Information Sciences 506 (2020) 19–36

Contents lists available at ScienceDirect

Information Sciences

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

Cross-resolution face recognition with pose variations via multilayer locality-constrained structural orthogonal procrustes regression

Guangwei Gao[a,b,c,*], Yi Yu[b], Meng Yang[d], Heyou Chang[e], Pu Huang[a], Dong Yue[a]

[a] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China
[b] Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo, Japan
[c] Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou, China
[d] School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China
[e] Key Laboratory of Trusted Cloud Computing and Big Data Analysis, Nanjing XiaoZhuang University, Nanjing, China

ARTICLE INFO

Article history:
Received 24 April 2019
Revised 31 July 2019
Accepted 2 August 2019
Available online 2 August 2019

Keywords:
Face recognition
Low-resolution
Pose variations
Nuclear norm

ABSTRACT

In real video surveillance scenes, the extracted face regions generally have low-resolution (LR) and are sensitive to pose and illumination variations; these flaws undoubtedly degrade the subsequent recognition task. To overcome these challenges, we propose an approach named multilayer locality-constrained structural orthogonal Procrustes regression (MLCSOPR). The proposed MLCSOPR not only learns the pose-robust discriminative representation features to reduce the resolution gap between the LR image space and the high-resolution (HR) one but also strengthens the consistency between the LR and HR image space. In particular, several contributions are made in this paper: (i) Inspired by the orthogonal Procrustes problem (OPP), a matrix approximation is exploited to find an optimal correction between two data matrices. (ii) The nuclear norm constraint is applied to the reconstruction error to maintain the structural property. (iii) Based on the abovementioned learned resolution-robust representation features, a linear regression-based classification strategy is adopted to recognize the LR input face images. Experiments on commonly used face databases have shown the effectiveness of the proposed method on cross-resolution face matching with pose variations.

1. Introduction

Over the past decades, face recognition technology has achieved significant progress [6,7,28–30,32,34,35,38,39]. In particular, pose-robust recognition tasks have been performed on cases where both the gallery and probe faces have the same resolution [32,38]. However, because of the long distance between the object and the surveillance system, the extracted face regions usually have low-resolution (LR), so they lack detailed features that enable recognition. This research scenario is regarded as cross-resolution face recognition, to which previous face recognition methods cannot be directly applied. Generally, existing approaches can be classified into three classes: (i) matching the acquired LR images to the down-sampled high-resolution (HR) galleries; (ii) matching the up-sampled LR images to the HR galleries; (iii) extracting discriminative fea-

* Corresponding author at: Institute of Advanced Technology, No.9 Wenyuan Road, Nanjing 210046, China
E-mail address: csggao@gmail.com (G. Gao).

https://doi.org/10.1016/j.ins.2019.08.004
0020-0255/

tures in the respective manifold to perform the matching task on the obtained feature space. The down-sampling strategy may reduce the resolution gap, but this procedure inevitably loses many useful discriminative facial details for recognition.

Super-resolution based methods: Researchers have exploited extensive learning-based super-resolution (SR) approaches to acquiring the desired HR face image from observed LR image to reduce the resolution discrepancy. Barker and Kanade [4] first introduced the term "face hallucination" and presented a Bayesian formulation based method. Jian et al. [14] presented a singular value decomposition (SVD)-based mapping to preserve the domain information in the HR face image space. Recently, many smaller patches based SR techniques have been proposed. Unlike previous methods adopting a fixed number of neighbors for representation learning, Li and Liang [21,22] introduced sparse representation technology to shape the prior model, which guides the reconstruction process. Lately, since the position prior becomes crucial in face reconstruction, Ma et al. [25] presented a local patch-based face hallucination approach. To alleviate over-fitting in [25], Jung et al. [15] incorporated sparsity before further enhancing the face reconstruction qualities. Then, Jiang et al. [16,17] presented a locality-constrained iterative neighbor embedding scheme to reveal the true topology of the nonlinear manifold. More recently, Jiang et al. [18] introduced smooth prior to obtain stable reconstruction weights. A robust locality-constrained bi-layer representation model is proposed by Liu et al. [23] to hallucinate the desired face images and to remove noise simultaneously. These abovementioned learning-based SR methods aim to obtain the optimal representation on a given training set instead of acquiring the discriminative features in the face super-resolution procedure.

