Regularized Discriminative Spectral Regression Method for Heterogeneous Face Matching

Xiangsheng Huang, Zhen Lei, *Member, IEEE*, Mingyu Fan, Xiao Wang, and Stan Z. Li, *Fellow, IEEE*

***Abstract*— Face recognition is confronted with situations in which face images are captured in various modalities, such as the visual modality, the near infrared modality, and the sketch modality. This is known as heterogeneous face recognition. To solve this problem, we propose a new method called discriminative spectral regression (DSR). The DSR maps heterogeneous face images into a common discriminative subspace in which robust classification can be achieved. In the proposed method, the subspace learning problem is transformed into a least squares problem. Different mappings should map heterogeneous images from the same class close to each other, while images from different classes should be separated as far as possible. To realize this, we introduce two novel regularization terms, which reflect the category relationships among data, into the least squares approach. Experiments conducted on two heterogeneous face databases validate the superiority of the proposed method over the previous methods.**

***Index Terms*— Discriminative regularization, face recognition, heterogeneous data processing, spectral regression, subspace learning.**

I. Introduction

SUBSPACE learning is an important approach in face recognition. A large number of methods, including Principal Component Analysis (PCA) [1], Linear Discriminative Analysis (LDA) [2], Locality Preserving Projection (LPP) [3], Neighborhood Preserving Embedding (NPE) [4] and Marginal Fisher Analysis (MFA) [5], have been proposed to solve this problem. These methods have been proved effective in face recognition [6]. However, all of these methods are designed to apply the image data in only one modality. On the other hand, in practical face recognition systems, the image data can be captured from more than one modality, including the near infrared (NIR) modality, the visual (VIS) modality, the sketch modality and so on. Therefore, it is necessary to solve the heterogeneous matching problem, which enables the utilization of all the images in various modalities for recognition. The Coupled Spectral Regression (CSR) method for heterogeneous face recognition has been proposed in our previous work [7]. It deals with the face images captured from two different modalities (i.e., the NIR modality and the VIS modality). The method maps the heterogeneous images into a common subspace by two different projective mappings. However, CSR does not make use of the discriminative information among images from different classes sufficiently, and its performance can be further improved by introducing the category relationship into its objective function.

In this paper, we propose a new method, the Discriminative Spectral Regression (DSR), for heterogeneous face recognition to improve our previous work [7]. The DSR method finds the projective mappings which map heterogeneous face images of the same person to similar representations and map images from different persons to significantly different representations. Specifically, compared with the previous works, the proposed method has two advantages.

1) The proposed method provides a theoretical framework for heterogeneous image matching problems. It can handle data captured in multiple modalities. In comparison, most previous heterogeneous subspace learning methods, such as the CSR [7] method and the Canonical Correlation Analysis (CCA) based method [8], can work only on databases captured in two modalities.
2) We introduce two new regularization terms that can effectively make use of the class information in the training set. Therefore, in the DSR method, the class information is not only considered in the cost function but also integrated in the two novel regularization terms. Subsequently, the learned subspace is more discriminative than conventional CSR.

The remainder of the paper is organized as follows. In Section II, we describe the heterogeneous face matching problem and review related heterogeneous matching methods. A general model for homogeneous subspace learning problem is introduced in Section III and a general framework for heterogeneous subspace learning is presented in Section IV. Experimental results on VIS-NIR and photo-sketch databases are demonstrated in Section V and in Section VI, we conclude the paper.

Manuscript received September 9, 2011; revised August 10, 2012; accepted August 11, 2012. Date of publication August 27, 2012; date of current version December 20, 2012. This work was supported in part by the Chinese National Natural Science Foundation (NSF) under Project 61070146, Project 61175034, and Project 61103156, the National IoT R&D under Project 2150510, the Chinese Academy of Sciences under Project KGZD-EW-102-2, and the NSF of Zhejiang Province, under Grant LQ12F03004. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bulent Sankur.

X. Huang is with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: xiangsheng.huang@ia.ac.cn).

Z. Lei and S. Z. Li are with the Center for Biometrics and Security Research and the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zlei@cbsr.ia.ac.cn; szli@cbsr.ia.ac.cn).

M. Fan is with the Institute of Intelligent Systems and Decision, Wenzhou University, Wenzhou 325000, China (e-mail: fanmingyu@wzu.edu.cn).

X. Wang is with the School of Mathematical Sciences, Graduate University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: wangxiao@lsec.cc.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2215617

II. Related Works

In practical face recognition systems, face images may be captured in more than one modality. For examples, the NIR based face recognition methods [9] have been developed to overcome the illumination variation problem; sketch images drawn by artists based on the recollection of an eyewitness have been used in the retrieval of a sketch from the police mug-shot databases. Therefore, heterogeneous face recognition is a current topic of interest. Fig. 1 illustrates the heterogeneous face recognition problem. For each person, multiple images are captured through more than one modality, such as the NIR (N) modality, the VIS (V) modality and the sketch (S) modality. To realize robust classification, subspace learning should find a common discriminative subspace in which the representations of heterogeneous images from the same person are as close as possible while the representations of heterogeneous images from different persons are as far as possible.

Tang and Wang [10] proposed an eigen-transform based method for matching sketch images with photos. Let the matrix $P = [P_1, P_2, \ldots, P_N] \in \mathbb{R}^{D\times N}$ be the photo training set, and $S = [S_1, S_2, \ldots, S_N] \in \mathbb{R}^{D\times N}$ be the corresponding sketch training set, where D is the dimension and N is the number of images. For a new photo P_r, it first compute the reconstruction coefficients using the photo training set. Subsequently, the same combination coefficients are used to synthesize pseudo sketch image S_r with the corresponding sketch training images. Finally, S_r is used in the image retrieval from S. Their method has been proved effective in reducing the difference between photo and sketch.

The Common Discriminant Feature Extraction (CDFE) method was proposed by Lin and Tang [11]. Assuming c_i^q and c_i^r be the class labels of the query faces and the reference faces, this method transforms query faces $\{(x_i^q, c_i^q)\}_{i=1}^{N_q}$ and reference faces $\{(x_i^r, c_i^r)\}_{i=1}^{N_r}$ into a common discriminant subspace by minimizing the following objective function,

$$J(A^q, A^r) = \sum_{i=1}^{N_q}\sum_{j=1}^{N_r} u_{ij}\|A^q x_i^q - A^r x_j^r\|^2 + \sum_{i=1}^{N_q}\sum_{j=1}^{N_q} v_{ij}^q \times \|A^q x_i^q - A^q x_j^q\|^2 + \sum_{i=1}^{N_r}\sum_{j=1}^{N_r} v_{ij}^r \|A^r x_i^r - A^r x_j^r\|^2$$

where u_{ij} describes the intra-class compactness and inter-class dispersion, v_{ij}^q and v_{ij}^r reflect the affinity of nearby data points, A^q is the projective mapping for query faces and A^r is the mapping for reference faces. CDFE method is somewhat time-consuming because of its pairwise way to compute the scatter matrices.

