Signal Processing 124 (2016) 162–172

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

CDMMA: Coupled discriminant multi-manifold analysis for matching low-resolution face images

Junjun Jiang [a,b,*], Ruimin Hu [c], Zhongyuan Wang [c], Zhihua Cai [a,*]

[a] School of Computer Science, China University of Geosciences, Wuhan 430074, China
[b] Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China
[c] National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 11 May 2015
Received in revised form
29 August 2015
Accepted 18 September 2015
Available online 3 October 2015

Keywords:
Face recognition
Super-resolution
Multi-manifold
Discriminant analysis
Low-resolution

ABSTRACT

Face images captured by surveillance cameras usually have low-resolution (LR) in addition to uncontrolled poses and illumination conditions, all of which adversely affect the performance of face matching algorithms. In this paper, we develop a novel method to address the problem of matching a LR or poor quality face image to a gallery of high-resolution (HR) face images. In recent years, extensive efforts have been made on LR face recognition (FR) research. Previous research has focused on introducing a learning based super-resolution (LBSR) method before matching or transforming LR and HR faces into a unified feature space (UFS) for matching. To identify LR faces, we present a method called coupled discriminant multi-manifold analysis (CDMMA). In CDMMA, we first explore the neighborhood information as well as local geometric structure of the multi-manifold space spanned by the samples. And then, we explicitly learn two mappings to project LR and HR faces to a unified discriminative feature space (UDFS) through a supervised manner, where the discriminative information is maximized for classification. After that, the conventional classification method is applied in the CDMMA for final identification. Experimental results conducted on two standard face recognition databases demonstrate the superiority of the proposed CDMMA.

1. Introduction

Image analysis and understanding technologies have been received more and more attentions in recent years [1–4]. As one of the most successful applications of these technologies, face recognition (FR), has been an active research area for more than two decades and many promising practical face recognition systems have been developed [5,6]. Although most current FR systems can achieve advanced levels under controlled environments, e.g., the face region is large enough and contains sufficient information for recognition, it is well known that they confront with low-resolution (LR) problems [7,8]. In many surveillance scenarios, in which the individual of interest is often times far away from the cameras, a captured face image is usually LR and it lacks detailed facial features, which are of vital importance to face recognition. This issue is called low-resolution face recognition (LR FR). Although recognizing a LR face in the surveillance scenario contains a lot of other technologies, such as detection [9–11], registration [12,13], and noise reduction [14,15], in this paper we are just concerned about the classification given the extracted LR face image.

1.1. Prior work

In this paper, we study the problem of how to matching a LR probe image to a high-resolution (HR) gallery of enrolled faces. There are three standard approaches to this

* Corresponding authors.

http://dx.doi.org/10.1016/j.sigpro.2015.09.026
0165-1684/

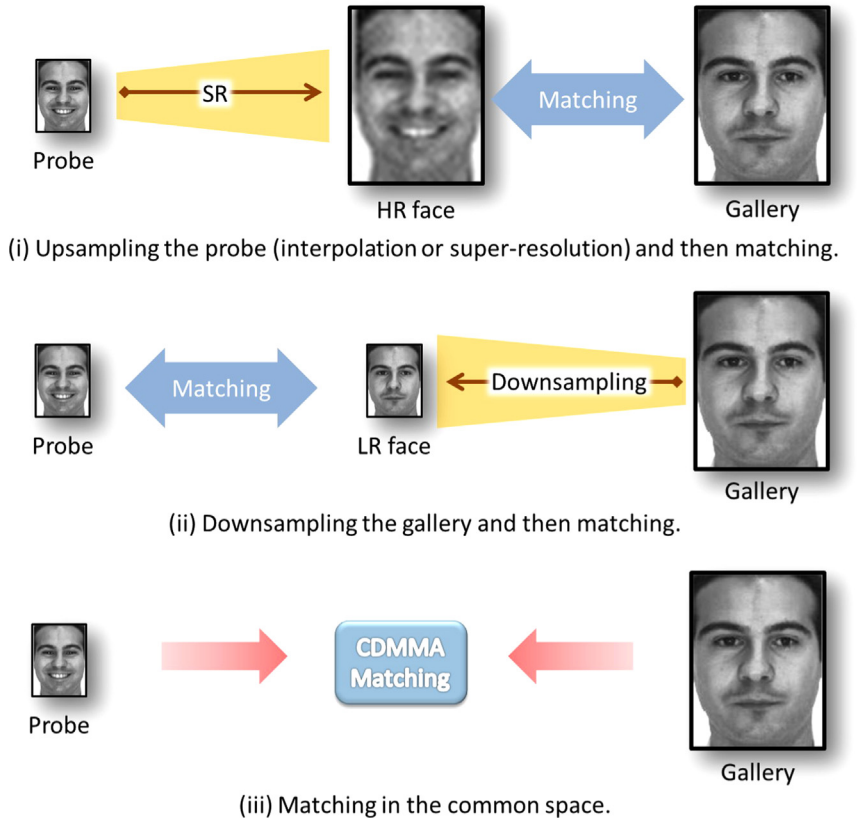

Fig. 1. Standard approaches to matching a LR probe to a HR gallery. (i) Upsampling the probe (interpolation or super-resolution) and then matching; (ii) downsampling the gallery and then matching; (iii) matching in the common space. The method proposed in this paper belongs to the third kind of approach.

problem (as shown in Fig. 1): (i) downsample the entire gallery and then perform matching in LR face image space; (ii) utilize the super-resolution (SR) or interpolation algorithm to obtain higher resolution version from the probe image and then perform matching; (iii) reduce the gap in appearance of different resolution and then apply the traditional face recognition methods on the common space.

The first class is very intuitive and simple. Recently, many learning based super-resolution (LBSR) methods have been developed to predict the corresponding high-resolution (HR) face from one input LR face to overcome LR problems. For example, with the assistance of training samples, Baker and Kanade [16] propose "*face hallucination*" to infer the HR face image from an input LR one under a Bayesian formulation. In [17], Liu et al. propose one two-step method. It integrates a global parametric Principal Component Analysis (PCA) model with a local non-parametric Markov Random Field (MRF) model for face image super-resolution. These are the pioneering works on hallucinating face image. Since the introduction of this work, a number of different methods and models have been introduced. In [18], Wang and Tang propose a holistic face hallucination method using an eigen-transformation. It employs PCA to represent a LR input image as a linear combination of LR training samples. Then, the target HR face image is reconstructed by replacing the LR training images with the corresponding HR ones, while using the same coefficients. Chakrabarti et al. [19] further utilize a Kernel PCA model to hallucinate face images. Park and Savvides [20] decompose a LR input face into many prototype faces via PCA and apply a recursive error back-propagation method to reconstruct the HR face. These mentioned methods are holistic and fail to render effectively the fine individual details. Recently, many techniques of decomposing a complete image into smaller patches have been introduced.

