28 August 2019; Accepted 12 February 2020; Published 16 March 2020

Academic Editor: Constantine Kotropoulos

Copyright © 2020 Guofeng Zou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a new multisupervised coupled metric learning (MS-CML) method for low-resolution face image matching. While coupled metric learning has achieved good performance in degraded face recognition, most existing coupled metric learning methods only adopt the category label as supervision, which easily leads to changes in the distribution of samples in the coupled space. And the accuracy of degraded image matching is seriously influenced by these changes. To address this problem, we propose an MS-CML method to train the linear and nonlinear metric model, respectively, which can project the different resolution face pairs into the same latent feature space, under which the distance of each positive pair is reduced and that of each negative pair is enlarged. In this work, we defined a novel multisupervised objective function, which consists of a main objective function and an auxiliary objective function. The supervised information of the main objective function is the category label, which plays a major supervisory role. The supervised information of the auxiliary objective function is the distribution relationship of the samples, which plays an auxiliary supervisory role. Under the supervision of category label and distribution information, the learned model can better deal with the intraclass multimodal problem, and the features obtained in the coupled space are more easily matched correctly. Experimental results on three different face datasets validate the efficacy of the proposed method.

## 1. Introduction

Image matching is an important task in computer vision and multimedia analysis, and numerous methods have been proposed to solve this issue under controlled conditions [1–4]. However, these methods often fail to gain a good performance in real-world degraded image matching. To address the problem of low resolution, some researchers usually improve image quality by image reconstruction [5–8]. However, image reconstruction methods usually require big data to train the model, which is complicated and time-consuming. More importantly, some invalid information is easily introduced into the reconstructed image, and sometimes the information is even interference for image matching. Therefore, to avoid invalid information interference in image reconstruction, some researchers proposed the image matching idea based on coupled metric learning.

In 2009, the linear coupled metric learning (LCML) [9–11] was proposed, which achieved good performance in low-resolution image matching. In order to better implement nonlinear data metric, some nonlinear coupled metric learning methods [12–14] combining the kernel technique were proposed. In addition, Zhang et al. [15] proposed the distance metric based on coupled edge discriminant mapping, which implemented the effective metric for different attribute data. Jiang et al. [16, 17] proposed coupled discriminant manifold analysis for low-resolution face recognition, which effectively improved the recognition performance. Since 2014, deep learning has become more and more widespread, which is introduced to improve the metric learning model [18, 19]. The deep coupled metric method maps samples to a coupled space by using deep networks, which can better extract nonlinear features and improve the performance of image matching. But deep networks require huge datasets for training and cannot be applied in small datasets.

Although there have been many studies on coupled metric learning, the existing metric methods still have some shortcomings. When only category labels are used as supervision, it is difficult to overcome the intraclass multi-modal problem, as shown in Figure 1; the distribution of samples cannot be well maintained in the coupled space, which will result in a decrease of image-matching accuracy. Therefore, to better use the spatial distribution information of samples to supervise the metric learning, this paper proposes a multisupervised coupled metric learning method fusing category label and distribution information of samples. Meanwhile, the proposed method is suitable for small sample datasets and avoids the deficiencies of deep coupled metric method.

In the proposed metric learning method, we first construct the category correlation matrix and distribution correlation matrix, respectively. Under the supervision of these two correlation matrices, the main objective function and auxiliary objective function are defined. Then, the general objective function is generated by fusing these two objective functions and the new correlation matrix fusing category label and distribution information is obtained. In the general objective function, the category label is used as main supervised information and distribution information is used as auxiliary supervised information. Finally, the objective function is transformed into the generalized eigenvalue problem to solve. The experiments on Yale-B, ORL, and UMIST face datasets demonstrate that the multisupervised coupled metric extends the distance metric methods and effectively improves the image matching performance. To sum up, there are two main contributions:

- (1) We propose a multisupervised coupled metric learning method fusing category label and distribution relationship, which can overcome the defect of existing coupled metric methods. The method can map data in different spaces to the coupled space, and at the same time, it can better maintain the original distribution relationship between samples in the new space.
- (2) We construct the linear and nonlinear multisupervised coupled metric learning objective function, respectively. The nonlinear coupled metric learning can better map sample in different spaces to the coupled space for more reliable image matching.

The rest of this paper is organized as follows. The definition and objective functions of coupled metric learning are described in Section 2. The multisupervised coupled metric learning method is described in detail in Section 3. The performance evaluation experiments based on three different face datasets are implemented in Section 4. Finally, Section 5 concludes our work.

## 2. Preliminary

Suppose that  $I_X$  and  $I_Y$  are two images with the same dimension. Conventionally, the image should be converted into a vector before distance measurement. Assuming that the vector forms are  $x$  and  $y$ , the distance between  $I_X$  and  $I_Y$  is defined as follows:

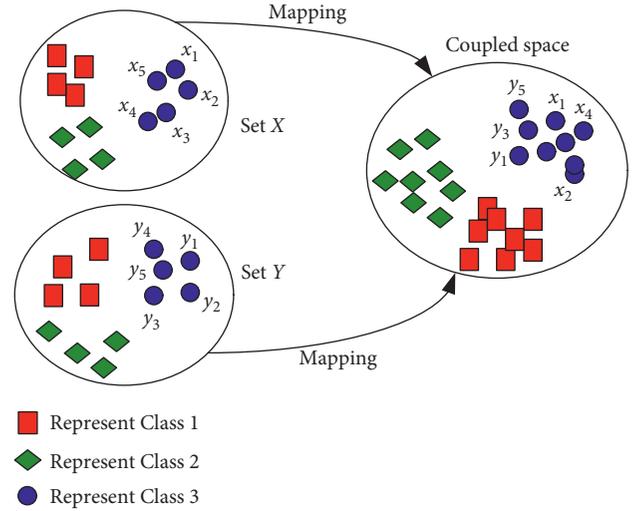


FIGURE 1: The coupled metric mapping supervised by category labels. There are thirteen samples in set X and set Y, respectively; all these samples belong to three categories, denoted as square, rhombus, and circle. In set X, sample  $x_1$  is close to sample  $x_2$ , away from sample  $x_4$ . However, sample  $x_1$  is close to sample  $x_4$ , away from sample  $x_2$  in the coupled space. Although the samples within the same category are grouped together, the distribution relationship between the samples is not maintained, which sometimes affects the image-matching accuracy.

$$d = \|x - y\|_A = \sqrt{(x - y)^T A (x - y)}, \quad (1)$$

where  $\|\cdot\|_A$  denotes norm operation and  $A$  is the distance metric matrix which is positive semidefinite matrix. It can be denoted as  $A = P^T P$ ; then,

$$d = \sqrt{(Px - Py)^T (Px - Py)}. \quad (2)$$

Obviously, the distance metric is converted to norm calculation after feature transformation. Therefore, the unified description of distance metric is as follows:

$$d(x, y) = d_p(Px, Py). \quad (3)$$

**2.1. Definition of Coupled Metric Learning.** The coupled metric is a distance function for samples from different datasets. Assuming that the two images  $I_X$  and  $I_Y$  have different resolutions, the corresponding vectors  $x \in R^{d_x}$  and  $y \in R^{d_y}$ . The coupled distance metric is defined as follows:

$$\begin{aligned} d(x, y) &= d_A(\tilde{x}, \tilde{y}) = d_A(f(x), g(y)) = \|f(x) - g(y)\|_A \\ &= \sqrt{(f(x), g(y))^T A (f(x) - g(y))} \\ &= \sqrt{(f(x) - g(y))^T P P^T (f(x) - g(y))} \\ &= \sqrt{(P^T f(x) - P^T g(y))^T (P^T f(x) - P^T g(y))} \\ &= \|P^T (f(x) - g(y))\|, \end{aligned} \quad (4)$$

where  $f(\cdot)$  and  $g(\cdot)$  are the transformation functions to map samples into coupled space and the matrix  $P$  denotes the transformation and distance measurement in the coupled space.