The Canonical Correlation Analysis (CCA) method for heterogeneous data is proposed by Yi et al. [8]. At the first step of their method, the PCA or LDA method is applied to find the low-dimensional representations X_n' and X_v' of the NIR images X_n and the VIS images X_v, respectively. At the second step, it computes the best correlational regression projections A_n and A_v between X_n' and X_v' by maximizing the following correlation function

$$\rho(A_n, A_v) = \frac{A_n^T C_{nv} A_v}{\sqrt{A_n^T C_{nn} A_n A_v^T C_{vv} A_v}}$$

where C_{nv}, C_{nn} and C_{vv} are the correlation matrices computed from the representations X_n' and X_v'. Let $x_n \in X_n'$ and $x_v \in X_v'$. At the third step, the correlation value $\frac{<A_n x_n \cdot A_v x_v>}{\|x_n\|\|x_v\|}$ is calculated as the matching score. In [12], Yang et al. proposed the regularized kernel CCA method to learn the relationship between subspaces of VIS images and 3D depth images.

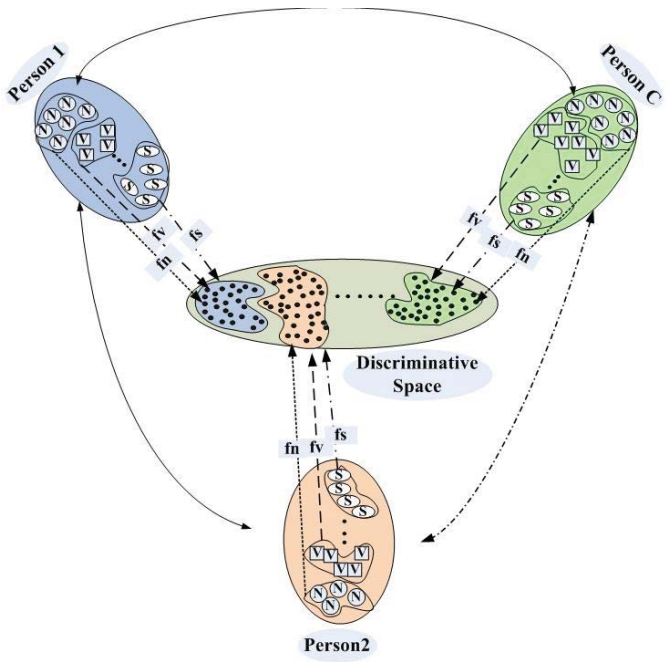


Fig. 1. Illustration of the heterogeneous face recognition problem. Images of different modalities [the NIR (N) modality, the VIS (V) modality, and the sketch (S) modality] from C persons are mapped into a common discriminative space.

In our previous work [7], the CSR method first models the properties of the NIR images X_n and the VIS images X_v separately and then learns two associated projective mappings for the heterogeneous images. Assuming A_n and A_v are linear projective matrices, the objective optimization of CSR method is formulated as follows

$$\{A_n, A_v\} = \arg\min_{A_n A_v} \left\{ \frac{1}{N_n}\|Y_n - A_n X_n\|^2 + \frac{1}{N_v}\|Y_v - A_v X_v\|^2 + \eta\|A_n - A_v\|^2 + \lambda(\|A_n\|^2 + \|A_v\|^2) \right\}$$

where N_n denotes the number of the NIR images, N_v is the number of the VIS images, Y_n and Y_v are pre-computed low-dimensional representations of X_n and X_v, respectively. Experiments on VIS-NIR face database prove that CSR method outperforms previous methods. However, CSR method does not consider the category relationship during the regression from the image data to the low-dimensional representations, so there is potential to extract the discriminative information more sufficiently.

III. Spectral Regression for Homogeneous Subspace Learning

In this section, we will consider the case when the data are captured through a single modality, such as by NIR or by VIS. Up to now, almost all the works on subspace learning have been studied for this case.

Without loss of generality, we assume that there are N data points in the training set, then X can be written in the matrix form $X = [x_1, \ldots, x_N]$. Suppose that $Y = [y_1, y_2, \ldots, y_N]$ is the low-dimensional embedding of X where $y_i \in \mathbb{R}^d$ ($d \ll D$) denotes the low-dimensional representation of x_i. In the graph embedding framework [5], the optimal embedding Y can be obtained as

$$\begin{aligned} Y^* &= \arg \min_{\mathrm{tr}(YBY^T)=1} \sum_{i,j} \|y_i - y_j\|^2 w_{ij} \\ &= \arg \min_{\mathrm{tr}(YBY^T)=1} \mathrm{tr}(YLY^T) \end{aligned} \tag{1}$$

where w_{ij} measures the similarity between data points x_i and x_j, $W = \{w_{ij}\}$, $L = B - W$ is the graph Laplacian, and B is a diagonal matrix with $B_{ii} = \sum_j w_{ij}$. The extra constraint $\mathrm{tr}(YBY^T) = 1$ is imposed to avoid the ill posed problem.

For the choice of W in LDA, assuming the t-th class has n_t data points and $\sum_{t=1}^{C} n_t = N$, the similarities between data points are defined as

$$w_{ij} = \begin{cases} 1/n_t, & \text{if } x_i \text{ and } x_j \text{ both belong to the } t\text{th class;} \\ 0, & \text{otherwise.} \end{cases}$$

As a block-wise diagonal matrix, it is easy to check that the rank of W equals to C, the number of classes, and there is only one nonzero eigenvalue equal to 1. It is straightforward to show that the corresponding eigenvectors of LDA based graph Laplacian are given as follows.

$$v_t = (\underbrace{0, \ldots, 0}_{\sum_{i=1}^{t-1} n_i}, \underbrace{1, \ldots, 1}_{n_t}, \underbrace{0, \ldots, 0}_{\sum_{i=t+1}^{c} n_i})^T, \qquad t = 1, \ldots, C.$$

Subsequently, we can get the $C-1$ useful orthogonal eigenvectors $\{\tilde{v}_t, t = 1, \ldots, C-1\}$ by implementing the Gram-Schmidt orthogonalization algorithm on $\{v_t, t = 1, \ldots, C\}$. As is shown in [13], the $C-1$ orthogonal eigenvectors are sufficient to represent a C class problem. So the low-dimensional representations are obtained as $Y^* = [\tilde{v}_1, \ldots, \tilde{v}_{C-1}]^T$, where $Y^* \in \mathbb{R}^{(C-1)\times N}$. In our methods, we first compute Y^* using the LDA based graph Laplacian method and then let $Y = [y_1, y_2, \ldots, y_N] = Y^*$ be the low-dimensional representation of data for regression.

