Coupled Spectral Regression for Matching Heterogeneous Faces

Zhen Lei Stan Z. Li*
Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100190, China.
{zlei,szli}@cbsr.ia.ac.cn

Abstract

Face recognition algorithms need to deal with variable lighting conditions. Near infrared (NIR) image based face recognition technology has been proposed to effectively overcome this difficulty. However, it requires that the enrolled face images be captured using NIR images whereas many applications require visual (VIS) images for enrollment templates. To take advantage of NIR face images for illumination-invariant face recognition and allow the use of VIS face images for enrollment, we encounter a new face image pattern recognition problem, that is, heterogeneous face matching between NIR versus VIS faces.

In this paper, we present a subspace learning framework named Coupled Spectral Regression (CSR) to solve this challenge problem of coupling the two types of face images and matching between them. CSR first models the properties of different types of data separately and then learns two associated projections to project heterogeneous data (e.g. VIS and NIR) respectively into a discriminative common subspace in which classification is finally performed. Compared to other existing methods, CSR is computational efficient, benefiting from the efficiency of spectral regression and has better generalization performance. Experimental results on VIS-NIR face database show that the proposed CSR method significantly outperforms the existing methods.

1. Introduction

Face recognition, as a challenging problem in image pattern recognition, has been researched for several decades. Although most current face recognition algorithms are able to achieve satisfactory performance on controlled environments, it is well known that they confront problems arising from poor lighting conditions.

Recently, near infrared (NIR) image based face recognition technology [8] has been developed to overcome illumination related problems encountered in visual (VIS) image based face technology. With active NIR illumination, the NIR approach is not only insensitive to environmental lighting changes, but also greatly increases the accuracy of face recognition for cooperative user applications.

However, the NIR approach performs matching between NIR faces. Therefore, it requires the enrolled face images captured using NIR as well. However, many applications require that enrollment be done using VIS face images. Therefore, the existing NIR based face recognition method is unsuitable in this situation.

A more powerful face technology would take advantage of the NIR approach being invariant to illumination changes while allowing the use of VIS face images to generate face templates. This leads to a new face image pattern recognition problem, that is, heterogeneous face matching between NIR and VIS faces. Solutions to such a problem are significant not only for NIR-VIS face matching but also in face surveillance and face query systems where characteristics of face images captured on spot are usually different or heterogeneous from that of the target faces stored in database.

Some methods have been proposed to deal with the problem. Tang and Wang [10] develop eigen-transform method to synthesis pseudo-sketch image from target photo and perform recognition between pseudo-sketch image and real probe sketch one. This method requires one photo vs. one sketch image under similar condition in training phase. Lin and Tang [9] propose a common discriminant feature extraction (CDFE) method to transform query faces captured using near infrared or sketch images and target faces of visible spectrum into a common discriminant feature subspace, where the difference of with-in scatter matrix and between scatter matrix are maximized. Because of the pairwise manner for with-in and between class scatter matrices computation, the CDFE method is somewhat time-consuming in training phase and only suitable to be applied on small or moderate database. Yi et al. [14] utilize canonical correlation analysis (CCA) to exploit the essential correlations in PCA [11] or LDA [2] subspaces of NIR and VIS im-

*Stan Z. Li is the corresponding author.

ages and Yang et al. [13] propose regularized kernel CCA to learn the relationship between VIS and 3D data spaces. However, their method doesn't consider class label information in CCA process and thus it may drop some important information helpful for classification.

In this paper, we develop a novel and effective subspace learning method, named Coupled Spectral Regression (CSR), for heterogeneous face recognition. As we know, heterogeneous data usually occupy different positions in data space. Using one projection (or feature extraction) to extract features for heterogeneous data would not produce sensible comparison. Therefore, like CDFE, we adopt to use two projections to project the different types of data, respectively, into a common discriminant subspace in which the classification is finally performed. Different from CDFE, we derive the solutions from the view of graph embedding [12] and spectral regression [4] which is easy to be combined with regularization techniques to improve the generalization performance and substantially reduces the computational complexity. The proposed algorithm is both efficient and effective and achieves much better accuracy than existing methods which will be shown in experimental section.