With the assumption that the HR image patches and LR image patches share the same local geometric structure, Change et al. [21] utilize locally linear embedding (LLE) [22] to learn the optimal reconstruction weights of K LR base elements to estimate the target HR patch representations. Most recently, inspired by the face analysis research which states that the position information is very important for face analysis and synthesis, Ma et al. [23] propose a position-patch based face hallucination method by performing collaboratively over the whole training image patches of the same position. To overcome the unstable solution of least squares problem in [23], Jung et al. [24] and Jiang et al. [25] propose to utilize sparse regularization and locality to obtain the optimal reconstruction weights for face hallucination respectively. In [26–28], Jiang et al. further propose an iterative neighbor embedding strategy to solve face super-resolution

problem. Recently, Hu et al. [29] propose a new framework by learning local-pixel structure to global image. The complexity of the algorithm is very high and the performance depends on training set.

However, although above-mentioned methods are the best and efficient in the sense of reduced mean square error (MSE) and high peak signal to noise ratio (PSNR), these methods aim at obtaining visually appealing results; moreover, these time-consuming SR methods are not suitable for real-world FR systems. Therefore, these existing LBSR based FR methods are unsuitable for the LR problems.

Multi-modal tensor super-resolution is the prior work and is targeted at super-resolving a discriminant HR face image [30]. In order to simultaneously recognize and super-resolve the LR faces, Hennings-Yeomans et al. [7] expressed the constraints between LR and HR images in a regularization formulation. Recently, Biswas et al. [31] proposed to use the multidimensional scaling to match LR face images. Inspired by some recent work of manifold learning [22], many manifold learning based common space methods have been proposed [32–34]. Huang and He [32] apply the canonical correlation analysis (CCA) to perform a nonlinear mapping between HR and LR face images in a coherent space. Li et al. [33] propose a coupled locality preserving mapping (CLPM) method to address LR problems from the perspective of manifold learning. They first explore the local geometric structure in the original space, and then preserve the local geometric structure for a unified feature space (UFS) shared by the LR and HR space. In particular, they explicitly learn two mappings, which can project the LR and HR resolution face images into UFS, and then perform matching in UFS.

However, CLPM seeks to model the face space in a single manifold through an unsupervised manner (dropping the label information) and identify neighbors according to Euclidean distance, and this will easily connect samples belonging to different manifolds (Note that,

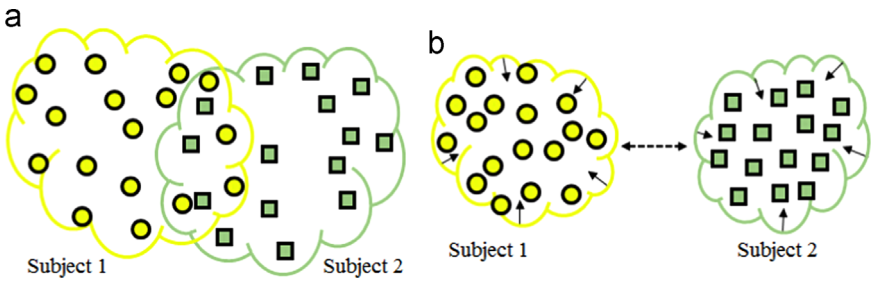

Fig. 2. Discriminant analysis on multi-manifold. (a) Shows the intersection of different manifolds formed by two subjects; (b) is the desired result of the proposed method, where samples of different subjects are well separated while those of the same subject become closer after the mappings. As a result, the manifold margin in the unified discriminative feature space is much larger and more discriminative information can be exploited for recognition.

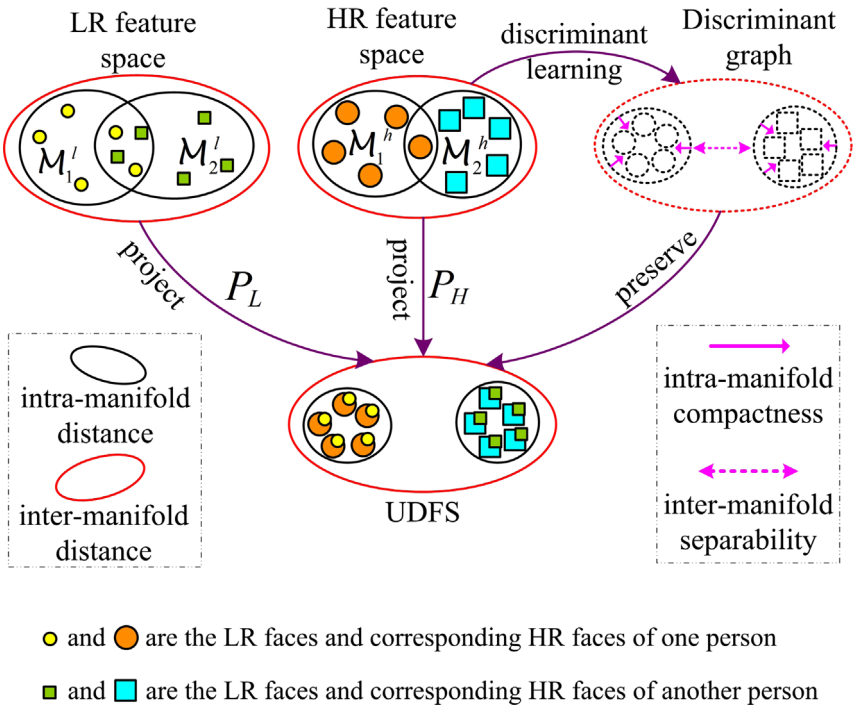

Fig. 3. A conceptual illustration of the proposed approach.

following [35], we assume that faces of the same person lie on the same manifold and those of different persons are associated with different manifolds), especially those near the intersection of different manifolds as shown in Fig. 2. Thus, it may drop some discriminative information helpful in classification.

In order to pursue discriminability, Wilman and Yuen [34] propose a discriminative super-resolution (DSR) method, which incorporates the class label information for SR. However, DSR treats all samples in the same manner during the calculation of within-class and between-class distances and thus captures the global geometric structure, dropping the local structure of manifold, which has recently been shown to be effective for the classification problem [36].

1.2. Motivation

To handle those drawbacks mentioned above, we develop a novel discriminative subspace learning method, namely coupled discriminant multi-manifold analysis (CDMMA). Due to the fact that different observation sets (e.g., HR faces and LR faces) might be located in different high-dimensional spaces, it is difficult to match the data in their original observation spaces. However, as these observations are from the same object, it is reasonable to assume that some common representations can be explored by an underlying unified space. Following CLPM, we adopt to use two explicit mappings to project the LR and HR faces into a unified discriminative feature space (UDFS) as shown in Fig. 3. Different from CLPM, our proposed CDMMA simultaneously learns from the neighboring information as well as the local geometric structure implied by the samples (identifying the neighbors of each sample from the manifold space rather than the entire Euclidean space) to minimize the intra-manifold distance and maximize the inter-manifold distance so that more discriminant information can be exploited for FR. Experiments on standard FR databases demonstrate the effectiveness of the proposed method.

Built upon our preliminary work reported in [37], this paper is an improved version with the following contributions: (i) the introduction section is rewritten to provide an extensive review of relevant work and to make our contributions clear; (ii) the geometrical structure of HR image manifold is very important for the proposed approach, so we give the analysis on the different neighbor numbers and the parameter settings of our proposed approach in this extended version.