The graph Laplacian method can only provide the low-dimensional representations for the training data. The mapping f from the observation space to the discriminative space is implicit. However, the mapping from x_i to y_i is essential to classify the newly introduced data. To address this problem, one need to find the mapping function $f: \mathbb{R}^D \to \mathbb{R}^{C-1}$, where the relationship holds:

$$y_i = f(x_i), \qquad i = 1, \ldots, N. \tag{2}$$

Here the mapping function f can be linear or nonlinear. In real applications, the equality in (2) may not hold exactly. To deal with this, we just require that $f(x_i)$ and y_i be as close as possible in the least squares sense, which is presented as

$$f^* = \arg \min_f \sum_{i=1}^{N} \|y_i - f(x_i)\|^2. \tag{3}$$

However, using the least squares approach in Eq. (3) is not enough for fitting a faithful mapping because that the least squares approach may lead to the problem of over fitting and thus has poor generalization ability.

To make the mapping f more discriminative, we modify the objective function in Eq. (3). On one hand, we hope that f maps data from the same class to the same area in low-dimensional space, i.e., their low-dimensional representations would be close to each other. Mathematically, we realize this goal through minimizing the sum of distances after mapping in low-dimensional subspace

$$\lambda \cdot \sum_{x_i, x_j \in \text{s.c.}} \|f(x_i) - f(x_j)\|^2$$

where "s.c." is the abbreviation for "same class" and λ is a nonnegative parameter which can be learned from the training set or just given some empiric value.

On the other hand, for the data from different classes, we wish to separate them as far as possible in the discriminative space. Contrary to the within-class case, we just need to maximize the distances among the low-dimensional representations. Equivalently, we minimize

$$-\eta \cdot \sum_{x_i, x_j \in \text{d.c.}} \|f(x_i) - f(x_j)\|^2.$$

Here, "d.c." is the abbreviation for "different class", η is also a nonnegative trade-off parameter.

In addition, in order to control the complexity of projective mapping, an extra normalization term $\xi\|f\|^2$ needs to be included, where $\xi > 0$ is a parameter. To sum up, the discriminative spectral regression model dealing with the homogeneous subspace learning is formulated as follows:

$$\begin{aligned} f^* = \arg \min_f \Bigg\{ &\sum_{i=1}^{N} \frac{1}{N} \|y_i - f(x_i)\|^2 + \lambda \cdot \sum_{\text{s.c.}} \|f(x_i) - f(x_j)\|^2 \\ &- \eta \cdot \sum_{\text{d.c.}} \|f(x_i) - f(x_j)\|^2 + \xi \|f\|^2 \Bigg\}. \end{aligned} \tag{4}$$

IV. Spectral Regression for Heterogeneous Subspace Learning

Intuitively different data modalities should be equipped with different projective mappings for feature extraction. Similar idea was introduced in [7]. In this section, we will give a new method to extract the discriminative feature for heterogeneous databases. This new method is designed to minimize the distances between low-dimensional representations of the same class and maximizes distances between low-dimensional representations of the different classes. Compared with previous methods, it can handle data captured from more than two modalities.

For simplicity, we introduce some concepts and notations. Assume that there are r kinds of data modalities. Accordingly, there are r projective mappings from the observation spaces to the discriminative space that we need to find. Denote them as $\{f^1, f^2, \ldots, f^r\}(r \geq 2)$. In the i-th kind of modality, we capture l_i data points, denoted as $\{x_1^i, \ldots, x_{l_i}^i\}(l_i \geq 1)$. Here f^i can only work on those data $\{x_j^i\}_{j=1}^{l_i}$. For each data point $x_j^i \in \mathbb{R}^D$ $(i = 1, \ldots, r; j = 1, \ldots, l_i)$, which is labeled, y_j^i is denoted as its pre-computed low-dimensional representation in $\mathbb{R}^d$. For W in the LDA, the vectors $y_j^i, (i = 1, \ldots, r; j = 1, \ldots, l_i)$ are learned via graph Laplacian in the preprocessing stage, as is described in the Section III.

To define a model which can describe the heterogeneous subspace learning problem well, some issues need to be considered. Firstly, since for any mapping function f^i, its motivation is to map high-dimensional data x_j^i close to its low-dimensional representation y_j^i. It is natural to minimize the distances between $f^i(x_j^i)$ and y_j^i for all the training data in the least squares sense

$$\sum_{i=1}^{r}\sum_{j=1}^{l_i}\frac{1}{l_i}\|y_j^i - f^i(x_j^i)\|^2. \tag{5}$$

Besides Eq. (5), it is also necessary to consider both within-class and between-class information. From the point of geometric distance, it is desirable that heterogeneous data from the same class should be forced to be close to each other in the discriminative space. So it is desirable that the projective mappings minimize

$$\sum_{x_j^i, x_l^k \in \text{s.c.}} \|f^i(x_j^i) - f^k(x_l^k)\|^2. \tag{6}$$

And also for data from different classes, it is expected that the projective mappings can maximize the distances among their low-dimensional representations. Therefore, it is reasonable to minimize

$$-\sum_{x_j^i, x_l^k \in \text{d.c.}} \|f^i(x_j^i) - f^k(x_l^k)\|^2. \tag{7}$$

Combining the cost function in (5), the two novel regularization terms in (6)-(7), and balancing their different contributions, and also considering the regularization on the projective mappings themselves, we propose the general model for solving the heterogeneous subspace learning problem as

$$\begin{aligned}\{f^1, \ldots, f^r\} = \arg\min \Bigg\{ & \sum_{i=1}^{r}\sum_{j=1}^{l_i}\frac{1}{l_i}\|y_j^i - f^i(x_j^i)\|^2 \\ & + \lambda \cdot \sum_{\text{s.c.}} \|f^i(x_j^i) - f^k(x_l^k)\|^2 \\ & - \eta \cdot \sum_{\text{d.c.}} \|f^i(x_j^i) - f^k(x_l^k)\|^2 \\ & + \xi \sum_{i=1}^{r} \|f^i\|^2 \Bigg\}. \end{aligned} \tag{8}$$

However, in the optimization problem (8), the superscripts i, k and subscripts j, l are correlated to each other. Because of this, it is impossible to transform Eq. (8) into a matrix optimization which can be solved efficiently. As follows, we will discuss several variations of the optimization problem (8).