The remainder of this paper is organized as follows. Section 2 briefly reviews the graph view of subspace learning. Section 3 describes the coupled spectral regression approach with its linear and nonlinear implementations in detail. Experimental results on VIS-NIR databases are demonstrated in Section 4 and in Section 5, we conclude the paper.

2. Graph View of Subspace Learning

Recently, Yan et al. [12] unify and re-interpret the subspace learning in a graph embedding framework. Let $G = \{\mathbf{X}, \mathbf{W}\}$ be a graph with vertex set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ and similarity matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$. Each vertex $\mathbf{x}_i$ in the vertex set represents a data point and the elements W_{ij} in the similarity matrix $\mathbf{W}$ is the weight of the edge connecting vertices $\mathbf{x}_i$ and $\mathbf{x}_j$ which describes certain relationship between them. The graph G under such definition can then be used for characterizing various statistical or geometric properties of the data set. The purpose of graph embedding is to find a low-dimension representation for each vertex while preserve the relationship among them.

For one dimension embedding case, suppose $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$ is the low dimension representations of the vertex set $\mathbf{X}$ where y_i represents the low-dimension embedding of $\mathbf{x}_i$. The optimal embedding $\mathbf{y}$ can then be obtained as

$$\mathbf{y} = \arg\min_{\mathbf{y}} \sum_{i,j} ||y_i - y_j||^2 W_{ij} = \arg\min_{\mathbf{y}} \mathbf{y}^T \mathbf{L} \mathbf{y} \quad (1)$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix, and $\mathbf{D}$ is a diagonal matrix where $D_{ii} = \sum_j W_{ij}$. The objective function gives a high penalty when the similar vertices $\mathbf{x}_i$ and $\mathbf{x}_j$ are mapped far apart. Therefore, it tries to preserve the neighborhood structure of the original data points in the low-dimension embedding.

To avoid trivial solutions to above optimization problem, a constraint is then imposed for the embedding, $\mathbf{y}^T \mathbf{D} \mathbf{y} = 1$, thus the minimization problem is finally formulated as

$$\mathbf{y} = \arg\min_{\mathbf{y}^T \mathbf{D} \mathbf{y} = 1} \mathbf{y}^T \mathbf{L} \mathbf{y} = \arg\min_{\mathbf{y}} \frac{\mathbf{y}^T \mathbf{L} \mathbf{y}}{\mathbf{y}^T \mathbf{D} \mathbf{y}} = \arg\max_{\mathbf{y}} \frac{\mathbf{y}^T \mathbf{W} \mathbf{y}}{\mathbf{y}^T \mathbf{D} \mathbf{y}} \quad (2)$$

The solution $\mathbf{y}$ can be obtained by solving the generalized eigen-problem corresponding to the maximum eigenvalue,

$$\mathbf{W}\mathbf{y} = \lambda \mathbf{D}\mathbf{y} \quad (3)$$

The embedding obtained above is only the solution for training data. It is unclear how to map new data from test set into the low-dimension space. To address this problem, it is turned to learn the relationship f between y_i and $\mathbf{x}_i$ instead of embedding y_i, such as $y_i = f(\mathbf{x}_i)$. Under linear assumption $y_i = f(\mathbf{x}_i) = \mathbf{x}_i^T \mathbf{a}$, Eq. 2 can be rewritten as

$$\mathbf{y} = \arg\max_{\mathbf{y}} \frac{\mathbf{y}^T \mathbf{W} \mathbf{y}}{\mathbf{y}^T \mathbf{D} \mathbf{y}} \xrightarrow{\mathbf{y} = \mathbf{X}^T \mathbf{a}} \mathbf{a} = \arg\max_{\mathbf{a}} \frac{\mathbf{a}^T \mathbf{X} \mathbf{W} \mathbf{X}^T \mathbf{a}}{\mathbf{a}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{a}} \quad (4)$$

and the solution $\mathbf{a}$ can be obtained by solving the generalized eigen-problem as

$$\mathbf{X}\mathbf{W}\mathbf{X}^T\mathbf{a} = \lambda \mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{a} \quad (5)$$

Once the relationship is learned from training data, it can be extended to unseen data. The approach mentioned above is called the linearization of graph embedding; linear subspace learning methods such as PCA [11], LDA [2], LPP [7] and NPE [6] etc. can all be interpreted in such framework by defining different similarity matrix $\mathbf{W}$. Specifically, for LDA, the similarity matrix $\mathbf{W}$ can be defined as follows:

$$W_{ij} = \begin{cases} 1/m_t & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to the } t\text{-th class} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where m_t is the number of samples in the t-th classes.