1.3. Organization of this paper

The remainder of the paper is organized as follows. In Section 2, we introduce the proposed CDMMA based face recognition method. Firstly, we give some notations, then introduce the details of constructing the discriminant graph on multi-manifold. Lastly, we show how to construct the objective function and how to solve it. We conduct the comparison experiments on two benchmark face database and analysis several important parameters of the proposed CDMMA in Section 3. Finally, we conclude in Section 4.

2. Coupled discriminant multi-manifold analysis

This section presents formulation of the proposed CDMMA method. It then describes the optimization algorithm and presents the algorithm pseudo code to summarize our proposed CDMMA method.

2.1. Preliminaries

As shown in Fig. 3, in order to exploit the discriminative information, we first learn the similarity relationship in the HR feature space, and then preserve the relationship (dotted line) for UDFS, enhancing the intra-manifold compactness and maximizing the inter-manifold separability simultaneously. Therefore, what we want is to project the HR and LR input images to a UDFS in such a way that the discriminative information in the projected space is maximized. The projection matrices can be learned by using a training set consisting of HR and LR face pairs of the same subjects. As mentioned above, we assume that faces of the same person lie on the same manifold and those of different persons are associated with different manifolds. Hence, N HR training images can be denoted by a multi-manifold space $\mathcal{M}^h = \{\mathcal{M}_c^h\}_{c=1}^{C}$, and the corresponding N LR images of the same subjects can be denoted by the other multi-manifold space $\mathcal{M}^l = \{\mathcal{M}_c^l\}_{c=1}^{C}$, where C is the class number, $\mathcal{M}_c^h = \{I_i^h\}_{i=1}^{N_c}$ and $\mathcal{M}_c^l = \{I_i^l\}_{i=1}^{N_c}$ are the HR and LR manifolds of the c-th person respectively. Here, N_c is the sample number of the c-th person in the training set, $\sum_{c=1}^{C} N_c = N$. Let x_h and x_l denote the features of the HR and LR images respectively. Given training data, our goal is to find two projection matrices P_H and P_L to make the distances between the projected HR and LR feature vectors in the new feature space as close as possible, which means the discriminative information is maximized.

2.2. Discriminant graph on multi-manifold

In CPLM [33], it directly preserves the local geometric structure of the HR feature space in an unsupervised manner, neglecting the label information and the multi-manifold structure of the training samples. Therefore, in the similarity graph of CPLM, samples on different manifolds may be connected, which will diffuse information across manifolds and be misleading.

Inspired by some recent work on spectral graph learning [38], in this paper, we try to explore the intrinsic geometric structure of the samples on multi-manifold in a supervised way to the task of detecting multiple low-dimensional manifolds embedded in high-dimensional space. The motivation of our method is to keep the label information and local geometric structure simultaneously after graph preserving (making it intrinsically different from CLPM [33] and DSR [34]). In other words, in the UDFS, we expect that the samples are still close if they are from the same manifold, and samples from different manifolds are as far from each other as possible. To this end, we define two similarity graphs on the multi-manifold, i.e., an intra-manifold

similarity graph based on intra-manifold neighbors and an inter-manifold similarity graph based on inter-manifold neighbors. While a large number of graph construction approaches have been proposed in the computer vision communities recently [39], we apply the effective and efficient k-nearest-neighbor (k-NN) method to calculate the similarity matrices, intra-manifold similarity matrix W^w and inter-manifold similarity matrix W^b, as follows:

$$w_{ij}^{w}=\begin{cases}\dfrac{1}{n_w+1} & \text{if } x_i^h\in N_{n_w}(x_j^h) \text{ or } x_j^h\in N_{n_w}(x_i^h)\\ 0 & \text{otherwise}\end{cases}, \tag{1}$$

$$w_{ij}^{b}=\begin{cases}\dfrac{1}{n_b} & \text{if } x_i^h\in N_{n_b}(x_j^h) \text{ or } x_j^h\in N_{n_b}(x_i^h)\\ 0 & \text{otherwise}\end{cases}, \tag{2}$$

where $N_{n_w}(x_i^h)$ is the local set comprising n_w nearest neighbors of x_i^h belonging to the same person, while $N_{n_b}(x_i^h)$ is the set of n_b nearest neighbors of x_i^h belonging to other persons.

2.3. The objective function

Let $f:\mathbb{R}^D\to\mathbb{R}^d$ denote the mapping from the $\mathbb{R}^D$ input space to the $\mathbb{R}^d$ embedded space, D and d are their dimensions. We consider the mapping $f=(f_1,\ldots,f_d)$ as a linear combination of t basis functions of the following form:

$$f_i(x;P)=\sum_{j=1}^{t}p_{ji}\phi_j(x), \tag{3}$$

where $\phi_j(x)$, $j=1,\ldots t$, can be a linear or non-linear function, and $[P]_{ij}=p_{ij}$ is the weight matrix of size $t\times d$ to be determined. The mapping defined by Eq. (3) can be written in a more compact manner as follows:

$$f(x;P)=P^{\top}\phi(x). \tag{4}$$

Here, "$\top$" denotes the transpose of a matrix. Let P_H and P_L denote the projection matrices for the HR and LR feature vectors respectively. The basis functions for the HR and LR feature vectors are given by

$$\phi^h(x^h)=[\phi_1^h(x^h),\ldots,\phi_{t_h}^h(x^h)]^{\top}, \tag{5}$$

$$\phi^l(x^l)=[\phi_1^l(x^l),\ldots,\phi_{t_l}^l(x^l)]^{\top}, \tag{6}$$

where t_h and t_l represent the numbers of basis functions of the HR and LR feature vectors respectively, which can be same or different. Thus the mappings corresponding to the LR and HR images can be written as

$$f(x^h;P_H)=P_H^{\top}\phi^h(x^h), \tag{7}$$

$$f(x^l;P_L)=P_L^{\top}\phi^l(x^l). \tag{8}$$

The goal of CDMMA is to transform the feature vectors of the HR and LR images to make the distance between the transformed feature vectors which is called as corresponding preserving objective (J_{CP}) as close as possible. So we want to find the projection matrices P_H and P_L which minimizes the following objective function:

$$J_{CP}(P_H,P_L)=\sum_{i=1}^{N}\left\|P_H^{\top}\phi^h(x_i^h)-P_L^{\top}\phi^l(x_i^l)\right\|^2. \tag{9}$$

As discussed above, in order to improve the classification performance, the objective function in Eq. (9) should be modified to include discriminative information of the training samples. Thus a discriminative analysis objective (J_{DA}) can be added to the J_{CP}, resulting in the following objective function:

$$\begin{aligned}J(P_H,P_L)&=J_{CP}(P_H,P_L)+J_{DA}(P_H,P_L)\\&=\sum_{i=1}^{N}\left\|P_H^{\top}\phi^h(x_i^h)-P_L^{\top}\phi^l(x_i^l)\right\|^2+\alpha\sum_{i=1}^{N}\sum_{j=1}^{N}\left\|P_H^{\top}\phi^h(x_i^h)\right.\\&\quad\left.-P_L^{\top}\phi^l(x_j^l)\right\|^2 w_{ij}^{w}-\beta\sum_{i=1}^{N}\sum_{j=1}^{N}\left\|P_H^{\top}\phi^h(x_i^h)\right.\\&\quad\left.-P_L^{\top}\phi^l(x_j^l)\right\|^2 w_{ij}^{b},\end{aligned} \tag{10}$$

where α is a scalar controlling the relative contribution of intra-manifold compactness term and β is a scalar controlling the relative contribution of inter-manifold separability term.