A. Linear Heterogeneous Subspace Learning

In linear case, the projective mappings are linear transformations given by

$$f^i(x) = A^i x$$

where $A^i \in \mathbb{R}^{d \times D}$ is the projective matrix for the i-th data modality. So the original optimization problem in Eq. (8) is specialized to the following minimization problem

$$\begin{aligned}\min_A \Bigg\{ & \sum_{i=1}^{r}\sum_{j=1}^{l_i}\frac{1}{l_i}\|y_j^i - A^i x_j^i\|^2 + \lambda \cdot \sum_{\text{s.c.}} \|A^i x_j^i - A^k x_l^k\|^2 \\ & - \eta \cdot \sum_{\text{d.c.}} \|A^i x_j^i - A^k x_l^k\|^2 + \xi \sum_{i=1}^{r} \|A^i\|^2 \Bigg\}. \end{aligned} \tag{9}$$

To formulate above problem Eq. (9) in a simple matrix-vector form, we just minimize or maximize (depending on wether they are of the same class or not) the distances between representations of the **same modality** instead of **all modalities**. This variation brings a concise but slightly different optimization problem for subspace learning. By matrix manipulation, we can check that the new optimization problem for heterogeneous subspace learning does not compromise our original motivation. We reformulate the data matrix $\tilde{X}$ in the form

$$\tilde{X} = \left[\tilde{X}^1, \tilde{X}^2, \ldots, \tilde{X}^r\right] \in \mathbb{R}^{D \cdot r \times N}$$

where $\tilde{X}^i = [\underbrace{0; \cdots; 0}_{(i-1)D}; X^i; \underbrace{0; \cdots; 0}_{(r-i)D}]$, $X^i = [x_1^i, \ldots, x_{l_i}^i]$ and N is the number of all the training data points. Correspondingly, we group the projective matrix and the representation matrix as

$$A = \begin{pmatrix} A^1 & 0 & \cdots & 0 \\ 0 & A^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A^r \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 & 0 & \cdots & 0 \\ 0 & Y_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Y_r \end{pmatrix} \tag{10}$$

where $Y_i = (y_1^i, \ldots, y_{l_i}^i)$ and $y_j^i \in \mathbb{R}^d$ is the pre-computed representation of x_j^i. The cost function of Eq. (9) can then be reformulated as

$$\frac{1}{N}\|Y - A\tilde{X}\|^2.$$

In the following, we replace the regularization terms (the summation of within or between scatters of all modalities) with new terms (the summation of within or between scatters of the same modality), which does not compromise our original motivation.

Let x_i denotes the i-th sample data, we define two symmetric matrices W^{sc}, $W^{dc} \in \mathbb{R}^{N\times N}$ as

$$W_{ij}^{sc} = \begin{cases} 1, & x_i, x_j \text{ belong to the same} \\ & \text{class and the same modality} \\ 0, & \text{otherwise} \end{cases}$$

$$W_{ij}^{dc} = \begin{cases} 1, & x_i, x_j \text{ belong to different} \\ & \text{classes and the same modality,} \\ 0, & \text{otherwise} \end{cases}$$

and two Laplacian matrices $L^{sc} = D^{sc} - W^{sc}$ and $L^{dc} = D^{dc} - W^{dc}$, where D^{sc} and D^{dc} are diagonal matrices with $D_{ii}^{sc} = \sum_j W_{ij}^{sc}$ and $D_{ii}^{dc} = \sum_j W_{ij}^{dc}$.

The following two regularization terms of Eq. (9)

$$\sum_{\text{s.c.}} \|A^i x_j^i - A^k x_l^k\|^2, \qquad \sum_{\text{d.c.}} \|A^i x_j^i - A^k x_l^k\|^2$$

are replaced, although not equivalent, with

$$\sum_{i,j} \|A\tilde{x}_i - A\tilde{x}_j\|^2 W_{ij}^{sc}, \quad \text{and} \quad \sum_{i,j} \|A\tilde{x}_i - A\tilde{x}_j\|^2 W_{ij}^{dc}$$

respectively, where $\tilde{x}_i$ denotes the i-th column of $\tilde{X}$.

Finally, the optimization problem described in (9) can be transformed to the matrix form as

$$\min_A \left\{ \frac{1}{N}\|Y - A\tilde{X}\|^2 + \lambda \cdot \sum_{i,j} \|A\tilde{x}_i - A\tilde{x}_j\|^2 W_{ij}^{sc} - \eta \cdot \sum_{i,j} \|A\tilde{x}_i - A\tilde{x}_j\|^2 W_{ij}^{dc} + \xi\|A\|^2 \right\}. \tag{11}$$

By algebraic manipulations, we can get the following matrix transformations

$$\sum_{i,j} \|A\tilde{x}_i - A\tilde{x}_j\|^2 W_{ij}^{sc} = 2 \times \text{tr}\left(A\tilde{X}L^{sc}\tilde{X}^T A^T\right)$$

$$\sum_{i,j} \|A\tilde{x}_i - A\tilde{x}_j\|^2 W_{ij}^{dc} = 2 \times \text{tr}\left(A\tilde{X}L^{dc}\tilde{X}^T A^T\right).$$

Therefore, through imposing the derivatives of the objective function with respect to A to zero, we obtain the solution by solving the following problem

$$A\left[\tilde{X}\tilde{X}^T + 2\lambda N\tilde{X}L^{sc}\tilde{X}^T - 2\eta N\tilde{X}L^{dc}\tilde{X}^T + \xi NI\right] = Y\tilde{X}^T. \tag{12}$$

The optimal projective matrix A can be computed by the direct inversion of a $Dr \times Dr$ matrix or by using the conjugate gradient method which needs $O(2CsDrN)$ computational time, where C is the number of classes and s is the number of iterations of LSQR algorithm [14]. The projective mappings $A^1, \ldots, A^r$ are obtained from the computed matrix A.

B. Kernel-Based Heterogeneous Subspace Learning

A strategy to extend methods to general nonlinear case is to take advantage of the *kernel trick* [15]. This technique is to map data into high dimensional feature space to make the problem solvable using linear methods.