3. Coupled Spectral Regression for Heterogeneous Subspace Learning

Heterogeneous data occupy different positions in observe space. Using the same projection (or feature extraction) to extract features for the two types of data would not produce optimal comparison and it is necessary that different projective directions should be used to map different types of data into a common subspace.

Motivated from this, we propose to use two projections to project the heterogeneous data, respectively, into a common subspace in which the classification is finally performed. Further, we require that such projections are most discriminative for the classification. Moreover, the computation should be efficient. The coupled spectral regression (CSR) is designed to achieve these three goals.

Like spectral regression [4], instead of obtaining the projective directions directly, we could accomplish the heterogeneous subspace learning by two steps: (1) get the common discriminant low embedding; and (2) learn the relationship between data in observe space and the low embedding. For the first step, it is equivalent to get the solution of $\mathbf{y}$ to Eq. 3. Fortunately, under the definition of $\mathbf{W}$ for c class LDA, the solution $\mathbf{y}$ to Eq. 3 in step (1) is straightforward as

$$\mathbf{y}_t = [\underbrace{0,\ldots,0}_{\sum_{i=1}^{t-1} m_i}, \underbrace{1,\ldots,1}_{m_t}, \underbrace{0,\ldots,0}_{\sum_{i=t+1}^{c} m_i}]^T; t = 1,\ldots,c \quad (7)$$

with the same eigenvalue $\lambda = 1$ and can be orthogonalized further by Gram-Shmidit method to get the $c-1$ useful solutions [4]. The left task is to explore the relationship between heterogenous data and the corresponding low embedding respectively.

3.1. Linear Coupled Spectral Regression (LCSR)

Suppose we have the embedding solution $\mathbf{y}$ (e.g. Eq. 7 for LDA) to Eq. 3 for the total data set, where the differences of samples from the same class are minimized while the differences of samples from different classes are maximized. Let $\mathbf{y}^g$ and $\mathbf{y}^p$ be the sub-vectors extracted from $\mathbf{y}$ corresponding to the two types of heterogeneous samples. The purpose of CSR is to explore the relationship between the heterogenous data sets $\{\mathbf{X}^g, \mathbf{X}^p\}$ and their low embedding $\{\mathbf{y}^g, \mathbf{y}^p\}$ respectively. Under linear assumption, this problem is simplified to find the projective vectors $\mathbf{a}^g$ and $\mathbf{a}^p$ for the two types of data to satisfy

$$\mathbf{y}^{gT} = \mathbf{a}^{gT}\mathbf{X}^g, \quad \mathbf{y}^{pT} = \mathbf{a}^{pT}\mathbf{X}^p \quad (8)$$

where $\mathbf{X}^g = [\mathbf{x}_1^g, \ldots, \mathbf{x}_{N^g}^g], \mathbf{X}^p = [\mathbf{x}_1^p, \ldots, \mathbf{x}_{N^p}^p]$ are the two heterogeneous data sets and N^g, N^p denote the sample sizes.

In practice, the solution $\mathbf{a}^g, \mathbf{a}^p$ to the above equations may not exist and a possible way is to solve it in least squares sense. Besides, to avoid the over-fitting problem, several regularization terms are imposed onto the objective function. First, we adopt shrinkage measures as in ridge regression [5] on each of $\mathbf{a}^g, \mathbf{a}^p$ respectively. Second, though we motivate to project heterogeneous data in different directions, both types of face data, e.g. VIS and NIR, come from the same kind of object, hence the projective directions $\mathbf{a}^g$ and $\mathbf{a}^p$ should not differ too much; therefore, we impose an additional regularizer to penalize the difference of $\mathbf{a}^g$ and $\mathbf{a}^p$.

