Coupled Discriminant Analysis for Heterogeneous Face Recognition

Zhen Lei, *Member, IEEE*, Shengcai Liao, Anil K. Jain, *Fellow, IEEE*, and Stan Z. Li, *Fellow, IEEE*

***Abstract*—Coupled space learning is an effective framework for heterogeneous face recognition. In this paper, we propose a novel coupled discriminant analysis method to improve the heterogeneous face recognition performance. There are two main advantages of the proposed method. First, all samples from different modalities are used to represent the coupled projections, so that sufficient discriminative information could be extracted. Second, the locality information in kernel space is incorporated into the coupled discriminant analysis as a constraint to improve the generalization ability. In particular, two implementations of locality constraint in kernel space (LCKS)-based coupled discriminant analysis methods, namely LCKS-coupled discriminant analysis (LCKS-CDA) and LCKS-coupled spectral regression (LCKS-CSR), are presented. Extensive experiments on three cases of heterogeneous face matching (high versus low image resolution, digital photo versus video image, and visible light versus near infrared) validate the efficacy of the proposed method.**

***Index Terms*—Face recognition, heterogeneous face recognition, coupled discriminant analysis, coupled spectral regression, locality constraint in kernel space.**

I. Introduction

FACE recognition has attracted much attention due to its potential value in security and law enforcement applications and its theoretical challenges. Although face recognition under controlled environments has been well addressed, its performance in many real world applications is still far from satisfactory. One of the main problems is that often the quality (modality) of probe images and gallery images differs so much that the face recognition performance does not meet the expected performance. For example, in the video surveillance scenario, the gallery images are usually high resolution photos, while the probe images are of low resolution and some times in near infrared (NIR) modality. In law enforcement applications, sketch images are usually used to compare with photos. These factors introduce a number of challenges for face recognition. Matching face images of different modalities is referred to as heterogeneous face recognition [1]. Fig. 1 shows some common heterogeneous face matching scenarios in real applications.

Fig. 1. Examples of heterogeneous face image matching. From left to right: visual light (VIS) versus near infrared (NIR), VIS versus 3-D depth, digital photo versus video, and photo versus sketch.

A number of face representation approaches have been introduced, including subspace based holistic features and local appearance features [2], [3]. Typical holistic features include the well known Principal Component Analysis (PCA) [4], Linear Discriminate Analysis (LDA) [5] and their many extensions like [6], [7]. Local appearance features, like Gabor [8], [9], local binary patterns (LBP) [10] and their combination [11], [12], have been shown to be more robust to illumination, expression and pose variations than holistic appearance features.

The framework of combining local features and holistic features (subspace learning) is one of state-of-the-art approaches in face recognition [3]. Generally speaking, the pipeline of this framework can be roughly divided into three stages (Fig. 2). First, face images are normalized in terms of their size and intensity. Second, effective local features robust to face variations are extracted. Finally, a discriminant subspace is learned for classification. Following this methodology, the purpose of heterogeneous face recognition can be formulated as finding a discriminant subspace which makes different classes separable for heterogeneous data.

Manuscript received March 26, 2012; revised June 21, 2012; accepted July 10, 2012. Date of publication July 24, 2012; date of current version November 15, 2012. This work was supported in part by the Chinese National Natural Science Foundation Project 61070146, 61105023, 61103156, 61105037, in part by the National IoT R&D Project 2150510, in part by Chinese Academy of Sciences Project KGZD-EW-102-2, in part by the Chinese Academy of Sciences Visiting Professorship for Senior International Scientists under Grant 2011T1G18, in part by European Union FP7 Project 257289 (TABULA RASA), and in part by AuthenMetric R&D Funds. The work of A. K. Jain was supported in part by the World Class University (WCU) program funded by the Ministry of Education, Science and Technology through the National Research Foundation of Korea (R31-10008). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fabio Scotti.

Z. Lei and S. Z. Li are with Center for Biometrics and Security Research, and 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).

S. Liao is with Michigan State University, East Lansing, MI 48824 USA (e-mail: scliao@msu.edu).

A. K. Jain is with Michigan State University, East Lansing, MI 48824 USA, and also with the Department of Brain and Cognitive Engineering, Korea University, Anamdong, Seongbukgu, Seoul 136-713, Republic of Korea (e-mail: jain@cse.msu.edu).

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/TIFS.2012.2210041

Fig. 2. Three stages in face recognition based on the framework of local features and subspace learning.

Different from traditional face recognition (visible to visible band face matching with similar image quality), the difficulty in heterogeneous face recognition mainly comes from the appearance differences between face images of different modalities. One intuitive idea is to reduce the difference in appearance from different modalities so that the heterogeneous face recognition performance can be improved. According to the processing pipeline in Fig. 2, the efforts in heterogeneous face recognition methods can be divided into three categories. In the first category, methods focus on the process in the first stage, which are usually called analysis by synthesis methods. The face samples of one modality are first transformed to another modality so that the face appearance difference is minimized. Traditional face recognition methods can be then applied to the samples of the same modality. The representative works in this category include [13]–[15]. Tang and Wang [13] developed the eigen-transform method to synthesize a sketch image from a target photo and then performed recognition between pseudo-sketch image and real probe sketch. Liu *et al.* [14] proposed a local linear preserving method to synthesize sketch images from photos and then used a nonlinear discriminant analysis to recognize the sketches. Wang and Tang [15] utilized MRF modeling to synthesize sketch/photo from photo/sketch images.

In the second category, researchers pay attention to the second stage and try to extract consistent features from heterogeneous face images. Proper texture descriptors are designed and applied to the heterogeneous images to reduce the feature gap between them. Liao *et al.* [16] first utilized the difference of Gaussian (DoG) filter to preprocess the visible light and near infrared images to reduce the appearance difference and then extracted multiblock local binary pattern (MBLBP) to represent faces. Klare and Jain [17] used HoG and LBP descriptors and learnt an ensemble of discriminant projections. In the matching phase, they incorporated sparse representation classifier (SRC) to improve the performance of heterogeneous face recognition. In [18], researchers proposed to extract SIFT and multiscale local binary patterns (MLBP) features from forensic sketches and mug shot photos, respectively. Multiple discriminant projections are then learned to improve the performance of forensic sketch-photo matching. Zhang *et al.* [19] proposed a learning based coupled information-theoretic encoding descriptor to capture a discriminant local structure for photo-sketch images and applied PCA+LDA classifier to compute the dissimilarity of samples. All the above methods try to reduce the gap between heterogeneous face images at the feature level and apply traditional face classification methods to realize the recognition task.