Remark. Through the HR and LR face pairs are highly related, the representations (observations or descriptions) are various due to the influence of pose, lighting, expression, etc. Thus, it is very difficult for us to find the correspondence between the data points in the original observation spaces (e.g., directly mapping the LR face images to the corresponding HR face images or vice versa). From the geometric perspective, each observation set forms a manifold. Given that these observations are from the same object, it is reasonable to assume that some common features across different observation spaces can be represented in an underlying common manifold. Therefore, we think that the projection from HR face space to LR face space (or vice versa) is uncertain while projecting the HR faces and LR faces to the underling common space is reasonable to some extent as we discussed above.

2.4. Implementation

To solve the objective function (10), we define $\phi^h(X^h)=[\phi^h(x_1^h),\ldots,\phi^h(x_N^h)]\in\mathbb{R}^{t_h\times N}$, $\phi^l(X^l)=[\phi^l(x_1^l),\ldots,\phi^l(x_N^l)]\in\mathbb{R}^{t_l\times N}$. Here, $\phi^h(x_i^h)\in\mathbb{R}^{t_h}$ and $\phi^l(x_i^l)\in\mathbb{R}^{t_l}$ are the HR and LR feature vectors. We consider the parameterization of P_H and P_L in the forms of $P_H=V\phi^h(X^h)$ and $P_L=U\phi^l(X^l)$ respectively. With some algebraic deductions, the objective function (10) can be simplified as

$$\begin{aligned}&J(P_H,P_L)\\&=\mathrm{tr}\left(\begin{bmatrix}U\\V\end{bmatrix}^T\begin{bmatrix}(\phi^l(X^l))^T & 0\\0 & (\phi^h(X^h))^T\end{bmatrix}\begin{bmatrix}\phi^l(X^l) & 0\\0 & \phi^h(X^h)\end{bmatrix}\right.\\&\quad\times\begin{bmatrix}I+\alpha D_L^w-\beta D_L^b & -I-\alpha W^w+\beta W^b\\-I-\alpha W^w+\beta W^b & I+\alpha D_H^w-\beta D_H^b\end{bmatrix}\\&\quad\left.\times\begin{bmatrix}\phi^l(X^l) & 0\\0 & \phi^h(X^h)\end{bmatrix}\begin{bmatrix}(\phi^l(X^l))^T & 0\\0 & (\phi^h(X^h))^T\end{bmatrix}\begin{bmatrix}U\\V\end{bmatrix}\right)\end{aligned}$$

$$= \mathrm{tr}\left(\begin{bmatrix} U \\ V \end{bmatrix}^T \begin{bmatrix} K^l & 0 \\ 0 & K^h \end{bmatrix} Q \begin{bmatrix} K^l & 0 \\ 0 & K^h \end{bmatrix}^T \begin{bmatrix} U \\ V \end{bmatrix} \right) = \mathrm{tr}(A^T ZQZ^T A), \tag{11}$$

where $K^l = (\phi^l(X^l))^\top \phi^l(X^l)$ and $K^h = (\phi^h(X^h))^\top \phi^h(X^h)$, they can be Gaussian kernel functions,

$$Q = \begin{bmatrix} I + \alpha D_L^w - \beta D_L^b & -I - \alpha W^w + \beta W^b \\ -I - \alpha W^w + \beta W^b & I + \alpha D_H^w - \beta D_H^b \end{bmatrix}, \quad Z = \begin{bmatrix} K^l & 0 \\ 0 & K^h \end{bmatrix}, \quad A = \begin{bmatrix} U \\ V \end{bmatrix},$$

I is the identity matrix, the diagonal matrices D_l^w and D_h^w are defined based on the weight matrix as $[D_L^w]_{ii} = \sum_j w_{ij}^w$ and $[D_H^w]_{jj} = \sum_i w_{ij}^w$ respectively. Similarly, $[D_L^b]_{ii} = \sum_j w_{ij}^b$ and $[D_H^b]_{jj} = \sum_i w_{ij}^b$. Therefore, U and V can be seen as the new projection matrices in the UDFS.

We introduce the scaling constraints, which remove an arbitrary scaling factor in the projection, and solve the optimization problem as

$$\min_A J(P_H, P_L) \quad \text{s.t. } A^\top ZZ^\top A = I. \tag{12}$$

We will now switch to a Lagrangian formulation of the problem. The Lagrangian is as follows:

$$L = \mathrm{tr}(A^T ZQZ^T A) - \lambda A^\top ZZ^\top A. \tag{13}$$

Requiring that the gradient of L vanish gives the following eigenvector problem:

$$ZQZ^\top A = \Lambda ZZ^\top A. \tag{14}$$

Therefore, given the dimension d' of the unified space, the solution of Eq. (11) with respect to A can be calculated by the $d'-$st smallest generalized eigenvectors of Eq. (14). Then, the obtained matrix A of size $2N \times d'$ can be divided into two projection matrices $U \in \mathbb{R}^{N \times d'}$ and $V \in \mathbb{R}^{N \times d'}$. Note that the matrix $ZZ^\top$ is usually noninvertible, and can be adjusted to $ZZ^\top + \tau I$, where τ is a small positive value (say $\tau = 10^{-6}$).

The features of the LR face set $F^l = P_L \phi^l(X^l) = U^\top K^l$ and the HR face set $F^h = P_H \phi^h(X^h) = V^\top K^h$ are all coupled and mapped into the UDFS. In the testing phase, the kernel feature of a LR face probe x_t^l (l_t^l) is expressed as $\phi^l(x_t^l)$. Then it is projected into the UDFS:

$$f_t^l = P_L^\top \phi^l(x_t^l) = \sum_{i=1}^{N} u_i (\phi^l(x_i^l))^\top \phi^l(x_t^l) = U^\top \mathrm{k}^l(\bullet,\ x_t^l), \tag{15}$$

where $\mathrm{k}^l(\bullet,\ x_t^l) = [(\phi^l(x_1^l))^\top \phi^l(x_t^l), \ldots, (\phi^l(x_N^l))^\top \phi^l(x_t^l)]$. Finally, we apply the k-NN classifier ($k=1$) to query the identity of the HR face registration list F^h. The entire LR face recognition process is summarized in Algorithm 1.

Algorithm 1. LR face recognition via CDMMA.