Resolution-robust feature extraction based methods: Compared to the visual quality of the reconstructed faces, resolution-robust feature extraction based methods take face recognition performance into account. Moutafis and Kakadiaris [27] jointly optimized metric learning and representation learning and then used the learned metric for matching. Bhatt et al. [1] improved LR-HR face matching performance by applying ensemble-based co-transfer learning. Yang et al. [40] presented a united face hallucination and recognition framework via sparse representation. Mudunuri and Biswas [28] applied a multidimensional scaling (MDS) scheme to learn a joint transformation function to preserve the distance between LR and HR faces. Inspired by manifold learning, many common space-based approaches have been presented. Wang et al. [36] wanted to use a coupled mapping to depict the non-linearity between LR and HR faces in terms of a low dimensional embedding. Inspired by the supervised approach, Jiang et al. [19] embedded the face pairs into a discriminative feature space. Lu et al. [24] proposed to simultaneously extract the discriminative features and learn the mapping function from the LR to the HR feature space via a semi-coupled representation feature learning. A discriminant correlation analysis method that was introduced by Haghighat and Abdel-Mottaleb [12] projects the features extracted from LR and HR images into a common space. Recently, Banerjee and Das [2] proposed a generative adversarial network-based deep method to reconstruct HR images from LR probe images for recognition. Additionally, Gao et al. [8] proposed obtaining resolution-robust representation features for cross-resolution face recognition with occlusions. However, when learning the representation features, these methods do not consider pose variations, which degrade the stability of the learned features.

In this work, we design a multilayer locality-constrained structural orthogonal Procrustes regression (MLCSOPR) scheme to improve cross-resolution face matching with pose variations. Existing methods usually perform the recognition task on super-resolved faces or perform resolution-robust feature extraction on both LR and HR faces directly. In contrast, our proposed method first uses a structural orthogonal Procrustes regression scheme to learn discriminative representation features that are robust to pose variations in their respective HR and LR face space. Then our method performs recognition based on these learned resolution-robust features. Specifically, a matrix approximation method is used to find an optimal transformation matrix to adjust the pose from one image to that of another. Furthermore, by taking into account the illumination variations and possible structural noises (e.g., disguise or occlusion) in the probe images, the structure constraint (i.e., nuclear norm) is applied to the reconstruction error to maintain the image's appearance properties. Additionally, we incorporate locality-constrained regularization into our method to strengthen the discriminability of the learned features. Based on these learned resolution- and pose-robust discriminative features in their respective image spaces, the sparse representation induced method has a better capability of recognizing the input LR probe. Experimental tests on some commonly used face databases have verified the superiority of our proposed method.

This paper is an extension of our previous conference paper [9] with the following improvements: (i) a MLCSOPR model to further improve performance, (ii) an in-depth analysis of the proposed method, and (iii) extensive experimental evaluations of the method's performance. We organize the rest of our paper as follows: Some related works are briefly reviewed in Section 2. The proposed method is described in Section 3. A relevant analysis of the proposed method is given in Section 4. The performance evaluation is given in Section 5. Future work and conclusions are provided in Section 6.

2. Related work

2.1. Orthogonal procrustes problem

The orthogonal Procrustes problem (OPP) [10,13] is briefly discussed in this section. The OPP desires to find an optimal reflection that adjusts a matrix X to approximate another target matrix B.

Without loss of generality, the two-dimensional data and the orthogonal matrix can be denoted as X and $B \in \Re^{p \times q}$, $Q \in \Re^{p \times q}$, respectively. Then, the OPP can be denoted as

$$\min_{Q} \|XQ - B\|_F^2, \quad s.t. \quad Q^T Q = I_q, \tag{1}$$

where $||\cdot||_F$ is the Frobenius norm and I_q is a q-dimensional identity matrix. By applying the singular value decomposition (SVD): $USV^T = B^TX$, the analytical solution of objective (1) can be derived by $\hat{Q} = VU^T$ [33]. Additionally, the orthogonal transformation Q can be placed on the left side of data matrix X:

$$\min_{Q} \|QX - B\|_F^2, \quad s.t. \quad Q^TQ = I_p. \tag{2}$$

Here, I_p is a p-dimensional identity matrix. By using the SVD scheme: $USV^T = XB^T$, the analytical solution of objective (2) can be easily obtained by $\hat{Q} = VU^T$[33]. As mentioned in [32], we mainly focus on the horizontal direction pose in this job.