Methods in the third category focus on the subspace learning stage and try to find a common discriminant subspace to classify heterogeneous data. Coupled projections for samples from different modalities are learned and used to project the data onto the common discriminant subspace. Lin and Tang [20] proposed

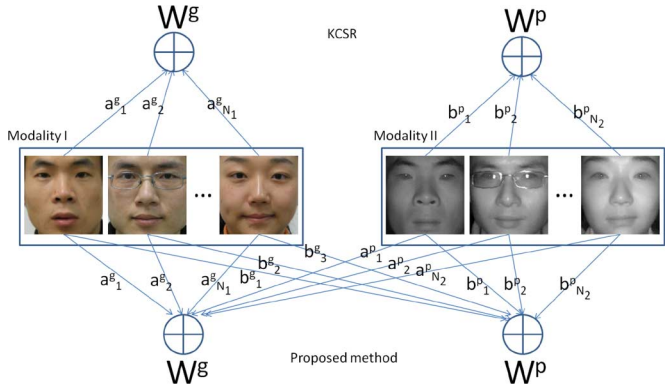


Fig. 3. Difference in projection representations between the proposed method and KCSR [23]. In KCSR, the projections for modality I or II are represented based on the data from modality I or II, respectively. In the proposed method, the coupled projections are constructed based on all the data from both modalities I and II.

a common discriminant feature extraction (CDFE) method to transform query faces captured using near infrared or sketch images and target faces of visible spectrum onto a common discriminant feature subspace, where the ratio of between scatter matrix to within scatter matrix is maximized. Although CDFE achieves high recognition rate on training set, its generalization performance is poor. Yi *et al.* [21] utilized canonical correlation analysis (CCA) to exploit the essential correlations in PCA [4] and LDA [5] subspaces of VIS and NIR images and Yang *et al.* [22] proposed regularized kernel CCA to learn the relationship between VIS and 3-D face data spaces. However, their method does not consider class label information in CCA and thus it does not fully utilize the discriminative information helpful in classification. Lei and Li [23] proposed the coupled spectral regression (CSR) method to deal with heterogeneous face recognition problem and achieved a better generalization performance than previous methods.

This work belongs to the third category. There are three contributions of this paper. (i) Unlike previous work (e.g., [23]), where the projections for different heterogeneous modalities are constructed by the samples from the corresponding modality, we not only use the samples from the same modality, but also samples from another modality for the coupled projections. Therefore, we use all the samples from multiple modalities to form the coupled projections. Fig. 3 gives a visualization of the differences in the projection representation between the proposed method and the previous ones. (ii) The existing methods do not adequately explore the locality information in the kernel space. In this work, we explore the locality information in the kernel space and incorporate it as a constraint into the discriminant analysis process to improve the generalization ability of the derived subspace. (iii) we present two locality constraint in kernel space (LCKS) based methods with discriminant analysis and spectral regression frameworks, namely LCKS based coupled discriminant analysis (LCKS-CDA) and LCKS based coupled spectral regression (LCKS-CSR), respectively, to deal with heterogeneous face matching problem. Preliminary results of this work have been published in [24].

The remainder of this paper is organized as follows. Sections II and III describe locality constraint in kernel space

and provide details of the two algorithms, namely LCKS-CDA and LCKS-CSR. Extensive experimental results and discussions on three different heterogeneous face matching scenarios are presented in Section IV. In Section V we conclude the paper.

II. Locality Constraint Based Coupled Discriminant Analysis

Without loss of generality and for ease of representation, we take VIS versus NIR faces matching as an example to describe our algorithm. Denote the sample data of the two modalities as $X^g = [X_1^g, X_2^g, \ldots, X_{N_g}^g]$ and $X^p = [X_1^p, X_2^p, \ldots, X_{N_p}^p]$, where g and p are indicators of different modalities and N_g and N_p are the number of samples. We first perform a nonlinear mapping to transform the data to a high-dimensional kernel space where the data is usually considered to be more linearly separable. $\Phi^g = [\phi(X_1^g), \phi(X_2^g), \ldots, \phi(X_{N_g}^g)]$ and $\Phi^p = [\phi(X_1^p), \phi(X_2^p), \ldots, \phi(X_{N_p}^p)]$ denote the data in the transformed kernel space. The purpose of coupled discriminant analysis is to find coupled projections in the kernel space, with which the transformed samples are projected onto a common discriminant subspace, where the low-dimensional embeddings are well classified. According to [25], the projections learned from the training samples lie in the space spanned by the training samples. Therefore, the projections W^g and W^p for VIS and NIR faces can be linearly represented as $W^g = \sum_{i=1}^{N_g} \alpha_i^g \phi(X_i^g) + \sum_{i=1}^{N_p} \alpha_i^p \phi(X_i^p)$ and $W^p = \sum_{i=1}^{N_g} \beta_i^g \phi(X_i^g) + \sum_{i=1}^{N_p} \beta_i^p \phi(X_i^p)$. Note that in previous methods like CDFE [20] and KCSR [23], the projections of VIS/NIR faces are supposed to be linearly represented by the face data belonging to VIS/NIR, respectively. That is $W^g = \sum_{i=1}^{N_g} \alpha_i^g \phi(X_i^g)$ and $W^p = \sum_{i=1}^{N_p} \beta_i^p \phi(X_i^p)$. In our formulation, all the samples from VIS and NIR make contributions to the coupled projections. Therefore, all available information contained in samples between VIS and NIR is utilized. Denoting $\Phi = [\Phi^g, \Phi^p]$, $A = [\alpha_1^g, \ldots, \alpha_{N_g}^g, \alpha_1^p, \ldots, \alpha_{N_p}^p]^T$, $B = [\beta_1^g, \ldots, \beta_{N_g}^g, \beta_1^p, \ldots, \beta_{N_p}^p]^T$, we have

$$\begin{aligned} W^g &= \Phi A \\ W^p &= \Phi B \end{aligned} \tag{1}$$

Similar to LDA [5], we can define the between and within class scatter in the reduced common subspace as follows:

$$\begin{aligned} S_b &= S_b^{gg} + S_b^{gp} + S_b^{pg} + S_b^{pp} \\ S_w &= S_w^{gg} + S_w^{gp} + S_w^{pg} + S_w^{pp} \end{aligned} \tag{2}$$

where S_b^{gp} and S_w^{gp} are defined as

$$\begin{aligned} S_b^{gp} &= \sum_{i=1}^{C} n_i^g \left(W^{gT} m_i^g - W^{pT} m^p\right) \\ &\quad \times \left(W^{gT} m_i^g - W^{pT} m^p\right)^T \\ S_w^{gp} &= \sum_{i=1}^{C} \sum_{j \in L_i^g} \left(W^{gT} \phi\left(X_j^g\right) - W^{pT} m_i^p\right) \\ &\quad \cdot \left(W^{gT} \phi\left(X_j^g\right) - W^{pT} m_i^p\right)^T \end{aligned} \tag{3}$$

where C is the number of classes; L_i^g is the index set of samples belonging to the ith class in modality g; n_i^g is the number of samples in the ith class for modality g; m^g, m^p are the mean vectors for modality g and p in transformed kernel space and m_i^g, m_i^p are the corresponding mean vectors for the ith class in the kernel space.

Substituting (1) into (3), we have

$$\begin{aligned} S_b^{gp} &= \sum_{i=1}^{C} n_i^g \left(A^T \mu_i^g - B^T \mu^p\right) \left(A^T \mu_i^g - B^T \mu^p\right)^T \\ &= \sum_{i=1}^{C} n_i^g \Big[A^T \mu_i^g \mu_i^{gT} A + B^T \mu^p \mu^{pT} B \\ &\quad - A^T \mu_i^g \mu^{pT} B - B^T \mu^p \mu_i^{gT} A\Big] \\ &= A^T \left(\sum_{i=1}^{C} n_i^g \mu_i^g \mu_i^{gT}\right) A + B^T (N_g \mu^p \mu^{pT}) B \\ &\quad - A^T \left(\sum_{i=1}^{C} n_i^g \mu_i^g \mu^{pT}\right) B - B^T \left(\sum_{i=1}^{C} n_i^g \mu^p \mu_i^{gT}\right) A \\ S_w^{gp} &= \sum_{i=1}^{C} \sum_{j \in L_i^g} \left(A^T \zeta_j^g - B^T \mu_i^p\right) \left(A^T \zeta_j^g - B^T \mu_i^p\right)^T \\ &= \sum_{i=1}^{C} \sum_{j \in L_i^g} \Big[A^T \zeta_j^g \zeta_j^{gT} A + B^T \mu_i^p \mu_i^{pT} B - A^T \zeta_j^g \mu_i^{pT} B \\ &\quad - B^T \mu_i^p \zeta_j^{gT} A\Big] \\ &= A^T \left(\sum_{i=1}^{C} \sum_{j \in L_i^g} \zeta_j^g \zeta_j^{gT}\right) A + B^T \left(\sum_{i=1}^{C} n_i^g \mu_i^p \mu_i^{pT}\right) B \\ &\quad - A^T \left(\sum_{i=1}^{C} n_i^g \mu_i^g \mu_i^{pT}\right) B - B^T \left(\sum_{i=1}^{C} n_i^g \mu_i^p \mu_i^{gT}\right) A \end{aligned} \tag{4}$$