1: **Input**: N HR and LR training images, $\{\{I_i^h\}_{i=1}^{N_c}\}_{c=1}^{C}$ and $\{\{I_i^l\}_{i=1}^{N_c}\}_{c=1}^{C}$, a LR face probe I_t^l, the model parameters n_w, n_b, α, and β.
2: **Training Phase:**
Extract the features of training images, $\{\{x_i^h\}_{i=1}^{N_c}\}_{c=1}^{C}$ and $\{\{x_i^l\}_{i=1}^{N_c}\}_{c=1}^{C}$.
3: Calculate the intra-manifold similarity matrix W^w and inter-manifold similarity matrix W^b according to Eqs. (1) and (2) respectively.
4: Obtain the D_l^w and D_h^w through $[D_L^w]_{ii} = \sum_j w_{ij}^w$ and $[D_H^w]_{jj} = \sum_i w_{ij}^w$ respectively.
5: Obtain the matrix Q through $Q = \begin{bmatrix} I + \alpha D_L^w - \beta D_L^b & -I - \alpha W^w + \beta W^b \\ -I - \alpha W^w + \beta W^b & I + \alpha D_H^w - \beta D_H^b \end{bmatrix}$
6: Calculate K^l and K^h through $K^l = (\phi^l(X^l))^\top \phi^l(X^l)$ and $K^h = (\phi^h(X^h))^\top \phi^h(X^h)$.
7: Obtain the matrix Z through $Z = \begin{bmatrix} K^l & 0 \\ 0 & K^h \end{bmatrix}$.
8: Construct the mapping matrix A according to Eq. (14).
9: Calculate the two projection matrices U and V through $U = A(1{:}N, :)$ and $V = A(N+1{:}2N, :)$ respectively.
10: Obtain the projected features of the HR and LR face sets through $F^h = P_H \phi^h(X^h) = V^\top K^h$ and $F^l = P_L \phi^l(X^l) = U^\top K^l$.
Testing Phase:
11: Extract the features of the input LR image, x_t^l.
12: Obtain the kernel feature through Eq. (15).
13: Apply the k-NN classifier ($k=1$) to query the identity of the HR face registration list F^h.
14: **Output**: The label of I_t^l.

2.5. Relation to existence methods

It is worth mentioning that CDMMA aims to obtain a UDFS for manifold alignment, where coupled manifold information are used, and this makes CDMMA different from the existing multi-manifold discriminant analysis approaches [35,40]. Specifically, for the works in [35,40–42], they also introduced the within-class and between-class similarity graphs to characterize intra-class compactness and inter-class separability, based on a graph embedding framework. However, they tried to learn one mapping matrix to project homogeneous data (from one domain) to DFS. They project data from one space to the potential low-dimensional discriminant space by one mapping matrix. In contrast, we propose to learn two mapping matrices to project heterogeneous data (from cross domain, i.e., the LR and HR face image manifolds) to a UDFS through a supervised manner. In our proposed method, data from two different spaces are project to the potential common low-dimensional discriminant space by two mapping matrices, i.e., coupled mapping. As far as we know, we are the first to the topic of coupled discriminant learning on multi-manifold in the literatures.

3. Experiments

In this section, we describe the details of extensive experiments performed to evaluate the usefulness of the proposed method to face recognition. In all our experiments, we use the HR face images as the gallery, and use the LR face images (which is smoothed and down-sampled from its HR versions) as the probe. It is worth noting that all the face images are aligned by some recently proposed automatic alignment methods [43] and feature points matching methods [44,45].

Fig. 4. Samples of Extended Yale-B database.

Fig. 5. CMU PIE face database.

3.1. Dataset configuration

To investigate the performance of CDMMA, two standard FR databases are used: Extended Yale-B [46] and CMU PIE face [47].

Extended Yale-B database contains 16,128 face images of 38 subjects under 9 posed and 64 illumination conditions. In our experiment, we choose the frontal pose and use all the images under different illuminations, thus we get 2414 image in total. All the face images are manually aligned and cropped. They are resized to 32×28 pixels, with 256 gray levels per pixel.

CMU PIE face database contains 68 individuals with 41,368 face images as a whole. The face images were captured by 13 synchronized cameras and 21 flashes, under varying pose, illumination and expression. In our experiment, one near fontal poses (C27) are elected under different illuminations, lighting and expressions which leaves us 49 near frontal face images for each individual. Similarly, they are resized to 32×28 pixels, with 256 gray levels per pixel. Some samples are shown in Fig. 4.

For each database, we randomly divide it into two equal parts. One half (containing 32 images in Extended Yale-B database and 25 images in CMU PIE face database for each person) is used as the training set (HR gallery), and the other half (LR probe is smoothed down-sampled by an averaging filter H of size 4×4 and down-sampled by a factor of 4 to 8×7 pixels) is used for testing. Some samples are shown in Fig. 5.

3.2. Comparison with the baseline methods

The performance of directly comparing HR version of probe images against HR gallery images and comparing LR probe images with down-sampled gallery images are given as baselines for comparison, which are denoted in bold *HR* and bold *LR* respectively. Recently, sparse representation based classification (SRC) methods [48,49] have been widely used for FR. We also compare the performance of HR and LR gallery images with SRC FR engine, which are denoted as *HR*+SRC and *LR*+SRC respectively. Note that, in practical, the cases *HR* and *HR*+SRC do not exist given the LR observation.

Results and analysis: Tables 1 and 2 tabulate the recognition rates of the baseline methods with different dimensions on Extended Yale-B database and CMU PIE database respectively. From the results we learn that

- Compared with the *HR* method, there is a significant drop (32.1% in Extended Yale-B while 7.9% in CMU PIE face) in recognition accuracy from the LR image. This indicates that it is unwise to perform matching in the LR image space by down-sampling the gallery of HR face images;
- We list the performance of SRC based methods (*LR*+SRC and the idea case *HR*+SRC) similarly in Tables 1 and 2. They all exceed the *LR* and *HR* methods with 1-NN classifier especially in Extended Yale-B database. In addition, the performance of CDMMA is better than that of *LR*+SRC (2.2% improvement in Extended Yale-B while 4.6% improvement in CMU PIE face), and even better than that of *HR*+SRC in CMU PIE face database (0.6% improvement).

3.3. Comparison with the state-of-the-art methods

We compare CDMMA with CLPM [33], DSR [34] and some two-step LBSR methods, which include Bicubic, Wang and Tang [18] eigentransformation based super-resolution (ESR), Ma et al.'s [23] position patch based super-resolution (PSR) and Yang et al.'s [50] sparse representation based super-resolution (SSR). For CDMMA and CLPM, we directly

Table 1
Recognition rates of the baseline methods with different dimensions on Extended Yale-B database.

Dimensions	10	20	30	50	100	200
HR	0.0855	0.3118	0.4197	0.5373	0.6565	0.7363
LR	0.4150	0.4150	0.4150	0.4150	0.4150	0.4150
HR+SRC	0.3569	0.7341	0.8366	0.8978	0.9375	0.9639
LR+SRC	0.8797	0.8797	0.8797	0.8797	0.8797	0.8797
CDMMA	0.7608	0.8746	0.8969	0.8925	0.8612	0.7546

Table 2
Recognition rates of the baseline methods with different dimensions on CMU PIE database.