2.2. Semi-coupled locality-constrained representation

Researchers in [24] employed a locality-constrained representation (LCR) [20] scheme to simultaneously extract the discriminative representation features and study the transformation from the LR to the HR feature space. $X_h \in \Re^{t \times n}$ is the t-dimensional HR dataset; n is the size of the set; $D_h \in \Re^{t \times m}$ denotes the t-dimensional HR dictionary; m denotes the dictionary size; and $\varphi_h \in \Re^{m \times n}$ represents the HR LCR representation matrix. Their LR counterparts to the HR space are $X_l \in \Re^{d \times n}$, $D_l \in \Re^{d \times m}$, and $\varphi_l \in \Re^{m \times n}$.

The objective can be denoted by

$$\min_{D_h, D_l, \varphi_h, \varphi_l, W} \|X_h - D_h\varphi_h\|_F^2 + \|X_l - D_l\varphi_l\|_F^2 + \lambda_1 \sum_{i=1}^{n} \left\| l_h^i \otimes \alpha_h^i \right\|_2^2 + \lambda_2 \sum_{i=1}^{n} \left\| l_l^i \otimes \alpha_l^i \right\|_2^2 + \lambda_3 \|\varphi_h - W\varphi_l\|_F^2 + \lambda_4 \|W\|_F^2, \tag{3}$$

where symbol $\otimes$ represents the element-wise product, α_h^i and α_l^i denote the ith HR and LR representation features, respectively, l_h^i and l_l^i are m-dimensional vectors that contain the similarity between the ith input image and each dictionary atom in the HR and LR spaces, respectively, and W denotes the mapping function between the HR and the LR feature matrices. In such a semi-coupled learning strategy, both the HR and LR dictionaries and mapping function are optimized simultaneously.

3. The proposed MLCSOPR

We detail our proposed method in this section. Our main goal is to reduce the resolution gap between different image spaces by learning resolution-robust representation features. First, we present our representation feature learning formulation in Section 3.1. Second, based on these representation features, we introduce the subsequent classification procedure in Section 3.2. Third, we also present the complexity analysis in Section 3.3. Finally, to make the manifold geometry of the acquired LR image space consistent with that of the target HR one, we present our proposed MLCSOPR.

3.1. Overview of LCSOPR

The well-aligned training samples can be represented as $A_i \in \Re^{p \times q}$ $(i = 1,..., N)$, where p and q denote the size of each image. Then, the so-called linear mapping can be represented as

$$A(x) = \sum_{i=1}^{N} x_i A_i, \tag{4}$$

where x_i represents the reconstruction coefficient. Generally, the face probe image $y \in \Re^{p \times q}$, which has a frontal pose, can be linearly represented by $A_1, A_2, ..., A_N$: $y = A(x) + E$; $E \in \Re^{p \times q}$ denotes the reconstruction residual. Nevertheless, in real-world applications, the poses in the test image may differ from those in the training images. To address this problem, the probe image can be reformulated as yQ ($Q \in \Re^{q \times q}$, which has the same definition as that in Section 2.1). Then, the corrected probe yQ can be linearly represented by

$$yQ = \sum_{i=1}^{N} x_i A_i + E. \tag{5}$$

Researchers in [7,39] revealed that, compared with the L_2-norm or L_1-norm, nuclear norm regularization is more suitable for depicting structural noise. Based on this observation, we apply the nuclear norm constraint on the reconstruction residual to maintain the image's structural information. Then, the orthogonal Procrustes regression can be reformulated as

$$\min_{x,Q} \|yQ - A(x)\|_*, \quad s.t. \quad Q^TQ = I, \tag{6}$$

where $||\cdot||_*$ represents the nuclear norm, which accumulates the singular values of a matrix. To describe the prior information from neighbors, the local manifold constraint is also imposed on the weight via a distance between the input image and each training image. Finally, we formulate the proposed LCSOPR scheme as

$$\min_{x,Q} \|yQ - A(x)\|_* + \lambda \|d \otimes x\|_2^2, \quad s.t. \quad Q^TQ = I, \tag{7}$$

where $d = (d_1, \ldots, d_N)^T$ is a similarity vector with $d_i = \|y - A_i\|_F^2$ denoting the distance between the input and each training atom, $\otimes$ denotes the element-wise product and λ is a parameter.