After transformation, the samples with the same category label should be as close as possible in the coupled space. Therefore, the optimal objective function of coupled metric learning is written as

$$\min J = \min \sum_{(i,j) \in S} \left\| P^T (f(x_i) - g(y_j)) \right\|^2, \quad (5)$$

where  $S$  is a category correlation matrix between datasets  $X$  and  $Y$ , which is defined as follows: assuming that sample  $x_i \in X$ , its category label is  $c_{x_i}$ ; sample  $y_j \in Y$  and its category label is  $c_{y_j}$ . The element in the matrix  $S$  is defined as follows:

$$S_{ij} = \begin{cases} 1, & \text{if } c_{x_i} = c_{y_j}, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Obviously, based on different supervised information, we can obtain different matrix  $S$ , so as to achieve different coupled metric approaches. Additionally, the mapping functions  $f(\cdot)$  and  $g(\cdot)$  may be a linear or nonlinear

function, i.e., corresponding to linear or nonlinear coupled metric.

**2.2. Linear Coupled Metric Learning (LCML).** Assume that the datasets  $X \in R^{d_x}$  and  $Y \in R^{d_y}$  contain  $N$  samples, respectively.  $S \in R^{N \times N}$  is the correlation matrix between  $X$  and  $Y$ , and  $S_{ij}$  is the elements in matrix  $S$ . The optimal objective function of coupled metric learning is as follows:

$$\begin{aligned} J &= \sum_{(i,j) \in S} \left\| P^T (f(x_i) - g(y_j)) \right\|^2 \\ &= \sum_{i=1}^N \sum_{j=1}^N \left\| P^T f(x_i) - P^T g(y_j) \right\|^2 \cdot S_{ij}, \end{aligned} \quad (7)$$

where  $x_i \in X$  and  $y_j \in Y$ .

Assume that  $f(\cdot)$  and  $g(\cdot)$  are linear parameterized functions, which are defined as  $f(x_i) = W_x^T x_i$ ,  $g(y_j) = W_y^T y_j$ .

Let  $P_x = W_x P$  and  $P_y = W_y P$ , (7) is changed to

$$J(P_x, P_y) = \sum_{i=1}^N \sum_{j=1}^N \left\| P_x^T x_i - P_y^T y_j \right\|^2 \cdot S_{ij}. \quad (8)$$

After simplification, (8) can be rewritten as

$$\begin{aligned} J(P_x, P_y) &= \text{tr} \left[ P_x^T X S_x X^T P_x + P_y^T Y S_y Y^T P_y - P_x^T X S Y^T P_y - P_y^T Y S^T X^T P_x \right] \\ &= \text{tr} \left( \begin{bmatrix} P_x \\ P_y \end{bmatrix}^T \begin{bmatrix} X & Y \end{bmatrix} \begin{bmatrix} S_x & -S \\ -S^T & S_y \end{bmatrix} \begin{bmatrix} X & Y \end{bmatrix}^T \begin{bmatrix} P_x \\ P_y \end{bmatrix} \right), \end{aligned} \quad (9)$$

where  $S_x$  and  $S_y$  are diagonal matrices, the diagonal elements in  $S_x$  are the cumulative sum of the corresponding row of matrix  $S$ , and the diagonal elements in  $S_y$  are the cumulative sum of the corresponding column of matrix  $S$ .

**2.3. Nonlinear Coupled Metric Learning (NCML).** Assume that  $f(\cdot)$  and  $g(\cdot)$  are nonlinear functions; that is,  $f(x_i) = \phi_x(x_i)$ ,  $g(y_j) = \phi_y(y_j)$ . The samples  $x_1, x_2, \dots, x_N$  and  $y_1, y_2, \dots, y_N$  will be transformed to a high-dimensional feature space through the mapping function  $\phi: R^n \rightarrow F$ . We can get

the feature representation mode  $\phi_x(x_1), \phi_x(x_2), \dots, \phi_x(x_N)$ , and  $\phi_y(y_1), \phi_y(y_2), \dots, \phi_y(y_N)$ ; then (7) can be rewritten as

$$J = \sum_{i=1}^N \sum_{j=1}^N \left\| P^T \phi_x(x_i) - P^T \phi_y(y_j) \right\|^2 \cdot S_{ij}. \quad (10)$$

Let  $\Phi_x(x) = [\phi_x(x_1), \phi_x(x_2), \dots, \phi_x(x_N)]$  and  $\Phi_y(y) = [\phi_y(y_1), \phi_y(y_2), \dots, \phi_y(y_N)]$ . Based on the kernel space theory, the matrix  $P$  can be represented by  $\Phi_x(x)$  or  $\Phi_y(y)$ ; that is,  $P = \Phi_x(x)\alpha$  or  $P = \Phi_y(y)\beta$ , where  $\alpha$  and  $\beta$  are the coefficient matrices. Then, (10) can be rewritten as follows:

$$J(\alpha, \beta) = \text{tr} \left( \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^T \begin{bmatrix} \Phi_x(x)^T \\ \Phi_y(y)^T \end{bmatrix} \begin{bmatrix} \Phi_x(x) \\ \Phi_y(y) \end{bmatrix} \begin{bmatrix} H_x & -S \\ -S^T & H_y \end{bmatrix} \times \begin{bmatrix} \Phi_x(x) \\ \Phi_y(y) \end{bmatrix} \begin{bmatrix} \Phi_x(x) \\ \Phi_y(y) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right), \quad (11)$$

where  $H_x$  and  $H_y$  are diagonal matrices, and the diagonal elements in these two matrices are cumulative sum of the corresponding row of matrix  $S$  and the cumulative sum of the corresponding column of matrix  $S$ . Constructing the inner product relationship by kernel function, (11) is changed to

$$J(\alpha, \beta) = \text{tr} \left( \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^T \begin{bmatrix} K^x \\ K^y \end{bmatrix} \begin{bmatrix} H_x & -S \\ -S^T & H_y \end{bmatrix} \begin{bmatrix} K^x \\ K^y \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right), \quad (12)$$

where matrices  $K^x$  and  $K^y$  based on the Gaussian kernel  $K(x_i, x_j) = \exp(-(\|x_i - x_j\|^2/2\sigma^2))$  are defined as follows:

$$\begin{aligned}
K^x &= \begin{bmatrix} \langle \phi_x(x_1) \cdot \phi_x(x_1) \rangle & \langle \phi_x(x_1) \cdot \phi_x(x_2) \rangle & \cdots & \langle \phi_x(x_1) \cdot \phi_x(x_N) \rangle \\ \langle \phi_x(x_2) \cdot \phi_x(x_1) \rangle & \langle \phi_x(x_2) \cdot \phi_x(x_2) \rangle & \cdots & \langle \phi_x(x_2) \cdot \phi_x(x_N) \rangle \\ \cdots & \cdots & \ddots & \cdots \\ \langle \phi_x(x_N) \cdot \phi_x(x_1) \rangle & \langle \phi_x(x_N) \cdot \phi_x(x_2) \rangle & \cdots & \langle \phi_x(x_N) \cdot \phi_x(x_N) \rangle \end{bmatrix}, \\
K^y &= \begin{bmatrix} \langle \phi_y(y_1) \cdot \phi_y(y_1) \rangle & \langle \phi_y(y_1) \cdot \phi_y(y_2) \rangle & \cdots & \langle \phi_y(y_1) \cdot \phi_y(y_N) \rangle \\ \langle \phi_y(y_2) \cdot \phi_y(y_1) \rangle & \langle \phi_y(y_2) \cdot \phi_y(y_2) \rangle & \cdots & \langle \phi_y(y_2) \cdot \phi_y(y_N) \rangle \\ \cdots & \cdots & \ddots & \cdots \\ \langle \phi_y(y_N) \cdot \phi_y(y_1) \rangle & \langle \phi_y(y_N) \cdot \phi_y(y_2) \rangle & \cdots & \langle \phi_y(y_N) \cdot \phi_y(y_N) \rangle \end{bmatrix}.
\end{aligned} \tag{13}$$

### 3. The Multisupervised Coupled Metric Learning Method

**3.1. The Class Relation Matrix and Distribution Relation Matrix.** Through the analysis in Section 2, it is known that the matrix  $S$  reflects the correlativity between sets  $X$  and  $Y$ , which plays a decisive role in metric learning. In this work, we fuse category label and distribution information of samples to construct a novel correlation matrix  $S$ , where the category label supervises the distance metric between two datasets and the distribution information supervises the distance metric inside the same dataset. The correlation matrices are constructed based on the category label and distribution information, respectively.