Based on the above discussions, the objective function is formulated as follows

$$\begin{aligned}\{\mathbf{a}^g, \mathbf{a}^p\} = \arg\min_{\mathbf{a}^g,\mathbf{a}^p} & \frac{1}{N^g}||\mathbf{y}^g - \mathbf{X}^{gT}\mathbf{a}^g||^2 \\ & + \frac{1}{N^p}||\mathbf{y}^p - \mathbf{X}^{pT}\mathbf{a}^p||^2 + \eta||\mathbf{a}^g - \mathbf{a}^p||^2 \\ & + \lambda(||\mathbf{a}^g||^2 + ||\mathbf{a}^p||^2)\end{aligned} \quad (9)$$

where the first two terms are the approximation errors; the third one is the penalty that prevents the coupled projective directions differing too much or over-fitting the training data; and the last two ones are the shrinkage constraints, also known the Tikhonov regularizers [5], which help improve the generalization of the solutions. η and λ are the two parameters that control the balance between the fitting accuracy and the generalization performance. Through putting the derivatives of the objective function with respect to the projective vectors to zero, we obtain

$$\begin{aligned}(\mathbf{X}^g\mathbf{X}^{gT} + N^g(\lambda+\eta)\mathbf{I})\mathbf{a}^g &= \mathbf{X}^g\mathbf{y}^g + N^g\eta\mathbf{a}^p \\ (\mathbf{X}^p\mathbf{X}^{pT} + N^p(\lambda+\eta)\mathbf{I})\mathbf{a}^p &= \mathbf{X}^p\mathbf{y}^p + N^p\eta\mathbf{a}^g\end{aligned} \quad (10)$$

By solving the above two equations, we can finally get the projective vectors $\mathbf{a}^g$ and $\mathbf{a}^p$ as

$$\begin{aligned}\mathbf{a}^g &= [\Phi^g - N^gN^p\eta^2(\Phi^p)^{-1}]^{-1}[\mathbf{X}^g\mathbf{y}^g + N^g\eta(\Phi^p)^{-1}\mathbf{X}^p\mathbf{y}^p] \\ \mathbf{a}^p &= [\Phi^p - N^gN^p\eta^2(\Phi^g)^{-1}]^{-1}[\mathbf{X}^p\mathbf{y}^p + N^p\eta(\Phi^g)^{-1}\mathbf{X}^g\mathbf{y}^g]\end{aligned} \quad (11)$$

where Φ^g, Φ^p are defined as

$$\Phi^g = \mathbf{X}^g\mathbf{X}^{gT} + N^g(\lambda+\eta)\mathbf{I}, \ \Phi^p = \mathbf{X}^p\mathbf{X}^{pT} + N^p(\lambda+\eta)\mathbf{I} \quad (12)$$

The method described above is called linear coupled spectral regression (LCSR). Naturally, corresponding to d useful eigenvectors $\{\mathbf{y}_t\}_{t=1}^d$ to Eq. 3, we can get d couples of projective directions $\mathbf{A}^g = [\mathbf{a}_1^g, \mathbf{a}_2^g, \ldots, \mathbf{a}_d^g]$ and $\mathbf{A}^p = [\mathbf{a}_1^p, \mathbf{a}_2^p, \ldots, \mathbf{a}_d^p]$ to map the two types of heterogeneous data respectively into a discriminative common subspace to be classified.

3.2. Kernel based Coupled Spectral Regression (KCSR)

Face appearances usually lie in a nonlinear low-dimensional manifold. It is reasonable to utilize nonlinear embedding to achieve good classification performance. Following similar idea proposed in SVM [3], the CSR method can also be integrated with kernel trick to kernelize data into an implicit high or even infinite dimension feature space to make the problem better solved.

Suppose the two heterogeneous data sets in original space $\mathcal{R}^n$ be $\mathbf{X}^g = [\mathbf{x}_1^g, \ldots, \mathbf{x}_{N^g}^g], \mathbf{X}^p = [\mathbf{x}_1^p, \ldots, \mathbf{x}_{N^p}^p]$,

and their corresponding mapped data in feature space $\mathcal{F}$ are denoted as $\phi(\mathbf{X}^g) = [\phi(\mathbf{x}_1^g), \ldots, \phi(\mathbf{x}_{N^g}^g)], \phi(\mathbf{X}^p) = [\phi(\mathbf{x}_1^p), \ldots, \phi(\mathbf{x}_{N^p}^p)]$, where ϕ is the mapping function $\phi : \mathcal{R}^n \rightarrow \mathcal{F}$. Our purpose is to find the two projective vectors $\mathbf{w}^g$ and $\mathbf{w}^p$ in feature space $\mathcal{F}$ respectively to satisfy:

$$\mathbf{y}^{gT} = \mathbf{w}^{gT}\phi(\mathbf{X}^g),\ \mathbf{y}^{pT} = \mathbf{w}^{pT}\phi(\mathbf{X}^p) \tag{13}$$