For convenience, we can reformulate our objective (7) as follows:

$$\begin{aligned} &\min_{x,Q,E} \|E\|_* + \lambda\|d \otimes x\|_2^2 \\ &s.t. \quad yQ - A(x) = E. \end{aligned} \tag{8}$$

The alternating direction method of multipliers (ADMM) [7,39] can be used to optimize the above objective using the subsequent augmented Lagrangian form:

$$L_\mu(x, Q, E) = \|E\|_* + \lambda\|d \otimes x\|_2^2 + Tr\left(Z^T(yQ - A(x) - E)\right) + \frac{\mu}{2}\|yQ - A(x) - E\|_F^2, \tag{9}$$

where μ is a penalty parameter, $Tr(\cdot)\|$ denotes the trace operator and Z is the Lagrange multiplier. Next, we detail the procedure to solve problem (7).

(1) *Updating x:* Given E and Q, the objective can be rewritten as

$$x^{k+1} = \arg\min_x \|d \otimes x\|_2^2 + \frac{\mu}{2\lambda}\left\|yQ - E + \frac{1}{\mu}Z - A(x)\right\|_F^2. \tag{10}$$

Inspired by the literature [20], the analytical solution of objective (10) can be obtained by

$$x^{k+1} = \left(G + \tau D^2\right)\backslash ones(N, 1), \tag{11}$$

where $\tau = 2\lambda/\mu$, $N \times 1$ column vector $ones(N,1)$ is padded with entries of ones, $N \times N$ diagonal matrix D is padded with entries $D_{mm} = d_m$, and the operator "\" denotes the left matrix division operation, while $G = C^TC$ denotes the covariance matrix with

$$C = Vec\left(yQ - E + \frac{1}{\mu}Z\right)ones(N, 1)^T - H. \tag{12}$$

Here, $H = [Vec(A_1), Vec(A_2), \ldots, Vec(A_N)]$, where $Vec(\cdot)$ denotes the operator that converts a matrix to a column vector. The final optimal solution of x is rescaled to satisfy $\sum_{m=1}^{N} x_m = 1$.

(2) *Updating Q:* Given E and x, the objective can be rewritten as

$$Q^{k+1} = \arg\min_Q \frac{\mu}{2}\|yQ - P\|_F^2, \tag{13}$$

where $P = A(x) + E\text{-}Z/\mu$. By several simple matrix transformations, we obtain

$$\|yQ - P\|_F^2 = tr\left((yQ - P)^T(yQ - P)\right) = tr\left(Q^Ty^TyQ\right) - 2tr\left(yQP^T\right) + tr\left(P^TP\right) = \|y\|_F^2 - 2tr\left(yQP^T\right) + \|P\|_F^2. \tag{14}$$

Utilizing the SVD: $USV^T = P^Ty$, the solution of problem (13) is obtained via $Q^{k+1} = VU^T$.

(3) *Updating E:* Given x and Q, our objective can be reformulated as

$$E^{k+1} = \arg\min_E \left(\frac{1}{\mu}\|E\|_* + \frac{1}{2}\left\|E - \left(yQ - A(x) + \frac{1}{\mu}Z\right)\right\|_F^2\right). \tag{15}$$

Inspired by the literature [5], the optimal solution of problem (15) can be acquired by

$$E^{k+1} = UT_{\frac{1}{\mu}}[S]V, \tag{16}$$

where $(U, S, V^T) = svd(yQ\text{-}A(x) + Z/\mu)$.

The singular value shrinkage operation $T_\tau(\bullet)$ is formulated as

$$T_{\frac{1}{\mu}}[S] = diag\left(\left\{\max\left(0, s_{j,j} - \frac{1}{\mu}\right)\right\}_{1 \le j \le r}\right), \tag{17}$$

where variable r denotes the rank of matrix S.

The pseudo-code is summarized in Algorithm 1.

3.2. Recognition

Our main goal is to recognize the degraded low-resolution faces with pose variations. To recognize the input LR probe image y, we also need two datasets: the HR training set A_h and its LR version A_l. For the HR gallery dataset X_h, we can calculate its representation features over the HR training set A_h, denoted by $B = [B_1, B_2, \ldots, B_c]$, where $B_i = [b_{i,1}, b_{i,2}, \cdots, b_{i,n_i}] \in \Re^{N \times n_i}$, n_i denotes the sample number of the i th class in X_h, and c is the number of the class. For the input LR probe y, its LCSOPR

Algorithm 1 Solving Eq. (7) using the ADMM algorithm.