Dimensions	10	20	30	50	100	200
HR	0.6909	0.8432	0.8771	0.9031	0.9233	0.9360
LR	0.8574	0.8574	0.8574	0.8574	0.8574	0.8574
HR+SRC	0.8001	0.9349	0.9585	0.9709	0.9761	0.9744
LR+SRC	0.9357	0.9357	0.9357	0.9357	0.9357	0.9357
CDMMA	0.8484	0.9596	0.9769	0.9821	0.9551	0.9115

use the face images themselves as features to be classified (we normalize each vector to unit in all the experiments). For these two-step LBSR methods (as well as the BI and HR methods), we use the PCA initial dimension reduction to obtain the feature representations. Note that we can use more discriminative features and FR engines to test our method. We simply use the PCA coefficients and 1-NN classifier in this work. For CDMMA, we empirically set $n_w \in [4,7]$, $n_b \in [50,200]$, $\alpha \in [4,40]$, $\beta \in [1,10]$.

From our empirical study, good performance can be achieved when β/α takes a value between 0.2 and 0.5. This indicates that the intra-manifold compactness term is more important than the inter-manifold separability term. Please refer to the following subsection for more details about the parameter settings. The kernel mapping ϕ of CDMMA is set to identity (i.e., $\phi(x)=x$) to highlight the performance improvement due to the proposed learning approach. In our experiments, we find that there is only limited improvement (around 2% in Extended Yale-B and 0.3% in CMU PIE face) by using the RBF kernel metric (using Gaussian function as kernels) and the parameter tuning is very complex.

Results and analysis: In Figs. 6 and 7, the average recognition rates (all the experiments are repeated 10 times) with different feature dimensions on Extended Yale-B database and CMU PIE face database are drawn. On the other hand, the mean recognition rates (in percents) of these 10 runs with optimal dimension are reported in Figs. 8 and 9 respectively. We can observe that

- Not all LBSR methods benefit the FR task. Existing face SR algorithms such as PSR and ESR can slightly improve the performance compared with *LR* method, but are not significant, whereas Bicubic interpolation (can be seen as a super-resolution method) drops the performance. Note that SSR gains a similar performance of the idea case *HR* and this is due to the sparse prior, which has been proven to be effective for image recognition task.
- CLPM and DSR are better than existing LBSR methods except SSR and this can be attributed to the introduction of discriminative information in those methods. Through SSR is a two-step method, it can reconstruct very discriminative faces with the introduction of sparsity and surpass DSR.
- The proposed CDMMA method significantly improves the performance when compared to the LBSR approaches and outperforms CLPM and DSR. When compared with CLPM and DSR, the improvement of CDMMA is around 2.5% and 27.7% in Yale-B database and 2.1% and 7.2% in CMU PIE face database respectively. This can be mainly explained by that CDMMA successfully incorporates the label information as well as local geometric structure of the training samples, while CLPM and DSR consider only one certain aspect.
- For these PCA based feature representation method, e.g., BI, ESR, SSR, PSR, and DSR, with the increase of the feature dimensionality, more information will be introduced to the classification, thus the recognition rate will increase or be stable when the feature dimensionality reach some values.
- It can be seen that the performance of CDMMA decreased when the dimension of the feature rose. This seems to be not normal compared with other methods (e.g., the methods used to be compared) whose performance usually increases when the feature dimension increases. We attribute this to the exitance of shared UDFS of LR and HR training samples. This also proves the rationality of projecting the original LR and HR faces to the UDFS. We also explain this point from another perspective, CDMMA tries to explore the discriminant information of the face data space, and project the original data to the discriminant subspace for classification (similar with the idea of LDA). Thus, projected subspace with small dimensionality will be very hard to capture all the discriminant information, while projected subspace that has large dimensionality will incorporate non-discriminant information into the discriminant subspace.

3.4. Sensitivity to parameters

In this subsection, we study the performance variation of CDMMA with respect to the number of nearest neighbors, n_w and n_b, in inter-manifold similarity graph and intra-manifold similarity graph respectively. Here, we use the Extended Yale-B database for illustration. The experimental results are shown in Fig. 10. We can see from Fig. 10 that the performance can achieve a relatively high recognition rate when $n_w \in [5,10]$ and $n_b \in [50,150]$. When the value of n_w is set to large, the recognition rate will fall slightly (i.e., $n_w=30$, which means connecting all the samples of the same class and the recognition rate is 88.7%). This proves that the neighbor information of intra-manifold is very important for discriminative analysis.

In addition, we also study the performance variation of CDMMA with respect to the regularization parameter α and β in (10). We again use the Extended Yale-B database for illustration. Fig. 11 illustrates the accuracy of CDMMA when α and β take different values. From that figure, we

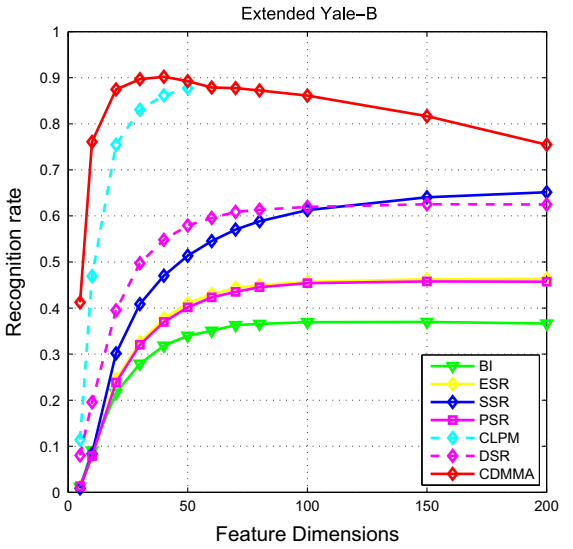

Fig. 6. Recognition results of different methods with different dimensions on Extended Yale-B database.

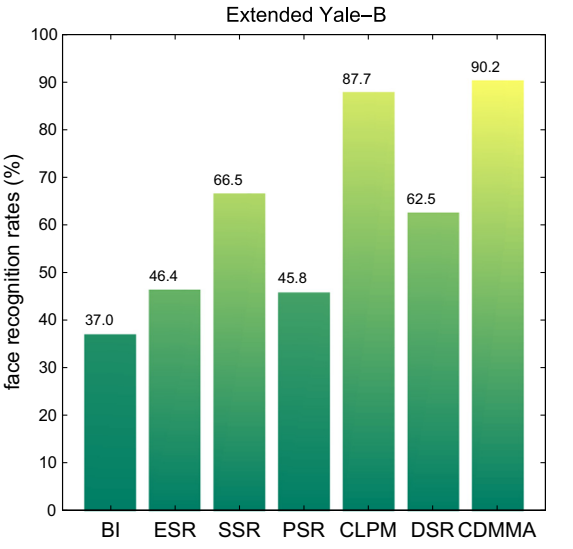

Fig. 8. Recognition results with the optimal dimension on Extended Yale-B database.