**3.1.1. Category Correlation Matrix.** Supposing that sample  $x_i \in X$ , its class label is  $c_{x_i}$ . Sample  $y_j \in Y$  and its class label is  $c_{y_j}$ . The category correlation matrix is  $C$ ; the element in this matrix is defined as follows:

$$C_{ij} = \begin{cases} 1, & \text{if } c_{x_i} = c_{y_j}, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

**3.1.2. Distribution Correlation Matrix.** Assuming that in set  $X$ , sample  $x_i$  is the  $k$ -nearest neighbor of sample  $x_j$ , we connect samples  $x_j$  and  $x_i$  to each other. Similarly, the same operation is also implemented in set  $Y$ . Let  $T_x$  and  $T_y$  be the distribution correlation matrix for set  $X$  and set  $Y$ , respectively, and the elements in  $T_x$  and  $T_y$  can be calculated as follows:

$$T_{xij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{t}\right), & \text{if } x_i \text{ connected } x_j, \\ 0, & \text{otherwise,} \end{cases} \tag{15}$$

$$T_{yij} = \begin{cases} \exp\left(-\frac{\|y_i - y_j\|^2}{t}\right), & \text{if } y_i \text{ connected } y_j, \\ 0, & \text{otherwise,} \end{cases} \tag{16}$$

where the parameter  $t$  is the average value of distance between all samples.

**3.2. The Linear Multisupervised Coupled Metric Learning (LMS-CML).** In the linear situation, in order to realize the joint supervision of category label and distribution information for coupled metric learning, we need to fuse the above correlation matrices. Therefore, two coupled metric objective functions need to be constructed: (1) category-label-based coupled metric objective function and (2) distribution-information-based coupled metric objective function. Then, the fusion of multisupervised information is realized through the fusion of these two objective functions.

To achieve the first objective function, combining (9) and (14), we construct the linear optimal objective function based on the category correlation matrix  $C$ :

$$J_L(P_x, P_y) = \text{tr} \left( \begin{bmatrix} P_x \\ P_y \end{bmatrix}^T \begin{bmatrix} X & \\ & Y \end{bmatrix} \begin{bmatrix} C_x & -C \\ -C^T & C_y \end{bmatrix} \begin{bmatrix} X & \\ & Y \end{bmatrix} \begin{bmatrix} P_x \\ P_y \end{bmatrix} \right), \tag{17}$$

where  $P_x$  and  $P_y$  are the coupled transformation matrices; the diagonal matrices  $C_x$  and  $C_y$  are similar to  $S_x$  and  $S_y$  in (9).

To achieve the second objective function, combining (9), (15), and (16), we construct the auxiliary optimal objective function for set  $X$  and set  $Y$  in the linear case, as follows:

$$J_{LX}(P_x, P_x) = \text{tr} \left( \begin{bmatrix} P_x \\ P_x \end{bmatrix}^T \begin{bmatrix} X & \\ & X \end{bmatrix} \begin{bmatrix} T_{xx} & -T_x \\ -T_x^T & T_{xx} \end{bmatrix} \begin{bmatrix} X & \\ & X \end{bmatrix} \begin{bmatrix} P_x \\ P_x \end{bmatrix} \right), \tag{18}$$

$$J_{LY}(P_y, P_y) = \text{tr} \left( \begin{bmatrix} P_y \\ P_y \end{bmatrix}^T \begin{bmatrix} Y & \\ & Y \end{bmatrix} \begin{bmatrix} T_{yy} & -T_y \\ -T_y^T & T_{yy} \end{bmatrix} \begin{bmatrix} Y & \\ & Y \end{bmatrix} \begin{bmatrix} P_y \\ P_y \end{bmatrix} \right), \tag{19}$$

where  $T_{xx}$  and  $T_{yy}$  are diagonal matrices, and the diagonal elements of these two matrices are the cumulative sum of corresponding row of matrix  $T_x$  and the cumulative sum of corresponding column of matrix  $T_y$ .

Then, we combine (17), (18), and (19) to obtain the linear general objective function:

$$J_{GL}(P_x, P_y) = J_L(P_x, P_y) + \gamma J_{LX}(P_x, P_x) + \eta J_{LY}(P_y, P_y) \\ = \text{tr} \left( \begin{bmatrix} P_x \\ P_y \end{bmatrix}^T \begin{bmatrix} X & \\ & Y \end{bmatrix} \begin{bmatrix} C_x + 2\gamma(T_{xx} - T_x) & -C \\ -C^T & C_y + 2\eta(T_{yy} - T_y) \end{bmatrix} \begin{bmatrix} X & \\ & Y \end{bmatrix}^T \begin{bmatrix} P_x \\ P_y \end{bmatrix} \right), \quad (20)$$

where  $C_x + 2\gamma(T_{xx} - T_x)$  and  $C_y + 2\eta(T_{yy} - T_y)$  are the novel correlation matrices fused category and distribution information. The coefficients  $\gamma$  and  $\eta$  are within [0-1], which are used to control the strength of distribution information to supervise coupled metric learning.

$$\text{Let } A = \begin{bmatrix} P_x & P_y \end{bmatrix}^T, \quad Z = \begin{bmatrix} X & \\ & Y \end{bmatrix}, \quad \Theta = \begin{bmatrix} C_x + 2\gamma(T_{xx} - T_x) & -C \\ -C^T & C_y + 2\eta(T_{yy} - T_y) \end{bmatrix}.$$

The linear objective function (20) is converted into

$$J_{GL}(A) = \text{tr} \left[ A^T Z \Theta Z^T A \right]. \quad (21)$$

The solution of coupled transformation matrix  $A$  is equivalent to the calculation of the following generalized eigenvalue equation:

$$(Z \Theta Z^T) A = \lambda (Z Z^T) A. \quad (22)$$

After obtaining matrix  $A$ , the transformation matrix  $P_x$  consists of the 1st to  $D_x$ -th rows of matrix  $A$ ; its size is  $D_x \times D_c$ , where  $D_c$  is the dimension of coupled space. The

transformation matrix  $P_y$  consists of the  $(D_x + 1)$ -th to  $(D_x + D_y)$ -th rows of matrix  $A$  and its size is  $D_y \times D_c$ .

**3.3. The Nonlinear Multisupervised Coupled Metric Learning (NMS-CML).** To implement the coupled metric of nonlinear data, we exploit the kernel technique to improve the linear general objective function. The nonlinear optimal objective function is constructed based on the category correlation matrix  $C$  and the distribution correlation matrices  $T_x$  and  $T_y$ .

Firstly, we combine (12) and (14) to get the nonlinear main objective function:

$$J_N(\alpha, \beta) = \text{tr} \left( \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^T \begin{bmatrix} K^x & \\ & K^y \end{bmatrix} \begin{bmatrix} H_x & -C \\ -C^T & H_y \end{bmatrix} \begin{bmatrix} K^x & \\ & K^y \end{bmatrix}^T \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right), \quad (23)$$

where  $\alpha$  and  $\beta$  are the coupled transformation matrices; the diagonal matrices  $H_x$  and  $H_y$  are similar to  $H_x$  and  $H_y$  in (12).