Denote the mapping operator from input space to feature space as

$$\phi : \mathbb{R}^D \to \mathcal{F}.$$

For the data subset $X_i = [x_1^i, \ldots, x_{l_i}^i]$ obtained by i-th sampling method, its corresponding mapped data in $\mathcal{F}$ is denoted as

$$\phi(X^i) = [\phi(x_1^i), \ldots, \phi(x_{l_i}^i)].$$

Then we perform the linear algorithm in $\mathcal{F}$. Define the projective mapping related with i-th sampling methods as V^i. Here V^i can be considered as a $d \times Dim$ transformation matrix, where Dim is the dimensionality of the feature space $\mathcal{F}$. Then the relationship can be described as

$$Y_i = V^i\phi(X^i).$$

The subspace learning problem in Eq. (8) becomes to the following optimization problem

$$\min_{V^1,\ldots,V^r} \left\{ \sum_{i=1}^{r}\sum_{j=1}^{l_i} \frac{1}{l_i}\|y_j^i - V^i\phi(x_j^i)\|^2 + \lambda \cdot \sum_{\text{s.c.}} \|V^i\phi(x_j^i) - V^k\phi(x_l^k)\|^2 - \eta \cdot \sum_{\text{d.c.}} \|V^i\phi(x_j^i) - V^k\phi(x_l^k)\|^2 + \xi \sum_{i=1}^{r}\|V^i\|^2 \right\}. \tag{13}$$

The reproducing kernel theory [16] implies that each row of V^i is in $\text{span}\{\phi(x_1^i), \ldots, \phi(x_{l_i}^i)\}$, equivalently, there exists a coefficient matrix A^i such that

$$V^i = A^i\phi(X^i)^T$$

which implies that the relationship between high dimensional data and their pre-defined low dimensional representations can be expressed as

$$y_j^i = A^i\phi(X^i)^T\phi(x_j^i).$$

Let $k(x, y)$ denote the inner product of feature points $\phi(x)$ and $\phi(y)$ and $K(x, X_i)$ denote the vector $[k(x, x_1^i), \cdots, k(x, x_{l_i}^i)]^T$, the subspace learning problem (13) can be rewritten as

$$\min_{A^1,\ldots,A^r} \left\{ \sum_{i=1}^{r}\sum_{j=1}^{l_i} \frac{1}{l_i}\|y_j^i - A^iK(x_j^i, X_i)\|^2 + \lambda \cdot \sum_{\text{s.c.}} \|A^iK(x_j^i, X_i) - A^kK(x_l^k, X_k)\|^2 - \eta \cdot \sum_{\text{d.c.}} \|A^iK(x_j^i, X_i) - A^kK(x_l^k, X_k)\|^2 + \xi \sum_{i=1}^{r}\|A^i\|_K^2 \right\}.$$

Similar to Sec. IV-A, denoting $K_i = [K(x_1^i, X^i), K(x_2^i, X^i), K(x_{l_i}^i, X^i)]$, $\tilde{K}_i = [\underbrace{0; \cdots; 0}_{\sum_{t=1}^{i-1} l_t}; K_i; \underbrace{0; \cdots; 0}_{\sum_{t=i+1}^{r} l_t}]$, and

$\tilde{K} = [\tilde{K}_1, \tilde{K}_2, \ldots, \tilde{K}_r]$, we reformulate the above objective function as

$$\min_A \left\{ \frac{1}{N}\|Y - A\tilde{K}\|^2 + \lambda \cdot \sum_{i,j=1}^{N} \|A\tilde{k}_i - A\tilde{k}_j\|^2 W_{ij}^{sc} - \eta \cdot \sum_{i,j=1}^{N} \|A\tilde{k}_i - A\tilde{k}_j\|^2 W_{ij}^{dc} + \xi \|A\|_{\tilde{K}}^2 \right\} \quad (14)$$

where $\tilde{k}_i$ is the i-th column of $\tilde{K}$; A and Y are defined in the same way as in Eq. (10).

Similarly, by algebraic manipulations, we can get the following matrix transformations

$$\sum_{i,j=1}^{N} \|A\tilde{k}_i - A\tilde{k}_j\|^2 W_{ij}^{sc} = 2 \times \mathrm{tr}\left(A\tilde{K}L^{sc}\tilde{K}^T A^T\right)$$

$$\sum_{i,j=1}^{N} \|A\tilde{k}_i - A\tilde{k}_j\|^2 W_{ij}^{dc} = 2 \times \mathrm{tr}\left(A\tilde{K}L^{dc}\tilde{K}^T A^T\right).$$

Therefore, to solve heterogeneous subspace learning problem (13), we can obtain A by solving

$$\min_A \left\{ \frac{1}{N}\|Y - A\tilde{K}\|^2 + 2\lambda \cdot \mathrm{tr}\left(A\tilde{K}L^{sc}\tilde{K}^T A^T\right) - 2\eta \cdot \mathrm{tr}\left(A\tilde{K}L^{dc}\tilde{K}^T A^T\right) + \xi tr\left(A\tilde{K}A^T\right) \right\}.$$

Imposing the derivative of the objective function with respect to A in the problem above to zero, it follows that

$$A = Y\tilde{K}^T[\tilde{K}\tilde{K}^T + 2\lambda N\tilde{K}L^{sc}\tilde{K}^T - 2\eta N\tilde{K}L^{dc}\tilde{K}^T + \xi N\tilde{K}]^{-1}$$

Naturally, for each input data $\tilde{x}$ from the i-th sampling method, it can be mapped into the low-dimensional representation by

$$\tilde{y} = A^i K(\tilde{x}, X_i).$$

V. Experimental Results

The following experiments evaluate the proposed DSR methods in comparison with several existing methods of LDA [2], CDFE [11], PCA+CCA [8], LCSR and KCSR [7]. In classification phase, we adopt the cosine distance to measure the dissimilarity of data points in the learned subspace and the nearest neighbor (NN) classifier is chosen to do the classification task, where the cosine distance is

$$d_{cos}(x, y) = -\frac{x^T y}{\sqrt{x^T x y^T y}}. \quad (15)$$

A. Data Sets Description

To evaluate the performance of Linear DSR (LDSR) and Kernel DSR (KDSR) algorithm, a VIS-NIR database is collected [17]. There are 2095 VIS images and 3002 NIR images from 202 persons in the database. We apply two test protocols to evaluate different methods, where the database is split into training set and test set randomly. In protocol I, there are 1062 VIS and 1487 NIR images from 202 persons in the training set, and the remaining data points are left as test set.

Fig. 2. (a) VIS images. (b) NIR images.

It should be noted that all persons in the test set are contained in training set. In protocol II, the training set consists of 1438 VIS and 1972 NIR images from 168 persons while the test set consists of 657 VIS and 1075 NIR images from 174 persons. Therefore, the persons in test set are partially contained in the training set. All the images are scaled, transformed and cropped to 128×128 size according to automatically detected eye coordinates. Fig. 2 shows some cropped VIS and NIR face samples from this data set.

On this database, we apply two kinds of features as input for the compared algorithms. One is the intensity feature where each image is resized into 32×32 and transformed to a 1024 dimension feature vector. The other is LBP feature, which contains 1000 LBP feature values and is extracted to represent each face image. Therefore, there are four combinations of different feature types and protocols. The results are reported in terms of the receiver operating characteristic (ROC) curve and recognition rate.