where $\zeta_j^g = [k(X_j^g, X_1^g), \ldots, k(X_j^g, X_{N_g}^g), k(X_j^g, X_1^p) \cdots, k(X_j^g, X_{N_p}^p)]^T$, in which $k(X_1, X_2)$ is the inner product function between $\phi(X_1)$ and $\phi(X_2)$; μ_i^g, μ_i^p are the mean vectors of ζ^g and ζ^p from the ith class and μ^g, μ^p are the mean vectors of all ζ^g and ζ^p vectors, respectively.

Defining $K_b^{gp}[1] = \sum_{i=1}^{C} n_i^g \mu_i^g \mu_i^{gT}$, $K_b^{gp}[2] = N_g \mu^p \mu^{pT}$, $K_b^{gp}[3] = \sum_{i=1}^{C} n_i^g \mu_i^g \mu^{pT}$, $K_b^{gp}[4] = \sum_{i=1}^{C} n_i^g \mu^p \mu_i^{gT}$, $K_w^{gp}[1] = \sum_{i=1}^{C} \sum_{j \in L_i^g} \zeta_j^g \zeta_j^{gT}$, $K_w^{gp}[2] = \sum_{i=1}^{C} n_i^g \mu_i^p \mu_i^{pT}$, $K_w^{gp}[3] = \sum_{i=1}^{C} n_i^g \mu_i^g \mu_i^{pT}$, $K_w^{gp}[4] = \sum_{i=1}^{C} n_i^g \mu_i^p \mu_i^{gT}$, (4) can be reformulated as

$$\begin{aligned} K_b^{gp} &= A^T K_b^{gp}[1] A + B^T K_b^{gp}[2] B \\ &\quad - A^T K_b^{gp}[3] B - B^T K_b^{gp}[4] A \\ K_w^{gp} &= A^T K_w^{gp}[1] A + B^T K_w^{gp}[2] B \\ &\quad - A^T K_w^{gp}[3] B - B^T K_w^{gp}[4] A \end{aligned} \tag{5}$$

Substituting (5) into (2), S_b and S_w can be reformulated as

$$\begin{aligned} S_b &= \tilde{A}^T K_b \tilde{A} \\ S_w &= \tilde{A}^T K_w \tilde{A} \end{aligned} \tag{6}$$

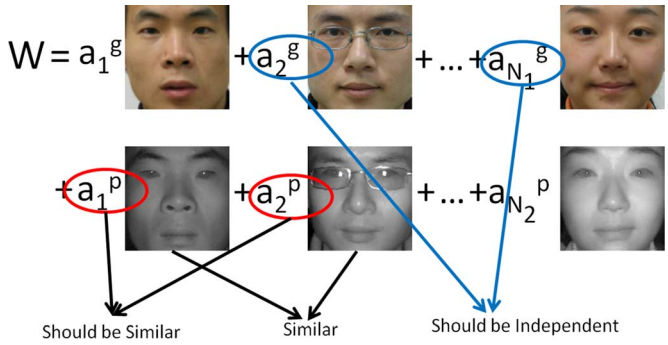


Fig. 4. Illustration of the principle of locality constraint in kernel space. If the two samples are similar, the corresponding combination coefficients should be similar; otherwise, the coefficients are independent.

where $\tilde{A} = [A; B]$, K_b and K_w are defined as

$$K_b = \begin{pmatrix} \sum_{i,j\in\{g,p\}} K_b^{ij}[1] & -\sum_{i,j\in\{g,p\}} K_b^{ij}[3] \\ -\sum_{i,j\in\{g,p\}} K_b^{ij}[4] & \sum_{i,j\in\{g,p\}} K_b^{ij}[2] \end{pmatrix}$$
$$K_w = \begin{pmatrix} \sum_{i,j\in\{g,p\}} K_w^{ij}[1] & -\sum_{i,j\in\{g,p\}} K_w^{ij}[3] \\ -\sum_{i,j\in\{g,p\}} K_w^{ij}[4] & \sum_{i,j\in\{g,p\}} K_w^{ij}[2] \end{pmatrix} \tag{7}$$

Like in LDA, the purpose of coupled discriminant analysis is to find projection $\tilde{A}$ that maximizes the ratio of S_b to S_w as

$$J = \frac{\tilde{A}^T K_b \tilde{A}}{\tilde{A}^T K_w \tilde{A}} \tag{8}$$

A. Locality Constraint

In practice, we always find that while the solution of (8) performs well on the training set, its generalization on unseen data is poor due to the limited number of training samples and high dimensionality of data. In order to deal with this problem, we impose some prior information on the objective function to limit the solution space to improve the generalization performance. As revealed in many previous studies [26], [27], locality information is an important clue in manifold learning. In most existing manifold learning methods (e.g., LPP, NPE), researchers try to find a subspace that best preserves the manifold structure in the data space. In this work, we do not aim to preserve the local structure of the data, but to exploit the discriminant subspace for classification. We adopt an alternative way to utilize the data manifold information to improve the generalization performance. As mentioned above, the learned projection can be represented as a linear combination of training samples. The solution to the projective vector can be transformed to find the linear combination coefficients for the projection. Generally, different samples make different contributions to the learned projection. That is, the coefficients of different samples used to form the projection are different. The principle of our idea is that if two samples are similar, their contributions to the projection should also be similar; otherwise, their contributions to the learned projection are independent to some extent (Fig. 4). Based on this idea, we impose the locality information in kernel space onto the process of coupled projection learning to alleviate the overfitting problem of its solution. Compared to the previous methods, there are two characteristics of the proposed locality information utilization. First, it is imposed on the combination of coefficients rather than the projected data, so that the projected data are allowed to change its neighboring structure, which is not inconsistent with the nonlinear mapping. Second, we explore the locality information in the kernel space rather than in the input data space. The manifold information in kernel space which is usually ignored in previous methods is exploited. In this way, information in both the input data space and kernel space is utilized. The locality constraint in kernel space was firstly proposed in [28] and has been shown to be helpful to improve the homogeneous face recognition performance. Specifically, suppose the similarity between samples i and j is S_{ij}, which is defined as,

$$S_{ij} = \begin{cases} \phi(X_i)^T \phi(X_j) & \text{if } \phi(X_i) \text{ and } \phi(X_j) \text{ are neighbors and from the same class} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Our idea can be formulated as minimizing the following criterion

$$\begin{aligned} J_l &= \frac{1}{2}\sum_i \sum_j [(\alpha_i - \alpha_j)^2 + (\beta_i - \beta_j)^2] S_{ij} \\ &= A^T L A + B^T L B \end{aligned} \tag{10}$$

where $L = D - S$ is the Laplacian matrix over the samples and D is a diagonal matrix in which $D_{ii} = \sum_j S_{ij}$. It is easy to verify that by minimizing J_l, the difference in coefficients whose corresponding samples are similar would be small, which is consistent with our motivation.