Input: Probe image $y \in \Re^{p \times q}$, a group of offline training images $A_1,...,A_N \in \Re^{p \times q}$, parameters λ and ε.
Initialize: $x = 0$, $E = 0$, $Q = 0$, $Z = 0$

1: Fix the others and set $C = (yQ^k - E^k + Z/\mu)ones(N,1)^T - H$, $G = C^TC$, $H = [Vec(A_1), Vec(A_2),..., Vec(A_N)]$.
•$x^{k+1} = (G + \tau D^2) \backslash ones(N, 1)$;
2: Fix the others and set $P = A(x^{k+1}) + E^k - Z^k/\mu$.
•By using the SVD: $USV^T = P^Ty$, $Q^{k+1} = VU^T$;
3: Fix the others and set $G = yQ^{k+1} - A(x^{k+1}) + Z^k/\mu$.
•$E^{k+1} = \arg\min_E (\frac{1}{\mu}\|E\|_* + \frac{1}{2}\|E - G\|_F^2)$;
4: Update the multipliers:
•$Z^{k+1} = Z^k + \mu(yQ^{k+1} - A(x^{k+1}) - E^{k+1})$;
5. Check for convergence:
•$\|yQ^{k+1} - A(x^{k+1}) - E^{k+1}\|_\infty > \varepsilon$,
Go to 1;

6. **Output:** Representation vector x^{k+1}.

representation feature over A_l can be represented as x_y. Thus, the combination coefficients of x_y over B can be calculated as follows:

$$\min_w \|x_y - Bw\|_2^2 + \eta\|w\|_1. \tag{18}$$

Here, η is the balance parameter. The homotopy [26,37] method can be used to solve Formula (18). The regularized reconstruction error for each class can be computed as

$$e_i(x_y) = \|x_y - B_i\delta_i(w^*)\|_2^2 / \|\delta_i(w^*)\|_2^2, \tag{19}$$

where function δ_i gathers the weights related to the i th class. Then, we assign the acquired LR probe image y to the class that receives the smallest representation error.

Algorithm 2 summarizes the pseudo-code in detail.

3.3. Complexity analysis

We discuss the computational complexity of our method in this section. Because the representation features of the gallery set can be learned offline, we only discuss the running time of the representation features' learning of the probe data and recognition stage. The major cost of Algorithm 1 results from performing these two tasks: (i) calculating the combination weights, and (ii) computing the SVD.

Algorithm 2 Cross-resolution face recognition via LCSOPR.

Input: LR probe y, gallery set X_h and training sets A_h, A_l.
1: Concerning each sample in X_h, obtain its representation feature over A_h by using Eq. (7) to form representation feature matrix B;
2: Concerning LR probe y, compute its representation feature x_y over A_l by using Eq. (7);
3: Recognize y based on Eq. (19);

Output: The label of LR probe y.

Following [16], it costs $O(pqN^3)$ to calculate the combination weights in step 1 of Algorithm 1. In steps 2 and 3, the cost of computing the SVD is $O(q^3)$ and $O(pq^2)$, respectively. Therefore, it costs $O(pqN^3 + q^3 + pq^2)$ at each iteration. By taking the maximum iteration times *maxIter* into consideration, the total time complexity for Algorithm 1 is $O(maxIter(pqN^3 + q^3 +$

Table 1
The NPR evaluations on the CMU PIE dataset.

Methods	LCSOPR	MLCSOPR
NPR	96%	98%

Table 2
The comparison results (%) of respective approaches on the CMU PIE dataset using pose C27 as the gallery set.

Methods	LINE[16]	SSR[18]	LcBR[23]	MDS[28]	DCA[12]	LR-GAN[2]	LCSOPR	MLCSOPR
C05	44.41	41.18	46.91	51.83	52.76	55.88	59.24	61.35
C29	45.03	38.09	45.58	48.25	49.16	51.52	54.97	56.24

Table 3
The comparison results (%) of respective approaches on the FERET dataset using frontal images as the gallery set.

Methods	LINE[16]	SSR[18]	LcBR[23]	MDS[28]	DCA[12]	LR-GAN[2]	LCSOPR	MLCSOPR
bd	28.00	29.50	28.50	35.50	36.50	38.50	42.00	43.50
bg	30.00	30.00	30.00	34.00	35.50	38.50	43.00	44.50

Table 4
The comparison results (%) of respective approaches on the CMU Multi-PIE dataset using the frontal images from pose 05_1 as the gallery set.