Then, combining (12), (15), and (16), the nonlinear auxiliary optimal objective function is constructed as follows:

$$J_{NX}(\alpha, \alpha) = \text{tr} \left( \begin{bmatrix} \alpha \\ \alpha \end{bmatrix}^T \begin{bmatrix} K^x & \\ & K^x \end{bmatrix} \begin{bmatrix} T_{xx} & -T_x \\ -T_x^T & T_{xx} \end{bmatrix} \begin{bmatrix} K^x & \\ & K^x \end{bmatrix}^T \begin{bmatrix} \alpha \\ \alpha \end{bmatrix} \right), \quad (24)$$

$$J_{NY}(\beta, \beta) = \text{tr} \left( \begin{bmatrix} \beta \\ \beta \end{bmatrix}^T \begin{bmatrix} K^y & \\ & K^y \end{bmatrix} \begin{bmatrix} T_{yy} & -T_y \\ -T_y^T & T_{yy} \end{bmatrix} \begin{bmatrix} K^y & \\ & K^y \end{bmatrix}^T \begin{bmatrix} \beta \\ \beta \end{bmatrix} \right), \quad (25)$$

where  $T_{xx}$  and  $T_{yy}$  are diagonal matrices, and the diagonal elements of these two matrices are the cumulative sum of the corresponding row of matrix  $T_x$  and the cumulative sum of the corresponding column of matrix  $T_y$ .

Finally, combining (23), (24), and (25), we get the nonlinear general objective equation:

$$J_{GN}(\alpha, \beta) = J_N(\alpha, \beta) + \gamma J_{NX}(\alpha, \alpha) + \eta J_{NY}(\beta, \beta) \\ = \text{tr} \left( \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^T \begin{bmatrix} K^x & \\ & K^y \end{bmatrix} \begin{bmatrix} H_x + 2\gamma(T_{xx} - T_x) & -C \\ -C^T & H_y + 2\eta(T_{yy} - T_y) \end{bmatrix} \begin{bmatrix} K^x & \\ & K^y \end{bmatrix}^T \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right), \quad (26)$$

where  $H_x + 2\gamma(T_{xx} - T_x)$  and  $H_y + 2\eta(T_{yy} - T_y)$  are the correlation matrices fusing category and distribution information in nonlinear situation. The coefficients  $\gamma$  and  $\eta$  control the supervised intensity of the distribution information.

$$\text{Let } W = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad K = \begin{bmatrix} K^x \\ K^y \end{bmatrix},$$

$$\Omega = \begin{bmatrix} H_x + 2\gamma(T_{xx} - T_x) & -C \\ -C^T & H_y + 2\eta(T_{yy} - T_y) \end{bmatrix}.$$

The solution of optimal objective function is transformed into the solution of the generalized eigenvalue problem:

$$(K\Omega K^T)w = \lambda(KK^T)w, \quad (27)$$

where  $\lambda$  is eigenvalue and  $w$  is the corresponding eigenvector. The matrix  $W \in R^{(2N) \times D_c}$  is constructed by the eigenvectors  $w$  corresponding to the minimum eigenvalue, the  $D_c$ -th eigenvalue. According to the definition of  $W = [\alpha \beta]^T$ , we get  $\alpha \in R^{N \times D_c}$  and  $\beta \in R^{N \times D_c}$ .

**3.4. Complete Steps of Algorithm Implementation.** Assume that there are two training datasets  $X$  and  $Y$ . The elements in set  $X$  are high-dimensional data corresponding to clear images; the elements in set  $Y$  are low-dimensional data corresponding to low-resolution images. Additionally, there are some low-dimensional data as test samples. The implementation steps of the proposed method are as follows.

#### 3.4.1. Training Process

*Step 1.* Based on the supervision of category label, establish the connection relations between set  $X$  and set  $Y$ ; then according to (14), the category relation matrix  $C$  is calculated.

*Step 2.* Establish local neighbor relation inside set  $X$  and set  $Y$ , respectively, and then the corresponding distribution correlation matrices  $T_x$  and  $T_y$  are computed according to (15) and (16).

*Step 3.* Construct main objective function  $J_L$  in (17) or  $J_N$  in (23), where  $J_L$  is the linear objective function and  $J_N$  is the nonlinear objective function.

*Step 4.* Construct the linear auxiliary objective functions  $J_{LX}$  and  $J_{LY}$  in (18) and (19); moreover, construct the nonlinear auxiliary objective functions  $J_{NX}$  and  $J_{NY}$  in (24) and (25).

*Step 5.* According to (20) and (26), the general objective functions in linear and nonlinear cases are constructed, respectively.

*Step 6.* Solve the optimal objective functions. The linear optimal function can be solved according to (22), and then the nonlinear optimal function can be solved according to (27).

Finally, we obtain the coupled metric matrix  $A = [P_x \ P_y]^T$  or  $W = [\alpha\beta]^T$ , and the features of training

samples in set  $X$  are calculated, namely,  $x_{if} = P_x^T x_i$  or  $x_{if} = \alpha^T x_i$ , where  $x_i \in X$ ,  $i = 1, 2, \dots, N$ .

#### 3.4.2. Testing Process

*Step 1.* Test samples are indentified. Let  $I_T$  denote one low-resolution test image. The coupled metric matrix  $P_y$  or  $\beta$  is used to extract the feature of test image; namely,  $I_f = P_y^T I_T$  or  $I_f = \beta^T I_T$ .

*Step 2.* After coupled mapping, the features of training sample and test sample have the same dimension, so the image matching is converted to the comparison of the features  $x_{if}$  and  $I_f$ .

Figure 2 illustrates the complete steps of the proposed method.

## 4. Experiment and Analysis

In this section, we evaluated the linear and nonlinear multisupervised coupled metric methods on recognition tasks involving inconsistent matching issues. The face datasets and experimental settings were briefly described in Section 4.1. The linear and nonlinear coupled metric experiments were implemented in Sections 4.2 and 4.3. Face matching experiments with different resolutions were implemented in Section 4.4. Finally, Section 4.5 shows comparative experiments between our method and other methods. Detailed experiments and results are described as follows.

**4.1. Datasets and Experimental Settings.** To prove the validity of the proposed coupled metric method, the Yale-B [20], ORL [21], and UMIST [22] face datasets were used for the low-resolution image matching experiments. Table 1 briefly shows the situation of these three face datasets.

In the Yale-B face dataset, each person's images are divided into 5 subsets based on the illumination changes. So, we randomly selected half number of images from the 5 subsets with illumination changes as training samples, a total of 2880 faces. The rest of the face images were used as the normal test samples.

In the ORL face dataset, we adopted 5 images of each volunteer as training samples, a total of 200 faces. The remaining 200 faces were used as normal test samples.

In the UMIST face dataset, we randomly selected 18 faces of each person as training samples, a total of 360 faces. The other faces were used as normal test samples.