B. Consistency Constraint

As introduced in [23], although the projections for heterogeneous modalities are different, they are generated and used to describe the same object (e.g., face), so the coupled projections should not differ too much. Similar ideas have also been adopted in many multiview learning methods like [29], [30]. In order to improve the robustness of the solution and avoid the overfitting problem, we impose a penalty on the difference between the coupled projections onto the objective function as

$$\begin{aligned} J_c &= \|W^g - W^p\|^2 = \|\Phi A - \Phi B\|^2 \\ &= A^T \Phi^T \Phi A - A^T \Phi^T \Phi B - B^T \Phi^T \Phi A + B^T \Phi^T \Phi B \end{aligned} \tag{11}$$

Combining (8), (10) and (11), we obtain the objective formulation of the coupled discriminant analysis as

$$J = \frac{\tilde{A}^T K_b \tilde{A}}{\tilde{A}^T (K_w + \lambda \tilde{L}^l + \eta \tilde{L}^c) \tilde{A}} \tag{12}$$

where $\tilde{A} = [A; B]$; $\tilde{L}^l$ and $\tilde{L}^c$ are defined as

$$\tilde{L}^l = \begin{pmatrix} L & 0 \\ 0 & L \end{pmatrix}$$
$$\tilde{L}^c = \begin{pmatrix} \Phi^T \Phi & -\Phi^T \Phi \\ -\Phi^T \Phi & \Phi^T \Phi \end{pmatrix} \tag{13}$$

Input: Heterogeneous face training samples, $X^g = [X_1^g, X_2^g, \cdots, X_{N_g}^g]$ and $X^p = [X_1^p, X_2^p, \cdots, X_{N_p}^p]$. $N = N_g + N_p$ is the total number of samples.
Output: Coupled discriminant projections: $A \in R^{N\times d}$ and $B \in R^{N\times d}$, where d is the reduced dimensionality of the subspace.
1: Compute the within and between class scatter matrices with heterogeneous data as in Eq. 7.
2: Construct the locality constraint and consistency constraint as in Eqs. 10 and 11.
3: Solve the generalized eigenvalue problem and obtain the eigenvectors $\tilde{A}$ with the d largest eigenvalues.
$$K_b\tilde{A} = \gamma(K_w + \lambda\tilde{L}^l + \eta\tilde{L}^c)\tilde{A}.$$
4: Split $\tilde{A}$ into $A = \tilde{A}(1:N,:)$, $B = \tilde{A}(N+1:2N,:)$.
5: **Return:** A and B

Fig. 5. Locality constraint in kernel space-based coupled discriminant analysis (LCKS) algorithm.

By solving the generalized eigen-value problem $K_b\tilde{A} = \gamma(K_w + \lambda\tilde{L}^l + \eta\tilde{L}^c)\tilde{A}$ with its leading eigenvalues, we can finally obtain the solution to LCKS-CDA by splitting the result $\tilde{A}$ into A and B appropriately. Fig. 5 shows the major steps of the LCKS-CDA algorithm.

III. Locality Constraint Based Coupled Spectral Regression

Spectral regression is an effective subspace learning framework [31], deducing from the graph embedding view of subspace learning. Different from the traditional subspace learning, which finds the discriminant subspace directly, spectral regression finds the subspace projections in two steps. First, it finds the most effective low-dimensional embeddings for the original sample data; second, it learns the projection between the low-dimensional embedding and the original data with regression techniques. Previous work has shown that spectral regression is an effective method to learn the discriminant and robust subspace for classification. Recently, Lei and Li [23] extended the spectral regression and proposed coupled spectral regression (CSR) for matching heterogeneous faces. Coupled projections are learned for different modalities respectively with appropriate regression methods. This work proposes an improved CSR method, namely locality constraint in kernel space based coupled spectral regression (LCKS-CSR) by incorporating the locality information in kernel space into the CSR learning.

In [31], it was proved that for LDA, LPP, or NPE, whose similarity matrix can be represented as block-dialog formulation, its low-dimensional embeddings $Y = [y_1; y_2; \ldots; y_C]$ can be constructed directly as

$$y_t = [\underbrace{0,\ldots,0}_{\sum_{i=1}^{t-1} n_i}, \underbrace{1,\ldots,1}_{n_t}, \underbrace{0,\ldots,0}_{\sum_{i=t+1}^{C} n_i}]; t = 1,\ldots,C \tag{14}$$

where C is the number of classes and n_i is the number of ith class samples. Since the samples of different modalities have different distributions in data space, their projections should also be different. CSR aims to learn the projections W^g and W^p for modality g and p, respectively to satisfy $Y^g = W^{gT}X^g$ and $Y^p = W^{pT}X^p$, where X^g and X^p denote samples from the two modalities and Y^g and Y^p are the corresponding low embedding representations extracted from Y.

As described in the last section, denoting the transformed sample data in high-dimensional kernel space as $\Phi^g = [\phi(X_1^g), \ldots, \phi(X_{N_g}^g)]$ and $\Phi^p = [\phi(X_1^p), \ldots, \phi(X_{N_p}^p)]$, we assume that the coupled projections are represented by all samples from different modalities. That is, $W^g = \Phi A$ and $W^p = \Phi B$, where $\Phi = [\Phi^g, \Phi^p]$ and A, B are the combination coefficient vectors for coupled projections. The objective function of CSR can then be formulated as

$$\begin{aligned} J &= \frac{1}{N_g}\|Y^g - W^{gT}\Phi^g\|^2 + \frac{1}{N_p}\|Y^p - W^{pT}\Phi^p\|^2 \\ &= \frac{1}{N_g}\|Y^g - A^T\Phi^T\Phi^g\|^2 + \frac{1}{N_p}\|Y^p - B^T\Phi^T\Phi^p\|^2 \end{aligned} \tag{15}$$

By incorporating the locality constraint in kernel space (10) and the consistency constraint (11), the objective function of LCKS-CSR is formulated as

$$\begin{aligned} J = \frac{1}{N_g}\|Y^g - A^T\Phi^T\Phi^g\|^2 + \frac{1}{N_p}\|Y^p - B^T\Phi^T\Phi^p\|^2 \\ +\lambda(A^TLA + B^TLB) + \eta\|\Phi A - \Phi B\|^2 \end{aligned} \tag{16}$$

where the first two terms are data fitting items and the last two terms are the locality and consistency constraints that help to improve the generalization performance of the solution. Parameters λ and η control the trade-off between the data fitting accuracy and the generalization capability. By setting the derivatives of objective function with respect to A and B to zero, we have

$$\begin{aligned} \Theta_1 A &= \Phi^T\Phi^g Y^{gT} + N_g\eta\Phi^T\Phi B \\ \Theta_2 B &= \Phi^T\Phi^p Y^{pT} + N_p\eta\Phi^T\Phi A \end{aligned} \tag{17}$$

where

$$\begin{aligned} \Theta_1 &= \Phi^T\Phi^g\Phi^{gT}\Phi + N_g\lambda L + N_g\eta\Phi^T\Phi \\ \Theta_2 &= \Phi^T\Phi^p\Phi^{pT}\Phi + N_p\lambda L + N_p\eta\Phi^T\Phi \end{aligned} \tag{18}$$

With proper matrix manipulation, we can obtain the solution A and B as

$$\begin{aligned} A &= \Omega_1^{-1}(\Phi^T\Phi^g Y^{gT} + N_g\eta\Phi^T\Phi\Theta_2^{-1}\Phi^T\Phi^p Y^{pT}) \\ B &= \Omega_2^{-1}(\Phi^T\Phi^p Y^{pT} + N_p\eta\Phi^T\Phi\Theta_1^{-1}\Phi^T\Phi^g Y^{gT}) \end{aligned} \tag{19}$$

where

$$\begin{aligned} \Omega_1 &= \Theta_1 - N_gN_p\eta^2\Phi^T\Phi\Theta_2^{-1}\Phi^T\Phi \\ \Omega_2 &= \Theta_2 - N_gN_p\eta^2\Phi^T\Phi\Theta_1^{-1}\Phi^T\Phi \end{aligned} \tag{20}$$

After obtaining A and B, one can get the projections W_g and W_p for different modalities via (1).