Methods	LINE[16]	SSR[18]	LcBR[23]	MDS[28]	DCA[12]	LR-GAN[2]	LCSOPR	MLCSOPR
13_0	20.25	19.00	20.25	30.50	31.50	34.00	38.50	41.55
14_0	53.75	52.25	54.25	57.25	57.75	60.75	63.75	65.25
05_0	44.00	42.75	44.00	47.25	47.75	49.75	52.75	54.25
04_1	17.50	17.50	17.75	29.25	30.50	33.25	37.75	40.75

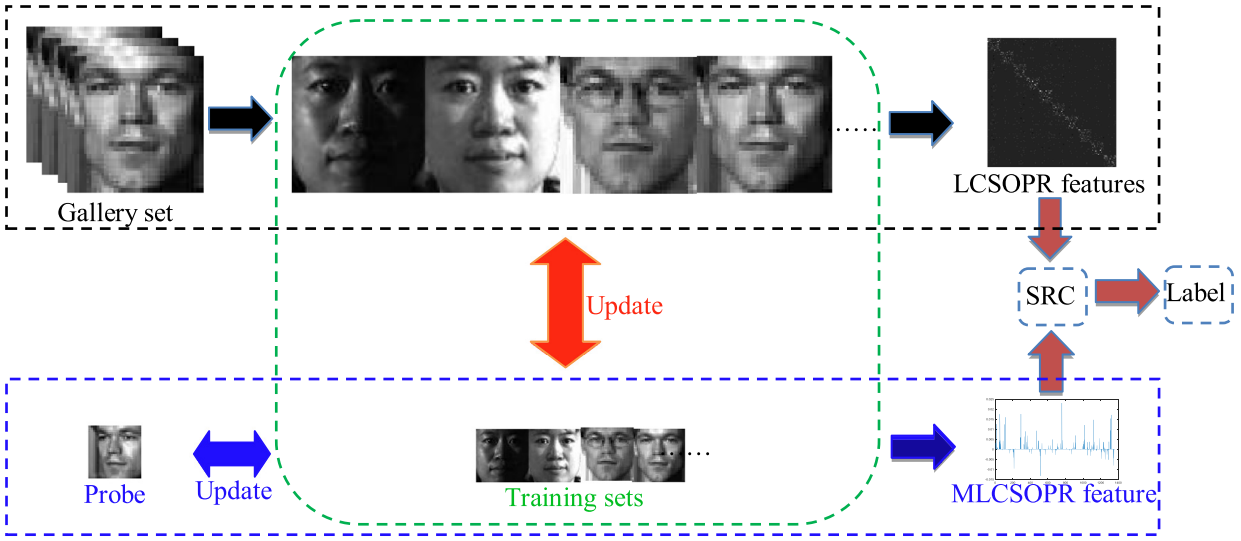

Fig. 1. The flow chart of the proposed MLCSOPR approach.

pq^2)). It should also be noted that the commonly used l_1-minimization solver, homotopy [26], has an empirical computational complexity of $O(N^2)$. Hence, the total time complexity of our method is about $O(maxIter(pqN^3 + q^3 + pq^2) + N^2)$.

3.4. Multilayer LCSOPR model

As discussed in Algorithm 2, the optimal pose-robust discriminative representation feature of LR input y over the LR training set A_l is obtained by minimizing the reconstruction error. However, there might be a discrepancy between the manifold geometries of the HR gallery image space and the LR one.

To handle this problem, the LCSOPR model is extended to its multilayer version (MLCSOPR), where the LR training dataset is updated continually. Thus, the goal of our MLCSOPR is formulated as:

$$\min_{x^b, Q^b, A_l^b} \left\| y^b Q^b - A_l^b(x^b) \right\|_* + \lambda \left\| d^b \otimes x^b \right\|_2^2, \quad s.t. \quad Q^T Q = I, \tag{20}$$

where $b = 0, 1, \ldots, B$ (B is the layer number). It should be noted that when $b = 0$, y^0 is the acquired LR probe and $A_l{}^0 = [A_{l1}, A_{l2}, \ldots, A_{lN}]$ is the given LR training dataset. Based on the iteratively updated LR training dataset, the discriminative representation feature of the LR probe can be learned in a much more consistent coupled space. Thus, compared with the single-layer LCSOPR, the performance of LCSOPR can be improved.