From the reproducing kernel theory, we know that $\mathbf{w}^g \in span\phi(\mathbf{X}^g)$,$\mathbf{w}^p \in span\phi(\mathbf{X}^p)$, which means

$$\mathbf{w}^g{=}\sum_{i=1}^{N^g} a_i^g\phi(\mathbf{x}_i^g){=}\phi(\mathbf{X}^g)\mathbf{a}^g,\ \mathbf{w}^p{=}\sum_{i=1}^{N^p} a_i^p\phi(\mathbf{x}_i^p){=}\phi(\mathbf{X}^p)\mathbf{a}^p \tag{14}$$

where $\mathbf{a}^g = [a_1^g, a_2^g, \ldots, a_{N^g}^g]^T$ and $\mathbf{a}^p = [a_1^p, a_2^p, \ldots, a_{N^p}^p]^T$. Substituting Eq. 14 into Eq. 13, we have

$$\mathbf{y}^{gT}{=}\mathbf{a}^{gT}\phi(\mathbf{X}^g)^T\phi(\mathbf{X}^g),\ \mathbf{y}^{pT}{=}\mathbf{a}^{pT}\phi(\mathbf{X}^p)^T\phi(\mathbf{X}^p) \tag{15}$$

Similar to LCSR, we derive the solutions in least squares sense and impose regularized items to avoid the over-fitting phenomenon. Using the kernel tricks like in SVM [3], we define three kernel matrices $\mathbf{K}^{gg} = \phi(\mathbf{X}^g)^T\phi(\mathbf{X}^g)$, $\mathbf{K}^{pp} = \phi(\mathbf{X}^p)^T\phi(\mathbf{X}^p)$, $\mathbf{K}^{gp} = \phi(\mathbf{X}^g)^T\phi(\mathbf{X}^p)$. Thus, the objective function is formulated as

$$\begin{aligned}\{\mathbf{a}^g, \mathbf{a}^p\} = &\arg\min_{\mathbf{a}^g,\mathbf{a}^p} \frac{1}{N^g}||\mathbf{y}^g{-}\mathbf{K}^{gg}\mathbf{a}^g||^2{+}\frac{1}{N^p}||\mathbf{y}^p{-}\mathbf{K}^{pp}\mathbf{a}^p||^2 \\ &+ (\eta+\lambda)(\mathbf{a}^{gT}\mathbf{K}^{gg}\mathbf{a}^g + \mathbf{a}^{pT}\mathbf{K}^{pp}\mathbf{a}^p) \\ &- \eta(\mathbf{a}^{gT}\mathbf{K}^{gp}\mathbf{a}^p + \mathbf{a}^{pT}\mathbf{K}^{gpT}\mathbf{a}^g)\end{aligned} \tag{16}$$

where λ and η are the two parameters that control the trade-off between fitting accuracy and generalization performance like LCSR. By putting the derivatives of above function with respect to $\mathbf{a}^g$ and $\mathbf{a}^p$ to zero, we get the final solution $\mathbf{a}^g$ and $\mathbf{a}^p$ as

$$\begin{aligned}\mathbf{a}^g =&[\Theta^g - N^gN^p\eta^2\mathbf{K}^{gp}(\Theta^p)^{-1}\mathbf{K}^{gpT}]^{-1} \\ &\cdot[\mathbf{K}^{gg}\mathbf{y}^g + N^g\eta\mathbf{K}^{gp}(\Theta^p)^{-1}\mathbf{K}^{pp}\mathbf{y}^p] \\ \mathbf{a}^p =&[\Theta^p - N^gN^p\eta^2\mathbf{K}^{gpT}(\Theta^g)^{-1}\mathbf{K}^{gp}]^{-1} \\ &\cdot[\mathbf{K}^{pp}\mathbf{y}^p + N^p\eta\mathbf{K}^{gpT}(\Theta^g)^{-1}\mathbf{K}^{gg}\mathbf{y}^g]\end{aligned} \tag{17}$$