IV. Experiments

As shown in (19) and (20), the solution to LCKS-CDA and LCKS-CSR can be represented as a series of inner product of

sample vectors. Therefore, we can use kernel trick as in SVM [32] to represent the data transformation ϕ implicitly. The RBF kernel $k(X_i, X_j) = \exp(-\|X_i - X_j\|^2/\sigma)$ is utilized in the following experiments.

We compare the proposed methods with state-of-the-art methods (CDFE [20], LDA + CCA [21], LCSR [23], KCSR [23] etc.) on different face databases. Three heterogeneous face recognition problems, including high resolution versus low resolution, digital photo versus video image and visible light (VIS) versus near infrared (NIR), are tested respectively to show the effectiveness of the proposed methods. In order to preserve the discriminant information as much as possible and compare different methods fairly, we preserve the dimensionality of subspace of all the compared methods to be $C - 1$, where C is the number of classes.

A. Parameter Selection

There are three parameters in the proposed LCKS-CDA and LCKS-CSR algorithms. One is σ in the RBF kernel and the other two are the trade-off parameters λ and η in (12) and (16). In this experiment, we determine these three parameter values on the PIE database in the case of high resolution versus low resolution heterogeneous face recognition problem.

The PIE database [33] consists of 41 368 images from 68 subjects under different poses, illumination and expression conditions. Five near frontal poses (C05, C07, C09, C27, C29) and all the images under different illuminations and expressions are selected. There are 170 images for each individual. The images are randomly partitioned into gallery and probe sets. Specifically, 5 images for each person are selected to construct the gallery set and the remaining images are used to construct the probe set. For every image, the high resolution image is cropped into 32×32 size and the low resolution image is obtained by first downsampling the high resolution one to the low resolution size and then upsampling it to the 32×32 size. Two low resolution sizes 16×16 and 8×8 are tested in the experiment. In the training phase, both the high and low resolution images in the gallery set are used. In testing phase, the high resolution images in gallery set are registered and low resolution images in probe set are tested. The random split is conducted 10 times and the mean recognition rate is reported.

For σ, we set it to $5e7$ empirically according to the average distance between the samples [34]. For λ and η, we select the values in the range $\{1e{-}12, 5e{-}12, 1e{-}11, 5e{-}11, 1e{-}10, 5e{-}10, 1e{-}9, 5e{-}9, 1e{-}8, 5e{-}8, 1e{-}7, 5e{-}7, 1e{-}6, 5e{-}6, 1e{-}5, 5e{-}4, 1e{-}4\}$. Fig. 6 shows the recognition rate trends of LCKS-CDA and LCKS-CSR with respect to λ and η for 16×16 and 8×8 resolution images, respectively. For LCKS-CDA, the optimal values of λ and η are $1e{-}9$ and $5e{-}8$, respectively, when the low resolution image size is 16×16. While for the 8×8 low resolution images, the optimal values of λ and η are $1e{-}8$ and $1e{-}9$, respectively. We finally set values of both λ and η, for LCKS-CDA, to $1e{-}9$ in the following experiments. For LCKS-CSR, in the case of 16×16 resolution, the best accuracy is achieved when both λ and η are chosen to be $1e{-}7$, while for 8×8 images, the optimal values of λ and η are $5e{-}7$ and $5e{-}8$, respectively. The optimal choice of parameters for 16×16 and 8×8 resolutions are similar and in the following experiments, the values of σ, λ, η for LCKS-CSR are fixed as $5e7$, $1e{-}7$ and $1e{-}7$, respectively.

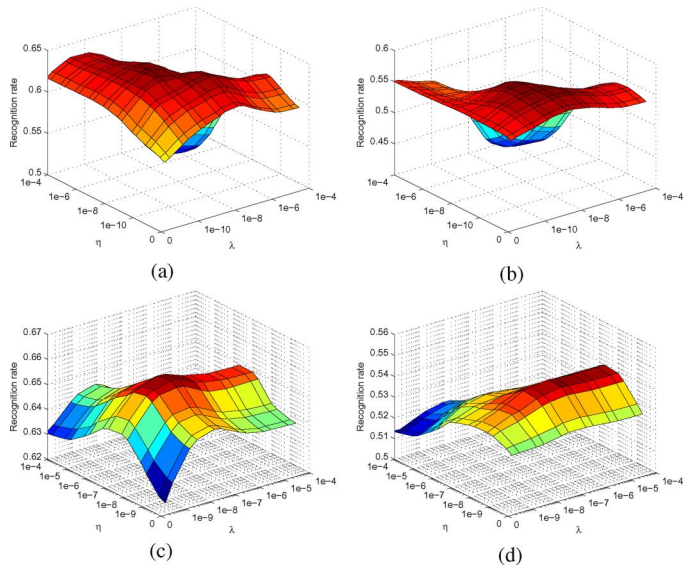


Fig. 6. Recognition rate trend of LCKS-CDA [(a), (b)] and LCKS-CSR [(c), (d)] with respect to λ and η. (a) LCKS-CDA; 16×16 resolution. (b) LCKS-CDA; 8×8 resolution. (c) LCKS-CSR; 16×16 resolution. (d) LCKS-CSR; 8×8 resolution.

B. Multi-PIE: High Resolution versus Low Resolution

In the proposed method, the coupled projections are learned from all available samples from different modalities, while in earlier studies, the projections are supposed to be represented by samples from the corresponding modality, respectively. In order to verify the advantage of the proposed method, we first compare the two coupled projection representations on the PIE database. For ease of representation, the CDA and CSR methods with these two projection representation assumptions are denoted as CDA^a, CSR^a and CDA^s, CSR^s, respectively, denoting CDA/CSR learning from all samples in different modalities or learning from samples in single modality. In this experiment, the values of λ and η in CDA and CSR are set to 0 in order to compare the two representations fairly. The training and testing protocols are the same as in Section IV-A.

Fig. 7 shows the recognition results of the two projection representation methods. It shows that the method that determines both the coupled projections based on all samples from different modalities outperforms the previous method, in which the coupled projections are determined by samples from single modality. These results further indicate that the proposed projection representation method helps to improve the heterogeneous face recognition performance.