A public available photo-sketch database [18] is also applied to evaluate the compared algorithms. This database contains 188 pairs of photo and sketch from 188 subjects. In this database, 88 photo-sketch pairs are applied as training data and the rest 100 photo-sketch pairs are applied as test data. All the images are cropped to 128×128 size according to the provided eye coordinates. Fig. 3 presents some photo-sketch pairs. We also apply two kinds of features as input for the compared algorithms. One is the intensity feature where each image is resized into 32×32 and then transformed to a 1024 dimension feature vector. The other is LBP features [19] which are extracted from 128×128 face images.

B. Parameters Selection and Experimental Settings

Parameters selection is a key issue for the compared algorithms. For CCA and CDFE methods, on both of the databases, data of each modality is first processed by PCA and 98 percent of data energy is preserved, meanwhile, the data are centralized according to their mean vector.

For CCA, CDFE, LCSR, and KCSR method, we optimize the parameters according to the recommended values in their papers. In LDSR and KDSR algorithms, each input data point is normalized to unit length. So the values of parameters are easy to choose. We fix $C\lambda = C(C-1)\eta/2 = 0.0001$. ξ is chosen from the set {0.0001, 0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9} and kernel width σ is chosen from

Fig. 3. Some photo-sketch examples from the CUHK face recognition database.

TABLE I
PARAMETERS SELECTION OF PROPOSED ALGORITHM

Database	LDSR	KDSR
Protocal I intensity	$\xi = 0.001$	$\xi = 0.001$, $\sigma = 0.7$
Protocal I LBP	$\xi = 0.1$	$\xi = 0.1$, $\sigma = 0.7$
Protocal II intensity	$\xi = 0.001$	$\xi = 0.001$, $\sigma = 0.7$
Protocal II LBP	$\xi = 0.5$	$\xi = 0.3$, $\sigma = 0.5$
CUHK face intensity	$\xi = 0.001$	$\xi = 0.001$, $\sigma = 0.7$
CUHK face LBP	$\xi = 0.001$	$\xi = 0.001$, $\sigma = 0.7$

the set $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ by performing cross-validation on the training data. The exact parameters settings of LDSR, and KDSR methods are given in Table I, where λ, η, ξ are corresponding parameters in their algorithms, and σ denotes the kernel width of the Gaussian kernel function.

C. Experimental Results

Tables II, III and IV demonstrate the recognition results with different configurations. Table II shows the results of the compared algorithms on image intensity and LBP features respectively following the protocol I on VIS-NIR database, Table III presents the corresponding results of compared algorithms following the protocol II, and Table IV demonstrates the recognition results of the compared algorithms on the CUHK face recognition database with image intensity and LBP features. Figs. 4 and 5 show the ROC curves of different methods with different configurations.

From the results, we can observe.

1) In Table II, we can see that the compared algorithms have very close performances following protocol I, where the recognition rates are all higher than 0.95. Although LDA method produces high recognition rates, it has the lowest Area Under Curve (AUC) score. This means that LDA method is unstable in finding the discriminative mapping. In term of both the recognition rate and the AUC score, the linear version of DSR method has better performance than linear CSR method. The ROC curves shown in Fig. 4(a) and (b) also support the conclusion that the proposed methods, LDSR and KDSR, outperform the compared algorithms following the protocol I.
2) As can be seen from Table III, LDA, CDFE and CCA methods have poorer performances following the protocol II in term of both the recognition rate and AUC socre. This is because that the training data do not contain all the classes contained in the test data. Therefore, the methods do not generalize to the untrained data well. The table indicates that the KDSR method has the best performances on both image intensity and LBP features. The ROC curves in Fig. 4(c) and (d) show that LDSR, KDSR, LCSR and KCSR have similar better performances than LDA, CDFE and CCA methods.
3) Table IV indicates LDSR and KDSR methods have superior performance on photo-sketch matching problems. As each data point is normalized to have unit length, the parameters of the proposed methods do not need to be carefully tuned. On the other hand, the parameters for other methods have to be changed to produce the

TABLE II
RECOGNITION RESULTS ON VIS-NIR DATABASE WITH PROTOCOL I

	PI Image Intensity		PI Image LBP	
	Accuracy	AUC	Accuracy	AUC
LDA	0.9801	0.9648	0.9874	0.9861
CDFE	0.9721	0.9889	0.9973	0.9950
PCA+CCA	0.9542	0.9862	0.9761	0.9929
LCSR	0.9748	0.9919	0.9940	0.9950
KCSR	0.9734	0.9899	0.9887	0.9950
LDSR	0.9754	0.9923	**0.9980**	0.9947
KDSR	**0.9834**	**0.9939**	0.9973	**0.9951**

TABLE III
RECOGNITION RESULTS ON VIS-NIR DATABASE WITH PROTOCOL II

	PII Image Intensity		PII Image LBP	
	Accuracy	AUC	Accuracy	AUC
LDA	0.6451	0.7639	0.7903	0.9106
CDFE	0.5487	0.7850	0.6282	0.9202
PCA+CCA	0.5109	0.8460	0.4612	0.8155
LCSR	0.7565	0.9357	0.9384	0.9925
KCSR	0.7306	0.9353	0.9523	0.9925
LDSR	0.7396	0.9177	0.9404	0.9910
KDSR	**0.7704**	**0.9393**	**0.9533**	**0.9936**

TABLE IV
RECOGNITION RESULTS BASED ON CUHK FACE RECOGNITION DATABASE

	CUHK Intensity		CUHK LBP	
	Accuracy	AUC	Accuracy	AUC
LDA	0.87	0.9803	0.88	0.9707
CDFE	0.75	0.9845	0.67	0.9786
PCA+CCA	0.79	0.9806	0.77	0.9829
LCSR	0.93	0.9944	0.89	**0.9918**
KCSR	0.83	0.9874	0.86	0.9875
LDSR	**0.95**	**0.9971**	**0.90**	0.9914
KDSR	**0.95**	0.9969	**0.90**	0.9903