Images in the above three face datasets are all high-quality samples, and there are not low-resolution faces. To verify the effect of our proposed method in low-resolution image matching, we need to artificially produce some low-quality images. In the experiment, we use the blurring and undersampling methods to produce low-resolution face images. For example, the resolution of clear face images is normalized to  $64 \times 64$ , and the resolution of corresponding low-quality face is  $16 \times 16$ . So, the training dataset includes two sample sets; namely, the high-quality images form the set  $X$  and low-quality images form the set  $Y$ . Then, these two

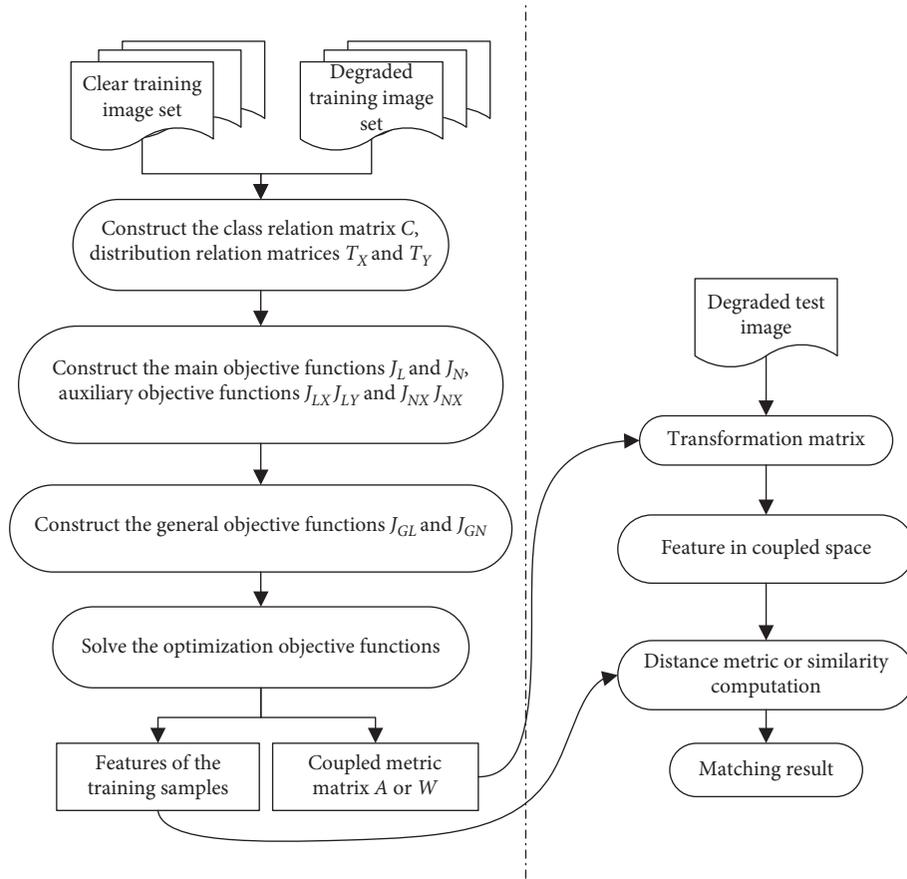


FIGURE 2: Basic steps of the proposed coupled metric learning.

TABLE 1: Introduction to three face datasets.

Face dataset	Dataset size	Number of categories	Resolution	Interference
Yale-B dataset	5760	10	$168 \times 92$	Pose; great illumination
ORL dataset	400	40	$112 \times 92$	Weak illumination and expression; pose; glasses
UMIST dataset	564	20	$92 \times 112$	Great pose variation; weak expression

sets are used to learn the coupled transformation matrices. In testing process, the test images are also low-resolution images, which can be produced by blurring or under-sampling from normal test face images.

In Sections 4.2, 4.3, and 4.5, the resolution of the degraded face image used in the training and testing process is set to  $16 \times 16$ . We repeated our experiments 10 times and took the average as the final matching results.

#### 4.2. Linear Multisupervised Coupled Metric Experiments.

In the proposed LMS-CML, there are three influencing factors: (1) the number  $k$  of nearest neighbors in the calculation of distribution correlation matrix, (2) the dimension of preserving features in solving generalized eigenvalue problems, and (3) coefficients  $\gamma$  and  $\eta$  in general objective function. Therefore, in the following experiments, we will discuss and analyze the impact of these factors on image recognition performance.

*4.2.1. Experiment 1.* Firstly, we analyze the joint influence of the number  $k$  of nearest neighbors and dimension of preserving features. Figure 3 illustrates the experiment results.

It can be seen that with the dimension increasing of preserving features, the recognition rate slowly decreases after a rapid rise. In the initial stage, the feature dimension increases; the effective classification features are more reserved, so the recognition rate rises rapidly. Then, when the feature dimension reaches the optimal value, the best recognition effect can be achieved. After that, the feature dimension continues to increase; the unnecessary interference factors are introduced in the classification feature, so the recognition rate will slowly decrease.

In Yale-B face dataset, when the dimension of preserved features is 15 and the number of nearest neighbors is 5, the highest recognition accuracy reaches 89.06%. In ORL face dataset, the dimension of reserved features is 35; the recognition rate is relatively high. When the number of neighbors is 7, the optimal recognition rate is 93%. In

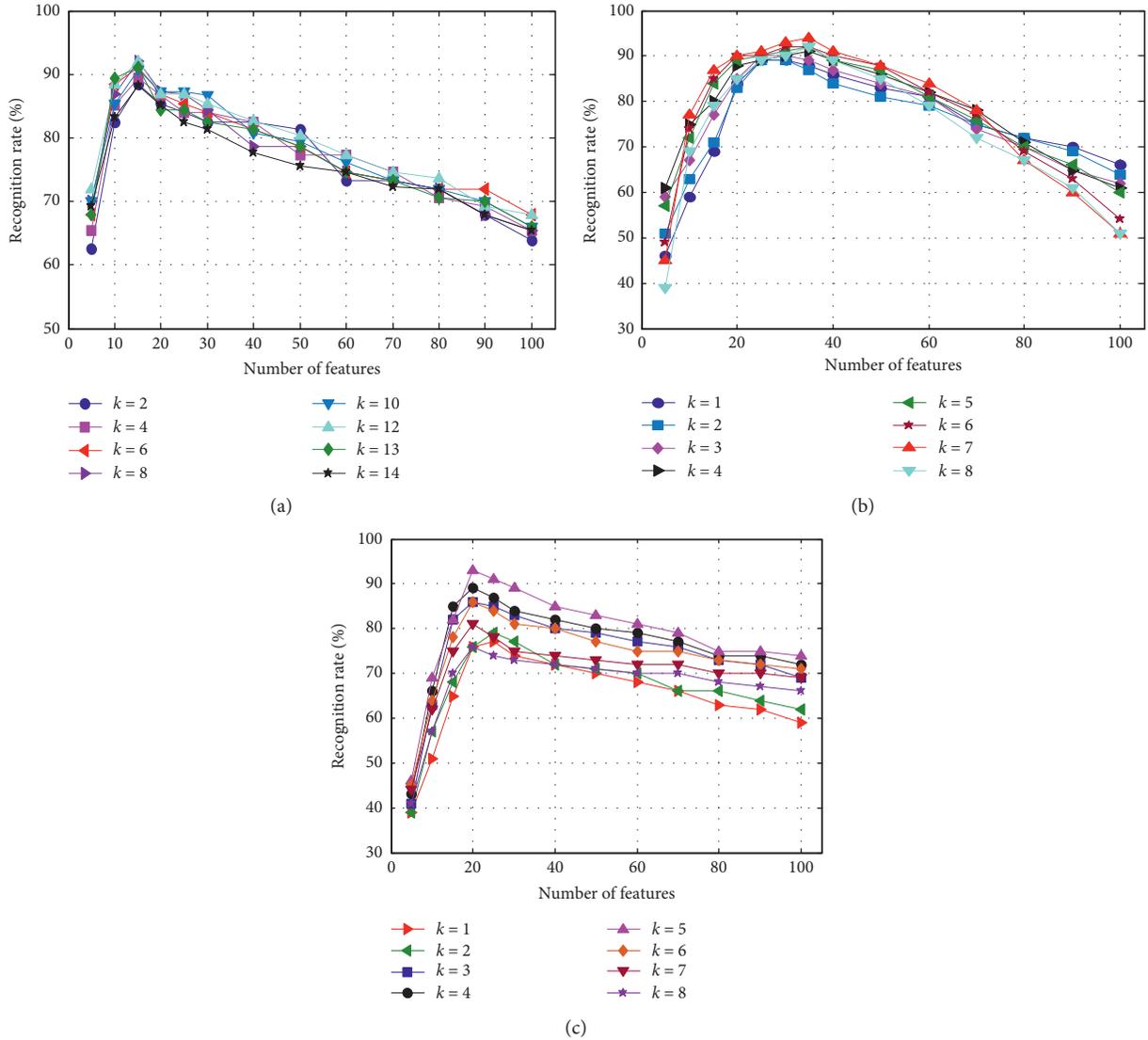


FIGURE 3: Recognition rate with different number of nearest neighbors and reserved features. (a) Experiment result based on Yale-B face database. (b) Experiment result based on ORL face database. (c) Experiment result based on UMIST face database.