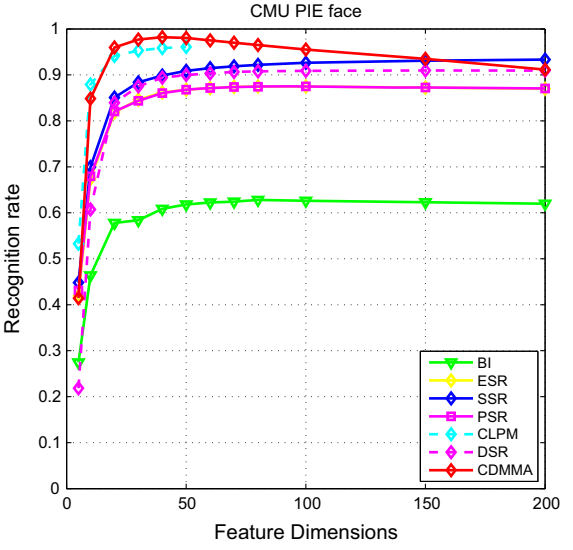

Fig. 7. Recognition results of different methods with different dimensions on CMU PIE face database.

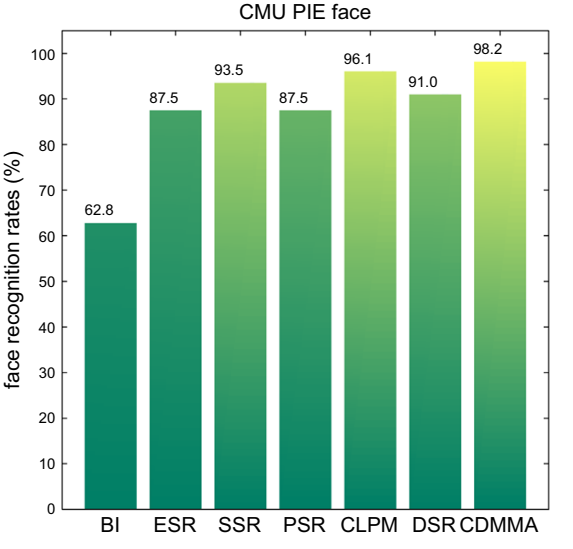

Fig. 9. Recognition results with the optimal dimension on CMU PIE face database.

can see that the smaller the α, the worse the performance will be. Larger α means that the local geometric structure of intra-manifold plays a more significant role in the learning procedure. Hence, we can conclude that the neighbor information of intra-manifold is very important. We can see from Fig. 11 that the performance can achieve at a stable level when $\alpha \in [0.4, 1]$ and $\beta \in [0.02, 0.1]$.

4. Conclusion

In this paper, we have proposed a novel coupled discriminant multi-manifold analysis (CDMMA) method to address the low-dimensional FR problem. Specifically, our approach employs the neighborhood information as well as local geometric structure of the manifold and derives two coupled mappings on multi-manifold space to project LR and HR feature spaces to UDFS, simultaneously maximizing a measure of discriminatory power and preserving the geometrical structure of the manifold. Experimental results on the same database have demonstrated the superiority of the proposed method over some state-of-the-art methods.

Finally, note that although all the experiments in this paper deal with low to high resolution FR problem, the proposed approach is very general and can be applied to any other heterogeneous face recognition (HFR) problem, such as photo-sketch recognition, near infrared (NIR) faces to visual (VIS) faces recognition.

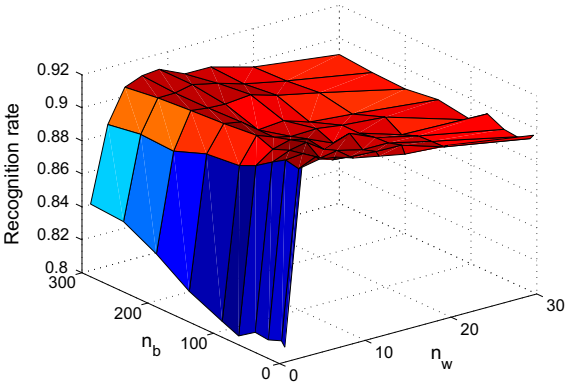

Fig. 10. The recognition rates of our proposed method with different values of n_w and n_b based on the Extended Yale-B database: n_w and n_b vary in the interval [1,30] and [5,300] respectively.

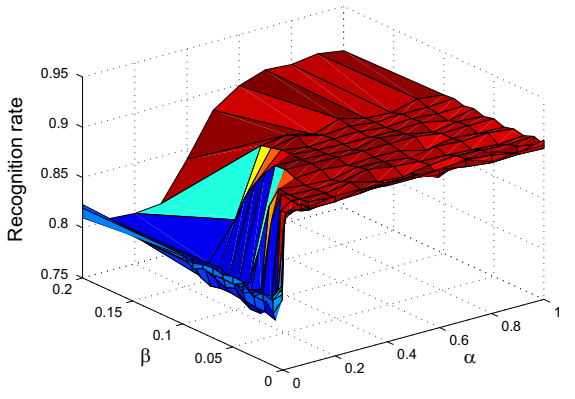

Fig. 11. The recognition rates of our proposed method with different values of α and β based on the Extended Yale-B database: α and β vary in the interval [0, 1] and [0, 0.2] respectively.

Acknowledgment

The research was supported by the National Natural Science Foundation of China under Grant 61501413, in part by the Fundamental Research Funds for the Central Universities, China University of Geosciences, Wuhan, China, in part by the National Natural Science Foundation of China under Grant 61503288, and in part by the Open Foundation of Hubei Provincial Key Laboratory of Intelligent Robot under Grant HBIR201404.