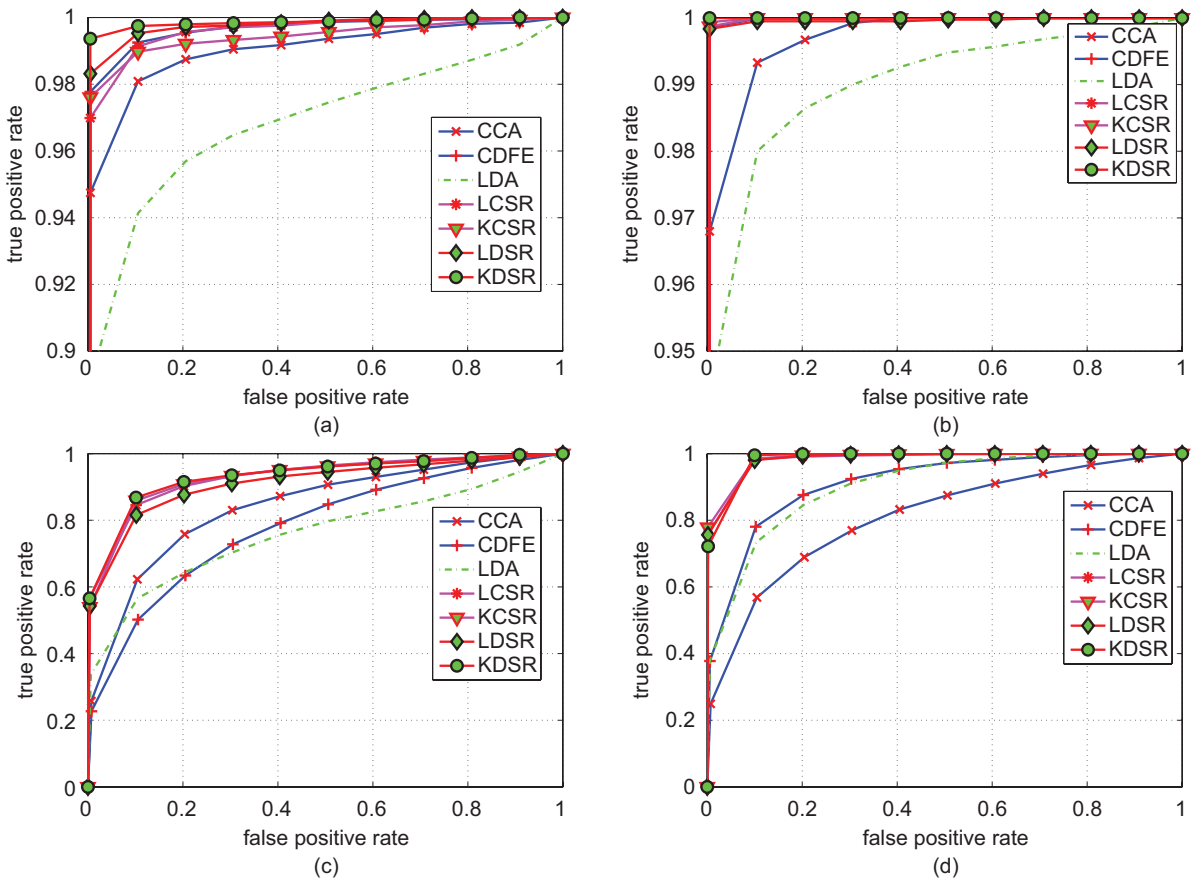


Fig. 4. Receiver operating characteristic (ROC) curves of different methods with four configurations on VIS-NIR database. (a) Image intensity + protocol I. (b) LBP + protocol I. (c) Image intensity + protocol II. (d) LBP + protocol II.

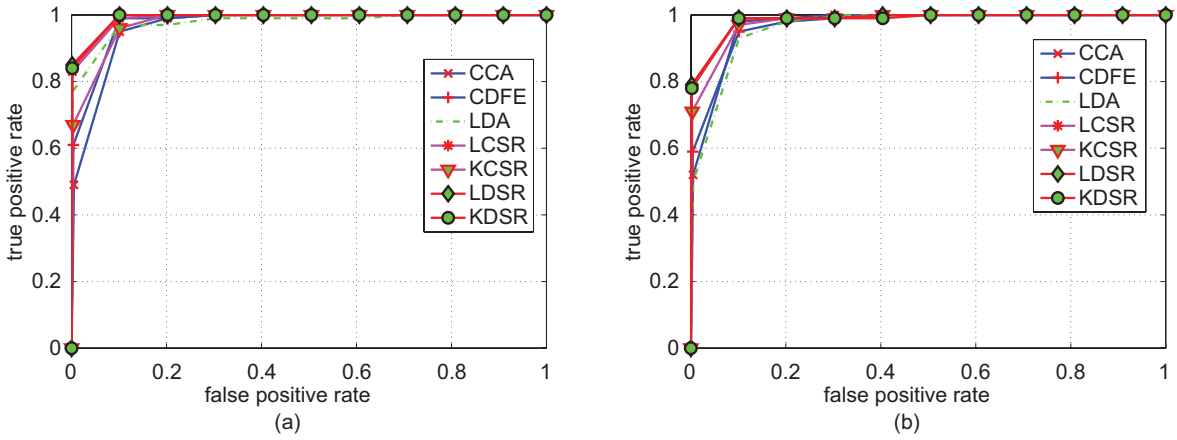


Fig. 5. Receiver operating characteristic (ROC) curves of different methods on photo-sketch database. (a) Image intensity. (b) LBP.

best results accordingly. On this database, linear methods have better results than nonlinear methods. The proposed methods attain 95% rank-1 recognition with the image intensity feature and 90% rank-1 recognition rate using LBP feature. Fig. 5 supports the conclusion that LDSR and KDSR are effective in the photo-sketch matching problems.

4) Feature is a key issue in heterogeneous face matching. For VIS-NIR face matching problems, whether following protocol I or II, the results of compared methods with LBP feature are much better than those with image intensity. This indicates LBP is a good feature descriptor to represent VIS and NIR faces. However, for photo-sketch matching problems, the results of compared methods with image intensity feature are better than those with LBP feature. This indicates image intensity feature is the right choice for photo-sketch matching.

We further conduct experiments to evaluate the impact of the training set size on the recognition accuracy. We randomly select p percent of the data points in each