where Θ^g and Θ^p are defined as

$$\Theta^g{=}[\mathbf{K}^{gg}{+}N^g(\lambda+\eta)\mathbf{I}]\mathbf{K}^{gg},\ \Theta^p{=}[\mathbf{K}^{pp}{+}N^p(\lambda+\eta)\mathbf{I}]\mathbf{K}^{pp} \tag{18}$$

Consequently, in test phase, for every input heterogeneous data pair $(\mathbf{z}^g, \mathbf{z}^p)$, it can be nonlinearly projected as:

$$\tilde{\mathbf{z}}^g = \sum_{i=1}^{N^g} a_i^g k(\mathbf{z}^g, \mathbf{x}_i^g),\ \tilde{\mathbf{z}}^p = \sum_{i=1}^{N^p} a_i^p k(\mathbf{z}^p, \mathbf{x}_i^p) \tag{19}$$

where $k(.)$ is the kernel function.

Naturally, corresponding to d useful eigenvectors $\{\mathbf{y}_t\}_{t=1}^d$ to Eq. 3, we can get d couples of projective directions $\mathbf{A}^g = [\mathbf{a}_1^g, \mathbf{a}_2^g, \ldots, \mathbf{a}_d^g]$ and $\mathbf{A}^p = [\mathbf{a}_1^p, \mathbf{a}_2^p, \ldots, \mathbf{a}_d^p]$ to project the heterogeneous data nonlinearly into a common discriminant subspace to be classified.

4. Experimental Results

The following experiments evaluate the proposed CSR methods in comparison with several existing methods of PCA [11], LDA [2], CDFE [9], PCA+CCA, LDA+CCA [14]. In classification phase, the cosine distance is adopt to measure the dissimilarity of feature in subspace and the nearest neighbor (NN) classifier is chosen to do the classification task.

Specifically, in CSR methods, the similarity matrix $\mathbf{W}$ is defined following LDA according to Eq. 6 in this experiment and hence the embedding solution $\mathbf{y}$ has the formula of Eq. 7. Gaussian kernel is utilized in KCSR. For LCSR and KCSR method, the regularized coefficients $\{\lambda, \eta\}$ are empirically set to $\{0.001, 0.01\}$ and $\{0.00005, 0.0005\}$ respectively. For other methods, the parameters are optimized according to the recommended values in their papers. For PCA and LDA, we combine the heterogenous data together and train a unique projection matrix for heterogenous data.

A VIS-NIR database is collected. There are totally 5097 images, including 2095 VIS and 3002 NIR ones from 202 persons in the database. Two test protocols are designed to evaluate different methods, where the database is split into training set and test set randomly. In protocol I, the training set contains 1062 VIS and 1487 NIR images from 202 subjects and the left ones constitute the test set. The persons in test set are all contained in training set. In protocol II, there are 1438 VIS and 1927 NIR images from 168 persons in training set and the test set contains 657 VIS and 1075 NIR ones from 174 persons. The persons in test set are partially contained in training set. All the images are cropped into 128×128 size according to the automatically detected eye coordinates. Fig. 1 shows some cropped VIS and NIR face examples from the database.

Two kinds of features are adopted as the input for different algorithms. One is intensity feature where each image is resized into 32×32 and transformed to form a 1024 dimension feature. The other is LBP [1] feature in which 1000 dimension LBP features are extracted to represent each face image. Consequently, there are four combinations of different feature types and test protocols for each algorithm in experiment. The results are reported in terms of three indices: rank-1 recognition rate, verification rate (VR) when the false accept rate (FAR) is 0.001 and equal error rate (EER).