In the following part, we compare the proposed LCKS-CDA and LCKS-CSR methods with LDA, CDFE, LDA + CCA, LCSR and KCSR methods in the case of high resolution

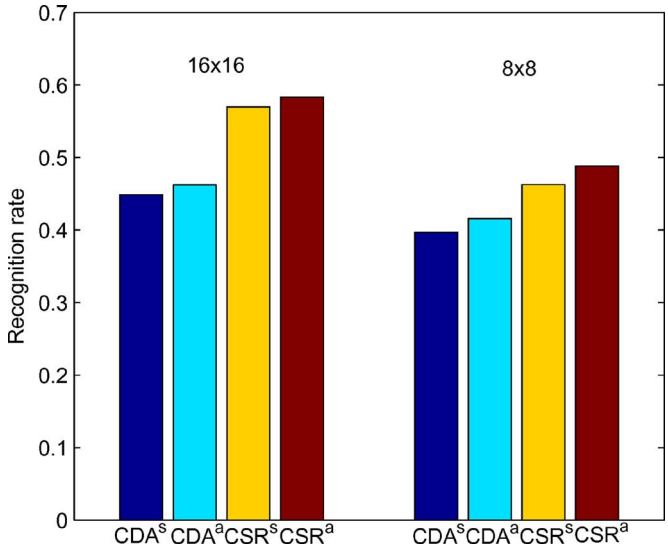


Fig. 7. Recognition rates of CDA^a, CDA^s, CSR^a, and CSR^s. The superscripts a and s denote, respectively, "learning from all samples" and "learning from single modality."

versus low resolution face recognition[1] on the Multi-PIE face database. The parameters of CDFE, LDA + CCA, LCSR and KCSR are adopted according to the recommended values in their original papers. For LDA, we combine the heterogeneous data (high resolution and low resolution images) together and train a single projection matrix for high and low resolution data.

The Multi-PIE database [35] is an extended version of PIE which contains 337 subjects from 4 sessions under different poses, illumination conditions and expressions. The frontal views with neutral expression under different illuminations are selected. There are 18 420 images in total from 337 subjects with 20–60 images per subject. In this experiment, we divide the database into three sets, namely training set, gallery set and probe set. In training set, 100 subjects with 20 images per person are selected. We select another 100 persons to construct the gallery and probe sets. For ease of representation, we use $\mathrm{G}N$ to denote that N images per subject are selected to construct the gallery set and the remaining images are used as the probe set. There is no intersection between the training set and gallery/probe set. In testing phase, the low resolution images in the probe set are compared with high resolution ones in the gallery set. The random split is run 10 times and the mean recognition rate along with the standard deviation are reported.

In our experiments, all the images are cropped to 32×32 size. For the low resolution $M \times M$ images, the images are first downsampled to $M \times M$ and then they are upsampled to 32×32 size. Fig. 8 shows high resolution image examples with four low resolution sizes 16×16, 8×8, 6×6 and 4×4, respectively.

Table I lists the recognition rates of LDA, CDFE, LDA+CCA, LCSR, KCSR, LCKS-CDA and LCKS-CSR on Mutli-PIE database. From the results, we can see that the performances of

[1]Note that for high resolution versus low resolution face recognition, there are many methods that utilize a super-resolution technique to synthesize high resolution images, followed by face recognition between high resolution images. In this experiment, we take high resolution versus low resolution as an example of heterogeneous face recognition and focus on subspace related methods.

Fig. 8. Cropped face image examples for three subjects. From left to right: high resolution (32×32), followed by low resolution (16×16, 8×8, 6×6, 4×4) images.

coupled projection based methods (CDFE, LDA+CCA, LCSR, KCSR, LCKS-CDA and LCKS-CSR) are significantly better than that of LDA, indicating that the coupled projection strategy for heterogeneous face recognition is effective. Generally, the spectral regression based methods (LCSR, KCSR, LCKS-CSR) achieve better recognition performance than traditional discriminant analysis methods (CDFE, LDA+CCA, LCKS-CDA). It indicates that the spectral regression is a good alternative framework of subspace learning and can lead to better generalization than the traditional frameworks. Comparing the proposed method with existing ones (LCKS-CDA versus CDFE, LDA+CCA and LCKS-CSR versus LCSR, KCSR), the performances are similar for 16×16 and 8×8 resolutions. However, in the lower resolution case of 6×6 and 4×4, the proposed methods (LCKS-CDA and LCKS-CSR) outperform their counterparts significantly. For example, LCKS-CSR improves the recognition rate of KCSR by 6–20 percent when the low resolution size is 4×4. In KCSR, the projections for high and low resolution data are derived from the high and low resolution samples, respectively. LCKS-CSR learns the coupled projections by utilizing high and low resolution images together. The good performance of LCKS-CSR, especially in lower resolution cases, validates that it is possible to utilize more discriminative information between heterogeneous data. This property along with the locality constraint are helpful to improve the generalization capability. Overall, LCKS-CSR achieves the best recognition performance among the various heterogeneous methods considered here.

C. Digital Photo versus Video Frame

The digital photo and video database was collected by us under the surveillance scenario. There are 311 subjects and for each subject, there are 5 photo images and 5 frames/images selected from videos. There are in total 1 555 digital photos and 1 555 video frames, respectively. The face images in these two modalities have significant pose and image quality variations. All the images are cropped into 35×30 size according to automatically detected eye coordinates. Fig. 9 shows some example images cropped from digital photo and video frame sets.

In this experiment, we randomly partition the original database into training and testing sets. The training set contains 150 persons with their digital photos and video images and the remaining images from 161 persons form the testing set. There is no intersection between training and testing sets in terms of subject or image. In the testing phase, the digital photos are registered in the gallery and the video images are considered as

TABLE I
PERFORMANCE COMPARISON (MEAN ACCURACY (%) ± STD) OF HIGH VERSUS LOW RESOLUTION FACE RECOGNITION ON MULTI-PIE DATABASE. FOUR LOW RESOLUTION CASES (16×16, 8×8, 6×6, AND 4×4) ARE TESTED. G2, G5 AND G10 MEAN 2, 5, 10 IMAGES PER SUBJECT ARE RANDOMLY SELECTED IN THE GALLERY SET

Methods	G2	G5	G10
LDA	63.67±2.62	77.19±1.69	84.66±0.94
CDFE	73.31±1.55	86.03±1.18	92.89±0.65
LDA+CCA	75.30±1.65	87.03±1.07	92.60±0.60
LCSR	82.52±1.41	92.58±0.69	96.18±0.20
KCSR	84.57±1.42	93.40±0.40	96.71±0.37
LCKS-CDA	80.97±1.79	91.50±0.50	95.62±0.49
LCKS-CSR	**85.53**±1.47	**94.02**±0.38	**97.08**±0.30

(a) 16×16

Methods	G2	G5	G10
LDA	33.53±2.43	41.48±2.87	47.26±2.19
CDFE	45.33±1.57	56.62±1.73	63.68±1.33
LDA+CCA	45.24±2.69	54.86±1.76	57.73±5.45
LCSR	64.54±1.11	77.09±1.25	82.45±0.70
KCSR	73.92±1.31	85.70±0.63	91.63±0.50
LCKS-CDA	71.28±1.52	83.40±0.48	89.52±0.80
LCKS-CSR	**74.52**±1.37	**86.71**±0.90	**92.55**±0.47

(b) 8×8

Methods	G2	G5	G10
LDA	15.21±1.21	17.84±1.88	20.37±1.22
CDFE	36.22±1.47	44.30±1.41	48.70±1.39
LDA+CCA	32.32±2.22	37.26±2.36	39.66±2.60
LCSR	39.96±1.59	48.66±1.15	51.51±1.18
KCSR	51.99±1.70	63.79±1.44	71.41±1.10
LCKS-CDA	54.84±1.98	67.00±1.26	74.10±0.49
LCKS-CSR	**57.15**±1.57	**70.54**±1.10	**79.34**±0.87

(c) 6×6

Methods	G2	G5	G10
LDA	9.50±1.04	10.61±1.23	11.07±0.93
CDFE	16.03±0.83	18.02±0.94	19.09±1.10
LDA+CCA	11.96±0.85	12.03±0.87	13.00±1.27
LCSR	17.83±1.03	20.15±0.99	20.81±1.21
KCSR	30.97±1.27	37.75±0.98	42.91±0.79
LCKS-CDA	29.07±1.63	35.33±1.19	40.80±1.40
LCKS-CSR	**36.26**±0.86	**50.73**±0.94	**62.35**±0.87

(d) 4×4

Fig. 9. Cropped face image examples of digital photo (first row) and video image/frame (second row).

the probe images, which is consistent with the practical applications. Both the rank-1 recognition performance and receiver operating characteristic (ROC) performance are reported.