UMIST face dataset, due to the great pose variation, it is necessary to reserve more features to obtain a better recognition effect. When the dimension of reserved features is 20 and the number of nearest neighbors is 9, the optimal recognition rate reaches 92.16%. Based on the above experiments, we see that the number of nearest neighbors has some influence on the recognition accuracy, but it does not change the overall trend of recognition rate curves.

**4.2.2. Experiment 2.** In the linear general objective function, the coefficients  $\gamma$  and  $\eta$  control the strength of the supervision of distribution information. So in experiment 2, we analyzed their impact on the recognition effect. Figure 4 shows the experiment results.

Obviously, in different face datasets, the recognition results reach the optimal value when parameters  $\gamma$  and  $\eta$  take different values, but the variation curves of recognition

results have a stronger overall regularity. The optimal value of parameter  $\gamma$  is within [0.6–0.9], and the optimal value of parameter  $\eta$  is within [0.3–0.5]. It can be seen that the impact of distribution information of high-resolution faces is stronger than that of low-resolution faces.

Additionally, we can see that the coefficients  $\eta = 0$  and  $\gamma = 0$ ; that is, the distribution information is not used as supervision; the recognition effect is not the worst, but it is not optimal. It indicates that the distribution information of samples as supervision effectively improves the performance of coupled metric method.

With the increasing of these two parameters, the recognition performance of LMS-CML decreases slightly. There are two reasons for this variation: (1) When the distribution information of samples is considered more, the main supervisory role of category label is weakened. (2) In construction of distribution correlation matrix, the neighbor

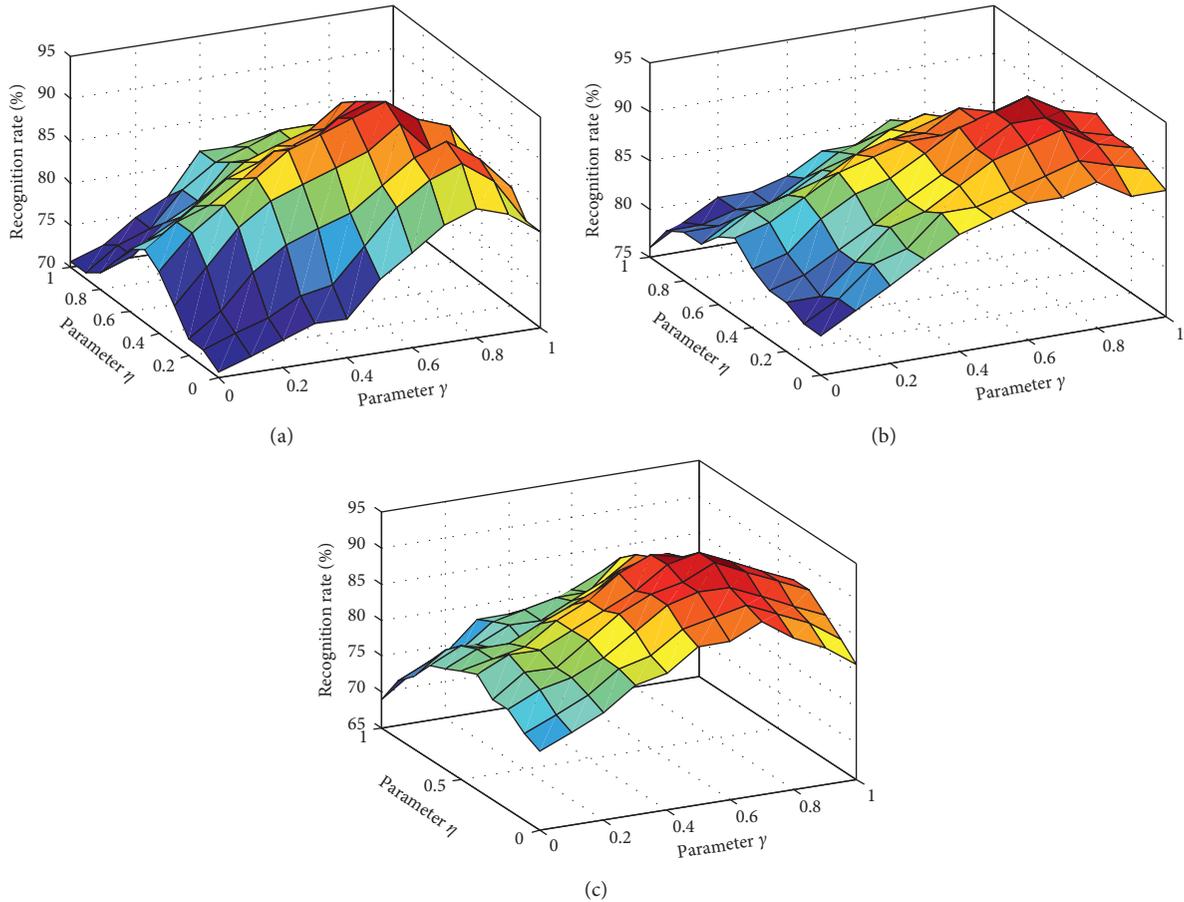


FIGURE 4: Recognition rate with different parameters  $\gamma$  and  $\eta$ . (a) Experiment result based on Yale-B face database. (b) Experiment result based on ORL face database. (c) Experiment result based on UMIST face database.

relation of samples with the same class is easily interfered by samples of other categories. Therefore, with the increasing of  $\eta$  and  $\gamma$ , the proportion of auxiliary function in the general objective function increases, which enlarges the interference intensity and affects the recognition effect.

#### 4.3. Nonlinear Multisupervised Coupled Metric Experiments.

Different from the linear coupled metric learning, there are four influencing factors in the nonlinear coupled metric learning method; in addition to the three influencing factors as described in Section 4.2, the adjustable factor  $\sigma$  of Gauss kernel function is the fourth influencing factor. In experiment 1 of this section, we mainly analyzed the joint influence of adjustable factor  $\sigma$  and dimension of reserved features. The influence of the coefficients  $\gamma$  and  $\eta$  in the general objective function is discussed in experiment 2. In addition, the experiment results show that the number of nearest neighbors in the nonlinear case is the same as that in the linear case.

**4.3.1. Experiment 1.** In the following experiment, we analyzed the impact on recognition rate based on different adjustable factor  $\sigma$  and varied dimensions of reserved features. The results are shown in Figure 5.

Figure 5 shows the average recognition rates of NMS-CML method on different datasets. Comparing with the experiment results in Section 4.2.1, we can see that the recognition effect is obviously improved in the nonlinear case. Furthermore, to achieve an optimal recognition rate, the dimension of reserved features in the nonlinear case is higher than that in the linear case. When the adjustable factor  $\sigma$  is near the optimal value, the recognition rate has an improvement dramatically, which indicates that the adjustable factor has a greater impact on the recognition performance.

The curves indicate that the dimension of features is 25 and  $\sigma=0.5$ ; the optimal face recognition rate of 91.15% is obtained in Yale-B face dataset. When the dimension of features is 30 and  $\sigma=0.3$ , the best recognition rate is 97% in ORL face dataset. Then, in UMIST face dataset, the dimension of features is 25 and  $\sigma=0.7$ ; the best recognition rate is 95.59%. These results illustrate that the NMS-CML can better extract the low-resolution and pose-varied face features, but the improvement in recognition effect of low-resolution images with severe illumination is limited.