Tables 1 - 4 illustrate the results of different configurations. Tables 1 and 2 illustrate the results of image inten-

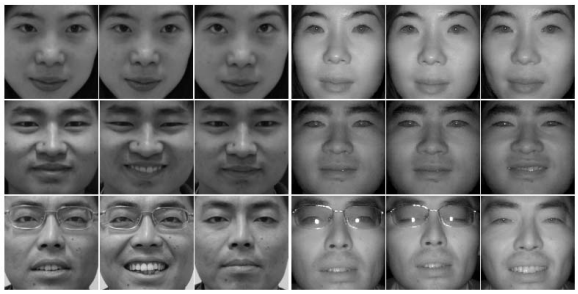

Figure 1. VIS and NIR face example images. (The left three columns are VIS images and the right three columns are NIR images.)

Table 1. Results based on image intensity in protocol I.

Method	Rank-1	VR@FAR=0.001	EER
PCA	0.1208	0.0271	0.3305
LDA	**0.9801**	0.8725	0.0609
CDFE	0.9721	0.9534	0.0201
PCA+CCA	0.9542	0.9196	0.0313
LDA+CCA	0.9774	0.9199	0.0310
LCSR	0.9748	0.9542	0.0191
KCSR	0.9734	**0.9653**	**0.0180**

Table 2. Results based on LBP features in protocol I.

Method	Rank-1	VR@FAR=0.001	EER
PCA	0.8061	0.3887	0.1091
LDA	0.9874	0.9268	0.0301
CDFE	**0.9973**	**0.9972**	**0.0017**
PCA+CCA	0.9695	0.9399	0.0214
LDA+CCA	0.9801	0.9579	0.0193
LCSR	0.9940	0.9795	0.0067
KCSR	0.9887	0.9775	0.0087

Table 3. Results based on image intensity in protocol II.

Method	Rank-1	VR@FAR=0.001	EER
PCA	0.1183	0.0124	0.3358
LDA	0.6451	0.2505	0.2890
CDFE	0.5487	0.1668	0.1927
PCA+CCA	0.5109	0.1640	0.1959
LDA+CCA	0.4155	0.1156	0.2062
LCSR	**0.7565**	0.4471	0.1310
KCSR	0.7306	**0.4473**	**0.1254**

sity and LBP features respectively following protocol I, and tables 3 and 4 are the corresponding results in protocol II. Fig. 2 plots the ROC curves of different methods with different configurations. For clarity, we omit the results of PCA and LDA. From the results, we can observe:

1. In protocol I, the performances of various methods are close. As a whole, the methods using different projective directions for heterogeneous data (CDFE, PCA+CCA, LDA+CCA, LCSR, KCSR) achieve higher accuracy than the unique one (PCA and LDA). It proves that it is feasible and effective to project heterogeneous data respectively into a common discriminant subspace.

2. For the previous methods like CDFE, PCA+CCA and LDA+CCA, they achieve satisfactory accuracy in protocol I. However, their performance degrades dramatically in protocol II, even much worse than PCA in the case of LBP feature. It is a severe problem in practice because in many cases, it is impossible to get images of all subjects to be trained. Comparatively, the proposed CSR methods, including LCSR and KCSR, owe to the effective regularization techniques, achieve significantly better results in protocol II where the subjects in training set and test set are not totally over-

Table 4. Results based on LBP features in protocol II.

Method	Rank-1	VR@FAR=0.001	EER
PCA	0.8221	0.3802	0.1119
LDA	0.7903	0.2644	0.1747
CDFE	0.6282	0.2615	0.1539
PCA+CCA	0.4612	0.1417	0.2479
LDA+CCA	0.3519	0.1008	0.3924
LCSR	0.9384	0.7305	0.0416
KCSR	**0.9523**	**0.7695**	**0.0373**