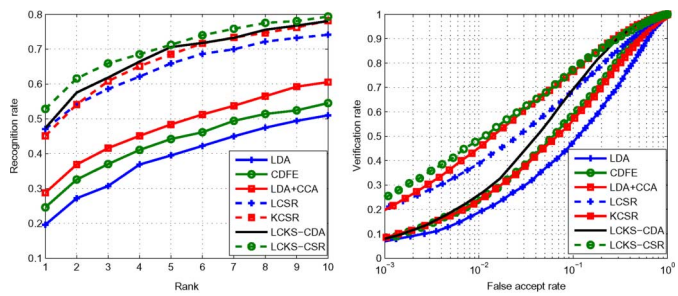
Table II lists the recognition performance of different methods for digital photo versus video frame recognition and Fig. 10 shows the corresponding rank and ROC curves. In particular, the rank-1 recognition rate and verification rates when the false accept rate is set at 0.1, 0.01, 0.001 are reported. From these results, one can see that the proposed LCKS based methods achieve better recognition performance than existing ones. LCKS-CDA improves the rank-1 performance of CDFE and LDA + CCA by more than 10 percents. LCKS-CSR enhances the rank-1 recognition rates of LCSR and KCSR by 5 percent. These results indicate that LCKS is effective in improving the heterogeneous face recognition performance of coupled subspace learning. Comparing the results of spectral regression based methods (LCSR, KCSR, LCKS-CSR) with other methods, one can find that the spectral regression based method always achieves better performance than others, indicating that spectral regression is an effective subspace learning framework which provides better generalization performance. Overall, the proposed LCKS-CSR method achieves the best performance in terms of all indices.

D. CASIA-HFB: VIS versus NIR

The CASIA-HFB database is an extended version of the HFB database [36] collected by CBSR for heterogeneous biometric

TABLE II
PERFORMANCE COMPARISON (%) ON DIGITAL PHOTO VERSUS VIDEO FRAME MATCHING

Method	Rec. Rate	VR@FAR=10%	VR@FAR=1%	VR@FAR=0.1%
LDA	19.61	47.63	19.03	6.90
CDFE	25.48	59.06	24.34	7.71
LDA+CCA	28.76	57.02	23.66	8.55
LCSR	47.06	69.07	38.88	20.76
KCSR	45.10	76.65	45.41	19.92
LCKS-CDA	47.32	69.67	26.14	8.13
LCKS-CSR	**52.81**	**77.59**	**48.65**	**24.78**



(a) (b)

Fig. 10. (a) CMC and (b) ROC curves of different methods on digital photo versus video frame database.

research. There are 300 subjects, for each of which there are 5 VIS images and 5 NIR images. In this experiment, we use the images of the first 150 subjects for training and the remaining 150 subjects constitute the testing set. There is no intersection between the training and testing sets in terms of subjects and images. In the testing phase, the VIS images of each subject are registered as the gallery set and the NIR ones are used as the probe set. The rank-1 recognition rate and the receiver operating characteristic (ROC) performance are reported. All images are cropped to 32×32 size according to the automatically detected eye coordinates. Fig. 11 shows some VIS and NIR face images from this database.

We compare the proposed LCKS-CDA and LCKS-CSR methods with CDFE [20], LDA + CCA [21], LCSR and KCSR methods and use the LDA as the baseline method. The

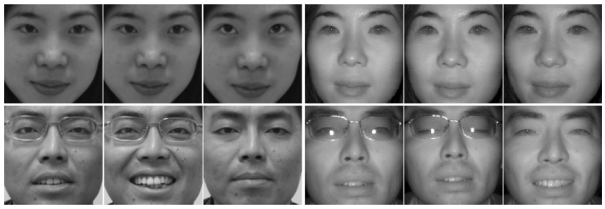
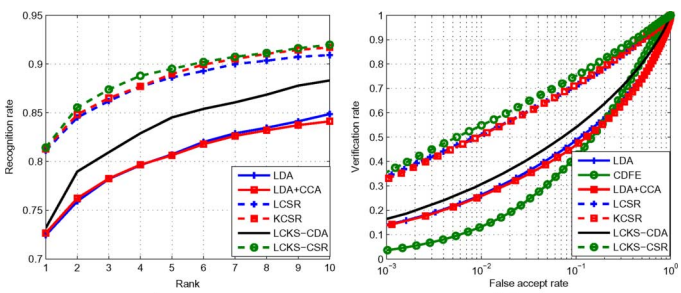


Fig. 11. Cropped VIS and NIR face image examples of three subjects. The left three columns are VIS images and the right three columns are the corresponding NIR images.

TABLE III
PERFORMANCE COMPARISON (%) ON VIS VERSUS NIR DATABASE

Method	Rec. Rate	VR@FAR=10%	VR@FAR=1%	VR@FAR=0.1%
LDA	72.43	48.75	26.55	14.04
CDFE	16.10	40.05	12.75	3.41
LDA+CCA	72.65	42.90	25.76	13.79
LCSR	81.12	71.28	51.07	33.98
KCSR	81.30	71.66	50.95	32.84
LCKS-CDA	73.18	54.11	31.21	16.61
LCKS-CSR	**81.43**	**75.18**	**54.81**	**35.69**



(a) (b)

Fig. 12. (a) CMC and (b) ROC curves of different face heterogeneous recognition methods on VIS-NIR database.

parameters of these methods are set to the recommended values in their papers. For the LDA method, the VIS and NIR images are combined together to learn a single projection for VIS and NIR images. Table III lists the performance (rank-1 recognition rate, verification rates (VR) at false accept rates (FAR) of 0.1, 0.01 and 0.001) of different methods and Fig. 12 shows the corresponding rank and ROC curves. We omit the rank curve of CDFE for ease in illustration. Surprisingly, CDFE and LDA + CCA methods perform worse than LDA, indicating that their generalization is very poor in this case. The proposed LCKS-CDA method outperforms LDA, but not by much. In contrast, the spectral regression based methods (i.e., LCSR, KCSR, LCKS-CSR), achieve significantly better results than others. It shows that the spectral regression is a good alternative to the traditional subspace learning and has better generalization performance. Overall, the proposed LCKS-CSR achieves the best performance in terms of all indices, supporting the conjecture that the locality information and the representation derived from all the samples is helpful to improve the heterogeneous face recognition performance.