**4.3.2. Experiment 2.** We experimented and discussed the effects of parameters  $\gamma$  and  $\eta$  in the nonlinear general

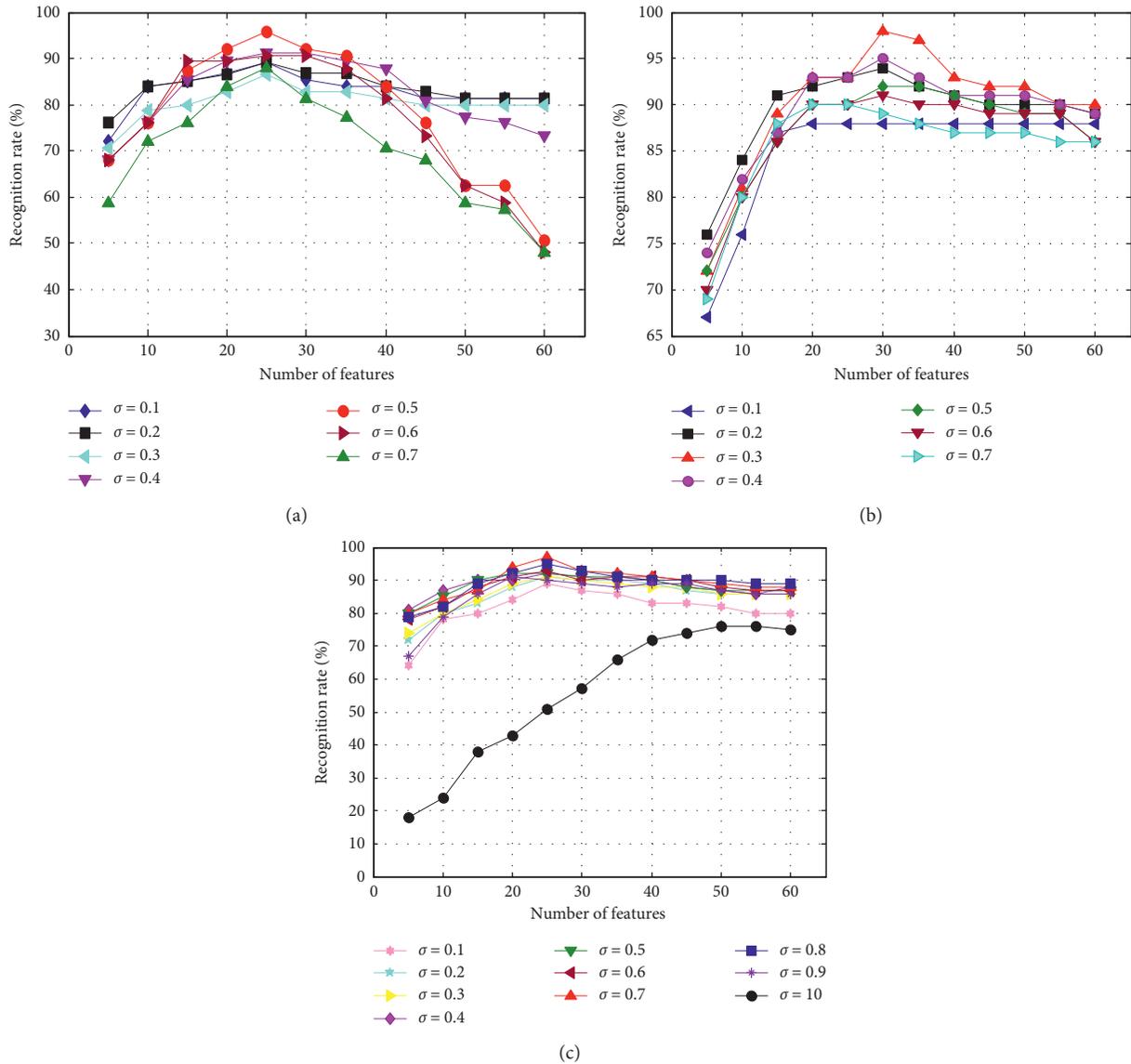


FIGURE 5: Recognition results based on different adjustable factor of kernel function and feature dimensions. (a) Experiment result based on Yale-B face database. (b) Experiment result based on ORL face database. (c) Experiment result based on UMIST face database.

objective function. The conclusion is that the supervised effect of the distribution relationship between high-resolution images is stronger than that between low-resolution images. The optimal values of  $\gamma$  and  $\eta$  are 0.65 and 0.3, respectively, in Yale-B face dataset. The optimal values of  $\gamma$  and  $\eta$  are 0.7 and 0.3 in ORL face dataset. The best values of  $\gamma$  and  $\eta$  are 0.7 and 0.45 in UMIST face dataset. In general, the category label has the strongest supervision effect in MS-CML, and the distribution information of samples in high-resolution image set has stronger supervision than that of low-resolution image set.

**4.4. Matching Experiments of Images with Different Resolutions.** In the above experiments, the resolution of high-quality face images is set to  $64 \times 64$  pixels, and the resolution of low-resolution face images is set to  $16 \times 16$

pixels. To better illustrate the effectiveness of MS-CML method, in this section, we performed coupled metric experiments of low-quality face images with different resolutions. The resolution of low-quality face images is set to  $32 \times 32$ ,  $16 \times 16$ , and  $8 \times 8$  pixels. Figure 6 shows some faces with different resolutions, and the experiment results are summarized in Table 2.

The experimental results show that the resolution of low-quality faces is set to different values; the recognition performance of MS-CML is affected to some extent. When the resolution of low-quality face image is set to  $8 \times 8$  pixels, the face image contains less effective classification information; thus, the recognition rate drops more. In addition, due to strong illumination interference in Yale-B face dataset and obvious pose variation in UMIST face dataset, the effective classification information further reduced, and the recognition rate decreased more than that of ORL face dataset. In

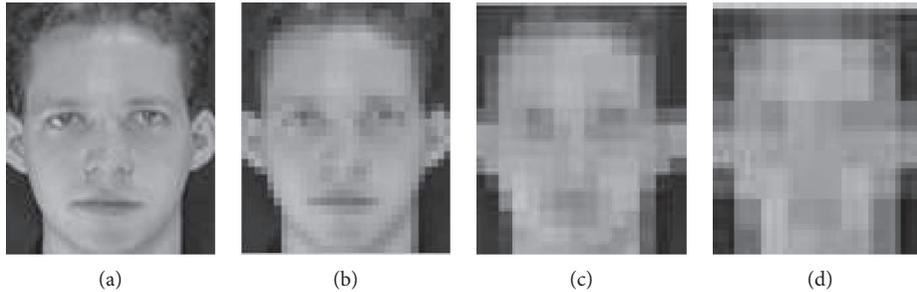


FIGURE 6: Different resolution images for one individual. (a)  $64 \times 64$ . (b)  $32 \times 32$ . (c)  $16 \times 16$ . (d)  $8 \times 8$ .

TABLE 2: Face recognition results under different resolutions.

Methods	Yale-B face dataset			ORL face dataset			UMIST face dataset		
	$32 \times 32$	$16 \times 16$	$8 \times 8$	$32 \times 32$	$16 \times 16$	$8 \times 8$	$32 \times 32$	$16 \times 16$	$8 \times 8$
LMS-CML	90.90	89.06	85.76	97.00	93.00	89.50	93.62	92.16	87.26
NMS-CML	93.64	91.15	87.26	98.50	97.00	94.50	97.06	95.59	90.69

TABLE 3: Comparison of degraded face recognition results between our method and other methods.