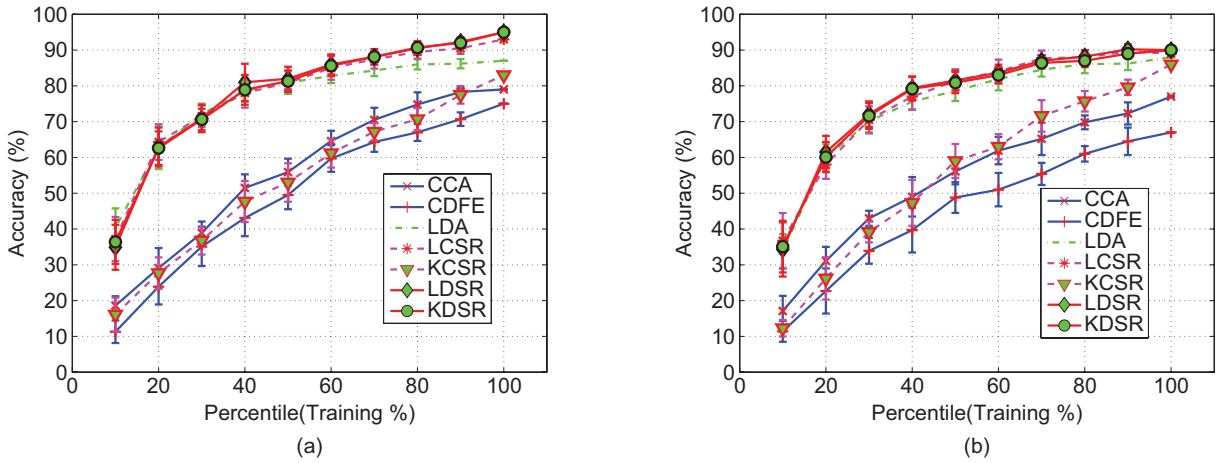


Fig. 6. Rank-1 recognition rate of different methods using various portions of training data on photo-sketch database. (a) Image intensity. (b) LBP.

class from the training set to learn the discriminative mapping, where p changes from 10 to 90. The experiment is repeated 10 times and the mean recognition accuracy with the standard deviation is reported. Fig. 6 shows the recognition results with different sizes of the training set on photo-sketch database. Generally, larger training set produces better recognition performance. The proposed LDSR and KDSR methods demonstrate competitive performance with other methods consistently.

VI. Conclusion

In this paper, we presented a general approach to deal with the heterogeneous face matching problem based on multiple spectral regressions. We extended the method in [7] to the case with multiple capturing ways. Moreover, the proposed new algorithm considers both the within-class information and between-class information. For the former case, we required the lower representations within the same class be as close as possible, while for the latter case, as far as possible. To represent those two cases, we add two regularization terms. The regularization factors can be obtained through regression. We test the new model on some problems with multiple image sources. The experiments results revealed that our new method can solve those problems effectively. It is an important aspect for us to evaluate the performance of the proposed DSR method over data sets that contain varying poses or expressions. However, we do not have heterogeneous face recognition data set that contains such variations. This is left as an open problem for the future, when more challenging data sets are collected.

References

[1] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1991, pp. 586–591.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[3] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[4] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 1208–1213.

[5] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[6] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, 2nd ed. New York: Springer-Verlag, Aug. 2011.

[7] Z. Lei and S. Z. Li, "Coupled spectral regression for matching heterogeneous faces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1123–1128.

[8] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li, "Face matching between near infrared and visible light images," in *Proc. IAPR/IEEE Int. Conf. Biometrics*, Jan. 2007, pp. 523–530.

[9] S. Z. Li, R. Chu, S. Liao, and L. Zhang., "Illumination invariant face recognition using near-infrared images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 627–639, Apr. 2007.

[10] X. Tang and X. Wang, "Face sketch recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 50–57, Jan. 2004.

[11] D. Lin and X. Tang, "Inter-modality face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 13–26.

[12] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Z. Li, "2D–3D face matching using CCA," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–6.

[13] D. Cai, X. He, and J. Han, "Spectral regression for efficient regularized subspace learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[14] C. C. Paige and M. A. Saunders, "Algorithm 583: LSQR: Sparse linear equations and least squares problems," *ACM Trans. Math. Softw.*, vol. 8, no. 2, pp. 195–209, 1982.

[15] K. Muller, S. Mika, G. Riitsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001.

[16] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, no. 3, pp. 337–404, 1950.

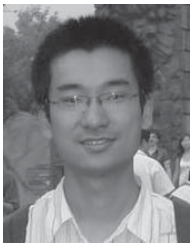
[17] S. Z. Li, Z. Lei, and M. Ao, "The HFB face database for heterogeneous face biometrics research," in *Proc. 6th IEEE Workshop Object Track. Classificat. Beyond Visible Spectrum*, Jun. 2009, pp. 1–8.

[18] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955–1967, Nov. 2009.

[19] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

Xiangsheng Huang received the B.S. degree in material science and the M.S. degree in computer science from Chongqing University, Chongqing, China, in 1998 and 2002, respectively, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2005.

He is currently an Associate Professor with the Institute of Automation, Chinese Academy of Sciences. From 2005 to 2010, he was with the Samsung Advanced Institute of Technology. He has authored or co-authored over 30 papers in international journals and conferences. He holds ten patents. His current research interests include face technology, 3-D imaging and registration, machine learning, and wavelet and filter banks.

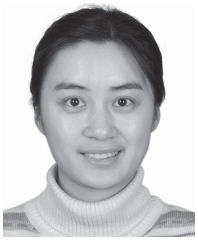


Zhen Lei (M'11) received the B.S. degree in automation from the University of Science and Technology of China, Hefei, China, in 2005, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently an Assistant Professor with the Institute of Automation, Chinese Academy of Sciences. He has authored or co-authored over 40 papers in international journals and conferences. His current research interests include computer vision, pattern recognition, image processing, and face recognition.

Mingyu Fan received the B.Sc. degree from the Central University for Nationalities, Beijing, China, and the Ph.D. degree from the Academy of Mathematics and System Science, Chinese Academy of Sciences, Beijing, in 2006 and 2011, respectively.

He is currently a Lecturer with the Department of Mathematics, Wenzhou University. His current research interests include manifold learning, feature selection, and information retrieval.



Xiao Wang received the Bachelor's degree from Shandong University, Shandong, China, and the Ph.D. degree from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China.

She is currently with the Graduate University of Chinese Academy of Sciences. Her current research interests include optimization methods, and software and applications.

Stan Z. Li (F'09) received the B.Eng. degree from Hunan University, Changsha, China, the M.Eng. degree from the National University of Defense Technology, Changsha, and the Ph.D. degree from Surrey University, Surrey, U.K.

He is currently a Professor with the National Laboratory of Pattern Recognition and the Director of the Center for Biometrics and Security Research, Institute of Automation, and the Director of the Center for Visual Internet of Things Research, Chinese Academy of Sciences, Beijing, China. He was a Researcher with Microsoft Research Asia, Beijing, from 2000 to 2004. He was an Associate Professor with Nanyang Technological University, Singapore. He has authored or co-authored over 200 papers in international journals and conferences, and authored or edited eight books. His current research interests include pattern recognition and machine learning, image and vision processing, face recognition, biometrics, and intelligent video surveillance.

Dr. Li was an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and was the Editor-in-Chief of the *Encyclopedia of Biometrics*. He was a Program Co-Chair for the International Conference on Biometrics in 2007 and 2009, the General Chair for the 9th IEEE Conference on Automatic Face and Gesture Recognition, and has been involved in organizing other international conferences and workshops.