V. CONCLUSION

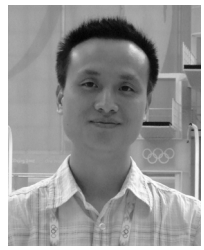
This paper incorporates locality constraint in kernel space into coupled subspace learning to solve the heterogeneous face recognition problem. Both the coupled projections proposed here are supposed to be represented by all available samples from different modalities, so that the mutual information between different modalities is sufficiently explored. The locality information in kernel space is modeled and imposed onto the combination coefficients properly. In this way, structures of the data in the input space and transformed kernel space are utilized, resulting in more discriminative information for heterogeneous face recognition. Two implementations, namely LCKS-CDA and LCKS-CSR are presented. Experiments on various databases demonstrate that the proposed LCKS based methods do improve the performance of heterogeneous face recognition.

REFERENCES

[1] S. Z. Li, , S. Z. Li, Ed., "Encyclopedia of Biometrics," in *Heterogeneous Face Biometrics*. New York: Springer, 2009.
[2] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surveys*, pp. 399–458, 2003.
[3] , S. Z. Li and A. K. Jain, Eds., *Handbook of Face Recognition*, 2nd ed. New York: Springer-Verlag, Aug. 2011.
[4] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Hawaii, Jun. 1991, pp. 586–591.
[5] 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.
[6] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[7] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.
[8] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.
[9] Z. Lei, S. Z. Li, R. Chu, and X. Zhu, "Face recognition with local gabor textons," in *Proc. ICB*, 2007, pp. 49–57.
[10] 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.
[11] W. Zhang, S. Shan, W. Gao, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," in *Proc. ICCV*, 2005, pp. 786–791.
[12] Z. Lei, S. Liao, M. Pietikäinen, and S. Z. Li, "Face recognition by exploring information jointly in space, scale and orientation," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 247–256, Jan. 2011.
[13] X. Tang and X. Wang, "Face sketch recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 50–57, Jan. 2004.
[14] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. CVPR*, 2005, pp. 1005–1010.
[15] 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.
[16] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," in *Proc. ICB*, 2009, pp. 209–218.
[17] B. F. Klare and A. K. Jain, "Heterogeneous face recognition: Matching NIR to visible light images," in *Proc. ICPR*, 2010, pp. 1513–1516.
[18] B. F. Klare, Z. Li, and A. K. Jain, "Matching forensic sketches to mug shot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, Mar. 2011.
[19] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proc. CVPR*, 2011, pp. 513–520.
[20] D. Lin and X. Tang, "Inter-modality face recognition," in *Proc. ECCV*, 2006, pp. 13–26.

[21] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li, "Face matching between near infrared and visible light images," in *Proc. ICB*, 2007, pp. 523–530.
[22] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Z. Li, "2D-3D face matching using cca," in *Proc. FGR*, Amsterdam, The Netherlands, 2008.
[23] Z. Lei and S. Z. Li, "Coupled spectral regressoin for matching heterogeneous faces," in *Proc. CVPR*, 2009, pp. 1123–1128.
[24] Z. Lei, C. Zhou, D. Yi, A. K. Jain, and S. Z. Li, "An improved coupled spectral regression for heterogeneous face recognition," in *Proc. ICB*, 2012, pp. 7–12.
[25] G. H. Golub and C. F. van Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
[26] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, no. 5500, pp. 2323–2326, Dec. 22, 2000.
[27] 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.
[28] Z. Lei, Z. Zhang, and S. Z. Li, "Feature space locality constraint for kernel based nonlinear discriminant analysis," *Pattern Recognit.*, vol. 45, no. 17, pp. 2733–2742, 2012.
[29] J. D. R. Farquhar, H. Meng, S. Szedmak, D. R. Hardoon, and J. Shawe-Taylor, "Two view learning: Svm-2k, theory and practice," in *NIPS*. Cambridge, MA: MIT Press, 2006.
[30] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 6, pp. 1438–1446, Dec. 2010.
[31] D. Cai, X. He, and J. Han, "Spectral regression for efficient regularized subspace learning," in *Proc. ICCV*, Rio de Janeiro, 2007.
[32] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[33] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.
[34] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 235–249, 2007.
[35] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.
[36] S. Z. Li, Z. Lei, and M. Ao, "The hfb face database for heterogeneous face biometrics research," in *Proc. 6th IEEE Workshop on Object Tracking and Classification Beyond and in the Visible Spectrum*, Miami, FL, 2009.

Zhen Lei (S'08–M'11) received the B.S. degree in automation from the University of Science and Technology of China (USTC), in 2005, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences (CASIA), in 2010, where he is now an assistant professor. His research interests are in computer vision, pattern recognition, image processing, and face recognition in particular. He has published over 40 papers in international journals and conferences.



Shengcai Liao received the B.S. degree in mathematics and applied mathematics from the Sun Yat-sen University, Guangzhou, China, in 2005, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently a Post Doctoral Fellow in the Department of Computer Science and Engineering, Michigan State University. His research interests include computer vision, pattern recognition, and machine learning, with a focus on image and video analysis, particularly face recognition, object detection and recognition, video surveillance, and sparse matrix factorization. He serves as a reviewer for several international journals including IJCV, T-PAMI, TIP, IVC, and Neurocomputing. He has also served as a program committee member or reviewer for several international conferences, including ICCV, CVPR, ECCV, ICPR, ICB, BTAS, etc.

Dr. Liao was awarded the Motorola Best Student Paper award and the 1st Place Best Biometrics Paper award at the International Conference on Biometrics in 2006 and 2007, respectively, for his work on face recognition.

Anil K. Jain (S'70–M'72–SM'86–F'91) is a university distinguished professor in the Department of Computer Science and Engineering at Michigan State University. His research interests include pattern recognition and biometric authentication. He served as the editor-in-chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (1991–1994). The holder of six patents in the area of fingerprints, he is the author of a number of books, including *Handbook of Fingerprint Recognition* (2009), *Handbook of Biometrics* (2011), *Handbook of Multibiometrics* (2006), *Handbook of Face Recognition* (2005), *BIOMETRICS: Personal Identification in Networked Society* (1999), and *Algorithms for Clustering Data* (1988). He served as a member of the Defense Science Board and The National Academies committees on Whither Biometrics and Improvised Explosive Devices.

Dr. Jain received the 1996 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award and the Pattern Recognition Society best paper awards in 1987, 1991, and 2005. He is a fellow of the AAAS, ACM, IAPR, and SPIE. He has received Fulbright, Guggenheim, Alexander von Humboldt, IEEE Computer Society Technical Achievement, IEEE Wallace McDowell, ICDM Research Contributions, and IAPR King-Sun Fu awards.

Stan Z. Li (M'92–SM'99–F'09) received the B.Eng. degree from Hunan University, China, the M.Eng. degree from National University of Defense Technology, China, and the Ph.D. degree from Surrey University, U.K.

He is currently a professor at the National Laboratory of Pattern Recognition and the director of the Center for Biometrics and Security Research (CBSR), Institute of Automation (CASIA), and the director of the Center for Visual Internet of Things Research (VIOT), Chinese Academy of Sciences. He worked at Microsoft Research Asia as a researcher from 2000 to 2004. Prior to that, he was an associate professor at Nanyang Technological University, Singapore. He was elevated to IEEE Fellow for his contributions to the fields of face recognition, pattern recognition, and computer vision. His research interest includes pattern recognition and machine learning, image and vision processing, face recognition, biometrics, and intelligent video surveillance. He has published over 200 papers in international journals and conferences, and authored and edited eight books. He was an associate editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and was acting as the Editor-in-Chief for the *Encyclopedia of Biometrics*. He served as a program cochair for the International Conference on Biometrics 2007 and 2009, a general chair for the 9th IEEE Conference on Automatic Face and Gesture Recognition, and has been involved in organizing other international conferences and workshops in the fields of his research interest.