Methods	Yale-B dataset	ORL dataset	UMIST dataset	
Image reconstruction	PCA [23]	57.95	69.00	36.76
	KPCA [25]	68.68	78.00	61.27
	LPP [24]	66.67	73.00	53.92
	KLPP [26]	75.04	87.00	75.49
	CML [9]	75.97	82.00	77.45
	CLPM [11]	76.35	87.00	80.39
	SLPP-KCML [13]	85.45	90.00	86.76
	KCML [14]	83.85	91.50	85.78
	CMDM [15]	88.61	93.50	90.68
	LMS-CML	89.06	93.00	92.16
NMS-CML	<b>91.15</b>	<b>97.00</b>	<b>95.59</b>	

addition, the nonlinear coupled metric has a stronger feature extraction capability, so the LMS-CML has a much lower recognition rate than NMS-CML.

**4.5. Comparison Experiments with Other Methods.** In order to evaluate the performance of proposed coupled metric methods, based on the methods of feature extraction [23–26] after the image reconstruction [27, 28], the methods in [9, 11, 13–15], we make the comparison experiments, respectively. The experiment results are shown in Table 3.

We can see that the recognition effect of feature extraction after image reconstruction is not ideal. The main reason is that some false information is introduced in the image reconstruction, which is not obvious for the improvement of recognition effect, and even interferes with the recognition rate. In addition, the same person’s face images have large illumination and pose changes; this intraclass multimodal problem will affect the metric learning and final classification results. The methods in [9, 14] are distance metric learning only based on category label information, which cannot overcome the intraclass multimodal problem, so the recognition rate is lower. The distance metric methods in [11, 13] are based on locality preserving relation in the same class, which are advantageous to solve the intraclass

multimodal problem, so comparing with CML and KCML, the recognition effect is improved. The CMDM [15] combines discriminant information with data distribution of local neighbors, so the recognition effect is similar to the proposed LMS-CML method. However, the CMDM does not deal with the nonlinear data, so the final recognition effect is not as good as our proposed NMS-CML method.

Through comparison experiments, we get two conclusions. (1) The fusion of category label and distribution information can better supervise the learning of the metric matrix. (2) The nonlinear coupled transformation can extract the nonlinear essential features more effectively. So, in this paper, we propose the linear and nonlinear MS-CML method, which make full use of the supervision of category label and distribution information. Finally, we get the excellent recognition result in face recognition with low resolution and pose variations.

## 5. Conclusions

In this paper, a multisupervised coupled metric learning method was proposed to obtain the coupled mapping matrix, which can be used for low-resolution image matching. In the metric learning, we constructed the linear and

nonlinear coupled metric learning objective functions, respectively. Compared with linear metric learning, the mapping matrix obtained by nonlinear metric learning can better improve the accuracy of low-resolution image matching. In addition, experimental results prove that the proposed MS-CML can overcome the intraclass multimodal problem of existing metric methods and effectively improve the matching accuracy of low-resolution images under the condition of small samples.

## Data Availability

The experimental data of face database setting and recognition rate used to support the findings of this study are included within the article. The download address of the three face databases has been added to the references; you can download it by yourself, if necessary.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This study was funded by the Visiting Project Funds of Shandong University of Technology (Grant no. 20180831); Integration Funds of Shandong University of Technology and Zhangdian District (Grant no. 118228); and the Natural Science Foundation of China (Grant no. 61601266).

## References

- [1] Y. Shen, Y. Yan, and H. Wang, "Recent advances on supervised distance metric learning algorithms," *Acta Automatica Sinica*, vol. 40, no. 12, pp. 2673–2686, 2014.
- [2] J. Lu, G. Wang, and P. Moulin, "Localized multifeature metric learning for image-set-based face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 529–540, 2016.
- [3] G. Wang, F. Zheng, C. Shi, J.-H. Xue, C. Liu, and L. He, "Embedding metric learning into set-based face recognition for video surveillance," *Neurocomputing*, vol. 151, pp. 1500–1506, 2015.
- [4] X. Lu, Y. Wang, X. Zhou, and Z. Ling, "A method for metric learning with multiple-kernel embedding," *Neural Processing Letters*, vol. 43, no. 3, pp. 905–921, 2016.
- [5] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proceedings of the British Machine Vision Conference 2012*, pp. 1–10, Oxford, UK, September 2012.
- [6] W. Wu, Z. Liu, and X. He, "Learning-based super resolution using kernel partial least squares," *Image and Vision Computing*, vol. 29, no. 6, pp. 394–406, 2011.
- [7] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
- [8] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *IEEE Transactions on Cybernetics*, vol. 47, no. 11, pp. 3991–4002, 2017.
- [9] B. Li, H. Chang, S. Shan, and X. Chen, "Coupled metric learning for face recognition with degraded images," in *Proceedings of the Asian Conference on Machine Learning*, Springer, Tokyo, Japan, pp. 220–233, 2009.
- [10] B. Li, H. Chang, S. Shan, and X. Chen, "Aligning coupled manifolds for face hallucination," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 957–960, 2009.
- [11] 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.
- [12] X. Ben, W. Meng, R. Yan, and K. Wang, "Kernel coupled distance metric learning for gait recognition and face recognition," *Neurocomputing*, vol. 120, pp. 577–589, 2013.
- [13] G. Zou, Y. Zhang, K. Wang, S. Jiang, H. Wan, and G. Fu, "An improved metric learning approach for degraded face recognition," *Mathematical Problems in Engineering*, vol. 2014, Article ID 724978, 10 pages, 2014.
- [14] K. Wang and T. Yan, "Kernel coupled metric learning and its application to gait recognition," *Pattern Recognition and Artificial Intelligence*, vol. 26, no. 2, pp. 169–175, 2013.
- [15] P. Zhang, X. Ben, W. Jiang, R. Yan, and Y. Zhang, "Coupled marginal discriminant mappings for low-resolution face recognition," *Optik*, vol. 126, no. 23, pp. 4352–4357, 2015.
- [16] J. Jiang, R. Hu, Z. Han, L. Chen, and J. Chen, "Coupled discriminant multi-manifold analysis with application to low-resolution face recognition," in *Proceedings of the International Conference on Multimedia Modeling*, Springer, Sydney, Australia, pp. 37–48, January 2015.
- [17] J. Jiang, R. Hu, Z. Wang, and Z. Cai, "CDMMA: coupled discriminant multi-manifold analysis for matching low-resolution face images," *Signal Processing*, vol. 124, pp. 162–172, 2016.
- [18] V. E. Liang, J. Lu, Y.-P. Tan, and J. Zhou, "Deep coupled metric learning for cross-modal matching," *IEEE Transactions on Multimedia*, vol. 19, no. 6, pp. 1234–1244, 2017.
- [19] J. Lu, J. Hu, and Y.-P. Tan, "Discriminative deep metric learning for face and kinship verification," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4269–4282, 2017.
- [20] Center for Computational Vision and Control, Yale University, Yale-b Face Dataset, 2001, <https://computervisiononline.com/dataset/1105138686>.
- [21] AT and T Laboratory of University of Cambridge, Orl face database, 2002, <https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.
- [22] H. Wechsler, J. P. Phillips, V. Bruce, F. F. Soulie, and T. S. Huang, *Face Recognition: From theory to applications*, vol. 163, Springer Science & Business Media, Berlin, Germany, 2012.
- [23] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [24] X. He and P. Niyogi, "Locality preserving projections," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 153–160, Vancouver, Canada, 2004.
- [25] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [26] D. Cai, X. He, and J. Han, "Spectral regression for dimensionality reduction," Department of Computer Science Technical Report No. 2856 (UIUCDCS-R-2007-2856), University of Illinois at Urbana-Champaign, Urbana, IL, USA, 2007.

- [27] Y. Kong, S. Zhang, and P. Cheng, "Super-resolution reconstruction face recognition based on multi-level ffd registration," *Optik*, vol. 124, no. 24, pp. 6926–6931, 2013.
- [28] J. Jiang, R. Hu, Z. Han, and Z. Wang, "Low-resolution and low-quality face super-resolution in monitoring scene via support-driven sparse coding," *Journal of Signal Processing Systems*, vol. 75, no. 3, pp. 245–256, 2014.