References

[1] D. Tao, L. Jin, W. Liu, X. Li, Hessian regularized support vector machines for mobile image annotation on the cloud, IEEE Trans. Multimed. 15 (4) (2013) 833–844.
[2] D. Tao, X. Lin, L. Jin, X. Li, Principal component 2-d long short-term memory for font recognition on single chinese characters, IEEE Trans. Cybern. PP (99) (2015) http://dx.doi.org/10.1109/TCYB.2015.2414920. 1–1.
[3] D. Tao, L. Jin, Y. Wang, Y. Yuan, X. Li, Person re-identification by regularized smoothing kiss metric learning, IEEE Trans. Circuits Syst. Video Technol. 23 (10) (2013) 1675–1685.
[4] C. Chen, R. Jafari, N. Kehtarnavaz, Improving human action recognition using fusion of depth camera and inertial sensors, IEEE Trans. Hum.–Mach. Syst. 45 (1) (2015) 51–61.
[5] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.
[6] X. Ben, W. Meng, R. Yan, K. Wang, Kernel coupled distance metric learning for gait recognition and face recognition, Neurocomputing 120 (2013) 577–589.
[7] P. Hennings-Yeomans, S. Baker, B. Kumar, Simultaneous super-resolution and feature extraction for recognition of low-resolution faces, in: IEEE Conference on Computer Vision and Pattern Recognition (2008), 2008, pp. 1–8. http://dx.doi.org/10.1109/CVPR.2008.4587810.
[8] Z. Wang, W. Yang, X. Ben, Low-resolution degradation face recognition over long distance based on cca, Neural Comput. Appl. 26 (7) (2015) 1–8.
[9] B. Du, L. Zhang, A discriminative metric learning based anomaly detection method, IEEE Trans. Geosci. Remote Sens. 52 (11) (2014) 6844–6857.
[10] B. Du, L. Zhang, Target detection based on a dynamic subspace, Pattern Recognit. 47 (1) (2014) 344–358.
[11] B. Du, Y. Zhang, L. Zhang, L. Zhang, A hypothesis independent sub-pixel target detector for hyperspectral images, Signal Process. 110 (2015) 244–249.
[12] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, J. Tian, Robust feature matching for remote sensing image registration via locally linear transforming, IEEE Trans. Geosci. Remote Sens. 53 (12) (2015) 6469–6481.
[13] J. Ma, J. Zhao, J. Tian, A.L. Yuille, Z. Tu, Robust point matching via vector field consensus, IEEE Trans. Image Process. 23 (4) (2014) 1706–1721.
[14] C. Chen, L. Liu, L. Chen, Y. Tang, Y. Zhou, Weighted couple sparse representation with classified regularization for impulse noise removal, IEEE Trans. Image Process. 24 (11) (2015) 4014–4026, http://dx.doi.org/10.1109/TIP.2015.2456432.
[15] L. Liu, C.L.P. Chen, Y. Zhou, X. You, A new weighted mean filter with a two-phase detector for removing impulse noise, Inf. Sci. 315 (2015) 1–16.
[16] S. Baker, T. Kanade, Hallucinating faces, in: Proceedings of IEEE Conference on Automatic Face and Gesture (FG), 2000, pp. 83–88.
[17] C. Liu, H.-Y. Shum, C.-S. Zhang, A two-step approach to hallucinating faces: global parametric model and local nonparametric model, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, 2001, pp. I-192–I-198.
[18] X. Wang, X. Tang, Hallucinating face by eigentransformation, IEEE Trans. Syst. Man Cybern. Part C 35 (3) (2005) 425–434.
[19] A. Chakrabarti, A. Rajagopalan, R. Chellappa, Super-resolution of face images using kernel pca-based prior, IEEE Trans. Multimed. 9 (4) (2007) 888–892.
[20] S.W. Park, M. Savvides, Breaking the limitation of manifold analysis for super-resolution of facial images, in: Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, 2007, pp. I-573–I-576.
[21] H. Chang, D. Yeung, Y. Xiong, Super-resolution through neighbor embedding, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, 2004, pp. 275– 282.
[22] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding 290 (5500) (2000) 2323–2326.
[23] X. Ma, J. Zhang, C. Qi, Hallucinating face by position-patch, Pattern Recognit. 43 (6) (2010) 2224–2236.
[24] C. Jung, L. Jiao, B. Liu, M. Gong, Position-patch based face hallucination using convex optimization, IEEE Signal Process. Lett. 18 (6) (2011) 367–370.
[25] J. Jiang, R. Hu, Z. Wang, Z. Han, Noise robust face hallucination via locality-constrained representation, IEEE Trans. Multimed. 16 (5) (2014) 1268–1281.
[26] J. Jiang, R. Hu, C. Liang, Z. Han, C. Zhang, Face image super-resolution through locality-induced support regression, Signal Process. 103 (2014) 168–183.
[27] J. Jiang, R. Hu, Z. Wang, Z. Han, Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning, IEEE Trans. Image Process. 23 (10) (2014) 4220–4231.
[28] J. Jiang, R. Hu, Z. Wang, Z. Han, J. Ma, Facial image hallucination through coupled-layer neighbor embedding, IEEE Trans. Circuits Syst. Video Technol. PP (99) (2015) http://dx.doi.org/10.1109/TCSVT.2015.2433538.
[29] Y. Hu, K.-M. Lam, G. Qiu, T. Shen, From local pixel structure to global image super-resolution: a new face hallucination framework, IEEE Trans. Image Process. 20 (2) (2011) 433–445.

[30] K. Jia, S. Gong, Multi-modal tensor face for simultaneous super-resolution and recognition, in: Proceedings of IEEE Conference on Computer Vision (ICCV), vol. 2, 2005, pp. 1683–1690.
[31] S. Biswas, K.W. Bowyer, P.J. Flynn, Multidimensional scaling for matching low-resolution face images, IEEE Trans. Pattern Anal. Mach. Intell. 34 (10) (2012) 2019–2030.
[32] H. Huang, H. He, Super-resolution method for face recognition using nonlinear mappings on coherent features, IEEE Trans. Neural Netw. 22 (1) (2011) 121–130.
[33] B. Li, H. Chang, S. Shan, X. Chen, Low-resolution face recognition via coupled locality preserving mappings, IEEE Signal Process. Lett. 17 (1) (2010) 20–23.
[34] W.Z. Wilman, P.C. Yuen, Very low resolution face recognition problem, IEEE Trans. Image Process. 21 (1) (2012) 327–340.
[35] A.B. Goldberg, X. Zhu, A. Singh, Z. Xu, R. Nowak, Multi-manifold semi-supervised learning, J. Mach. Learn. Res. – Proc. Track (2009) 169–176.
[36] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, D. Cai, Graph regularized sparse coding for image representation, IEEE Trans. Image Process. 20 (5) (2011) 1327–1336.
[37] J. Jiang, R. Hu, Z. Han, L. Chen, J. Chen, Coupled discriminant multi-manifold analysis with application to low-resolution face recognition, in: Proceedings of 21st International Conference on MultiMedia Modeling, MMM 2015, Part I, Sydney, NSW, Australia, January 5–7, 2015, 2015, pp. 37–48.
[38] Fan RK. Chung, Spectral graph theory, American Mathematical Soc. 92 (1997).
[39] L. Zhu, S. Zhu, Face recognition based on orthogonal discriminant locality preserving projections, Neurocomputing 70 (7–9) (2007) 1543–1546.
[40] M. Harandi, C. Sanderson, S. Shirazi, B. Lovell, Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2011, pp. 2705–2712.
[41] W. Yang, C. Sun, L. Zhang, A multi-manifold discriminant analysis method for image feature extraction, Pattern Recognit. 44 (8) (2011) 1649–1657.
[42] J. Lu, Y.-P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 39–51.
[43] J. Ma, W. Qiu, J. Zhao, Y. Ma, A.L. Yuille, Z. Tu, Robust l_2e estimation of transformation for non-rigid registration, IEEE Trans. Signal Process. 63 (5) (2015) 1115–1129.
[44] J. Ma, J. Zhao, Y. Ma, J. Tian, Non-rigid visible and infrared face registration via regularized gaussian fields criterion, Pattern Recognit. 48 (3) (2015) 772–784.
[45] J. Ma, J. Zhao, J. Tian, X. Bai, Z. Tu, Regularized vector field learning with sparse approximation for mismatch removal, Pattern Recognit. 46 (12) (2013) 3519–3532.
[46] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.
[47] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, Image Vis. Comput. 28 (5) (2010) 807–813.
[48] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[49] A.Y. Yang, A. Ganesh, Z. Zhou, S. Sastry, Y. Ma, A review of fast l1-minimization algorithms for robust face recognition, 2010, CoRR abs/1007.3753.
[50] J. Yang, H. Tang, Y. Ma, T. Huang, Face hallucination via sparse coding, in: Proc. IEEE Conf. Image Process. (ICIP), 2008, pp. 1264–1267.