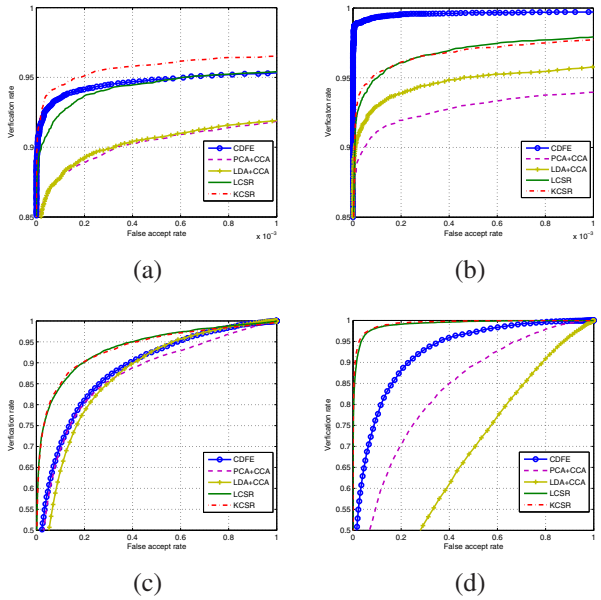

Figure 2. Receiver operating characteristic (ROC) curves of different methods with four configurations ((a) Image intensity + protocol I, (b) LBP + protocol I, (c) Image intensity + protocol II, (d) LBP + protocol II).

lapped and thus are more reliable and applicable in real world.

3. For most methods, whether following protocol I or II, the results with LBP feature are better than those with image intensity which suggests LBP is a good feature descriptor to represent faces. The kernel based method KCSR achieves better result than LCSR which shows the effectiveness of nonlinear kernel technology. In all of the methods, the KCSR with LBP feature achieves the best heterogeneous face recognition accuracy in protocol II.

Table 5. Computational time (s) of different methods with 1024 dimension intensity feature on training set A (protocol I) and training set B (protocol II).

	CDFE	PCA+CCA	LDA+CCA	LCSR
Training Set A	77.98 (s)	22.29 (s)	30.47 (s)	5.07 (s)
Training Set B	73.26 (s)	21.85 (s)	26.50 (s)	4.76 (s)

Table 5 illustrates the experimental computational cost of different methods on two training sets of protocol I and II. These four are all linear feature extraction methods, so the results are comparable in linear sense. Each reported result is the average of 20 running times with matlab code on a Core 2 Duo 2.4GHz and 2GB RAM PC. It can be seen the proposed LCSR method has the lowest computational cost, less than 1/15 of CDFE, 1/4 of PCA+CCA and 1/5 of LDA+CCA which indicates the computational efficiency of the CSR method.

5. Conclusions

In this paper, we have developed the couple spectral regression (CSR) as an effective and efficient framework for matching heterogeneous faces. The motivation is that heterogeneous face data occupy different positions in a feature space. Therefore, it is necessary that different projective directions should be used to map the heterogeneous data into a common subspace. The couple spectral regression framework provides such a way to solve this problem that it can integrate regularization technology to improve the generalization performance effectively and meanwhile greatly reduce the computational expenses. While this work deals with the two-modal heterogeneous data case, the CSR method can be extended to couple more than two modalities and this is a future direction to explore.

Acknowledgements

This work was supported by the following funding resources: National Natural Science Foundation Project #60518002, National Science and Technology Support Program Project #2006BAK08B06, National Hi-Tech (863) Program Projects #2006AA01Z192, #2006AA01Z193, and #2008AA01Z124, Chinese Academy of Sciences 100 people project, and AuthenMetric R&D Funds.

References

[1] T. Ahonen, A. Hadid, and M.Pietikainen. Face recognition with local binary patterns. In *Proceedings of the European Conference on Computer Vision*, pages 469–481, Prague, Czech, 2004. 4

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997. 1, 2, 4

[3] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, Pittsburgh, Pennsylvania, July 27-29 1992. 3, 4

[4] D. Cai, X. He, and J. Han. Spectral regression for efficient regularized subspace learning. In *Proceedings of IEEE International Conference on Computer Vision*, 2007. 2, 3

[5] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, 2001. 3

[6] X. He, D. Cai, S. Yan, and H. Zhang. Neighborhood preserving embedding. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1208–1213, 2005. 2

[7] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005. 2

[8] S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):627–639, April 2007. 1

[9] D. Lin and X. Tang. Inter-modality face recognition. In *Proceedings of the European Conference on Computer Vision*, pages 13–26, 2006. 1, 4

[10] X. Tang and X. Wang. Face sketch recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):50–57, 2004. 1

[11] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591, Hawaii, June 1991. 1, 2, 4

[12] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007. 2

[13] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Z. Li. 2d-3d face matching using cca. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2008. 2

[14] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li. Face matching between near infrared and visible light images. In *Proceedings of IAPR/IEEE International Conference on Biometrics*, pages 523–530, 2007. 1, 4