The iterative strategy is used to optimize the joint problem (20). Once the LR training set $A_l{}^b$ is fixed, the optimization problem reduces to Eq. (7). At each layer, by using a so-called *leave-one-out* trick, we desire to update the whole LR training dataset. For each LR training face $A_{lj}{}^b$ at the *b-th* layer, the updated LR training dataset is constructed with all other LR training images: $\hat{A}_l^b = \{A_{lm}{}^b \mid m = 1, 2, \ldots, j\text{-}1, j+1, \ldots, N\}$. Meanwhile, the corresponding new HR training dataset is $\hat{A}_h = \{A_{hm} \mid m = 1, 2, \ldots, j\text{-}1, j+1, \ldots, N\}$. The desired HR version A_{hi}^{*b} can be reconstructed (achieved by $\hat{A}_h(x^b)$) by $A_{lj}{}^b$, $\hat{A}_l^b$ and $\hat{A}_h$ using Algorithm 1. Fig. 1 illustrates the flow chart of the MLCSOPR method. We also tabulate the neighborhood preserving

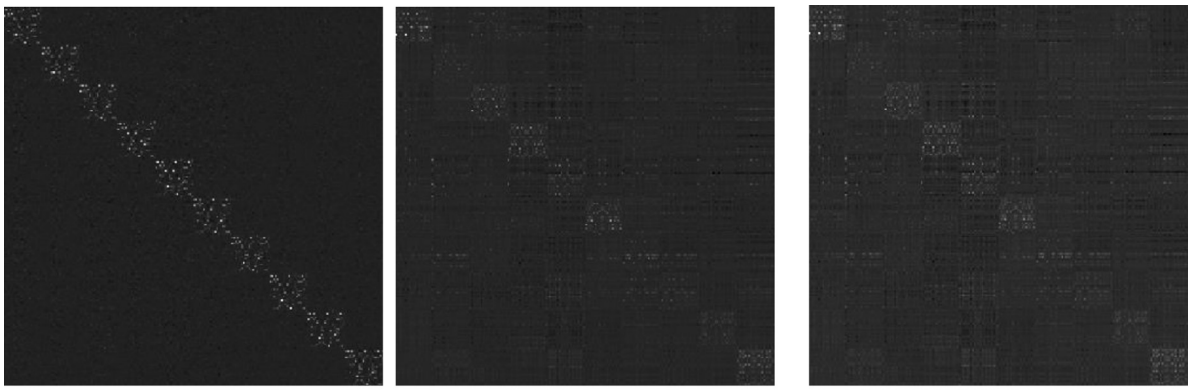

Fig. 2. The representation features obtained by LCSOPR and its two variations.

rate (NPR) [16] of the respective methods in Table 1, from which we can observe that MLCSOPR can preserve the coupled manifolds better.

3.5. Connections to related methods

It should be noted that LCSOPR aims at learning resolution-robust representation features in their respective spaces for pose-robust cross-resolution face recognition. The proposed LCSOPR method is different from the existing resolution-robust feature extraction based approaches in the following aspects:

(1) The MDS method [28] learns a common mapping matrix that simultaneously projects the facial features from both the LR and HR faces into a distance-preserving space. Nevertheless, because of the degradation from HR to LR manifold, the LR probe may have restricted the ability to distinguish its label. In contrast, we propose to learn pose-robust representation features in their respective spaces, and then match these resolution-robust features.
(2) Our proposed method inherits the merits of robust and expressive representation learning in face hallucination methods. In addition, the method transforms the cross-resolution face recognition problem into a resolution-robust feature learning one. However, our proposed LCSOPR method is quite different from the recently published low-rank representation and locality-constrained regression (LRRLCR) [8] method. When performing representation feature learning in the LR space, LRRLCR does not consider pose variations. In addition, we further propose a multilayer version of LCSOPR to make the manifold geometry of the LR image space more consistent with that of the HR one.

4. Further analysis of LCSOPR

In our proposed LCSOPR, the orthogonal Procrustes problem (OPP) is incorporated to alleviate the usual pose variations in the LR probe image. Furthermore, the proposed LCSOPR adopts the nuclear norm constraint to keep the structural property of the reconstruction error.

To further analyse the superiority of our proposed method, we list two variations here for comparison. In the first variation, we remove the orthogonal transformation matrix (i.e., $Q=I$) and the model is reformulated as

$$\min_{x} \|y - A(x)\|_* + \lambda \|d \otimes x\|_2^2. \tag{21}$$

This is a nuclear norm-based regression model, which can be optimized by the same procedure that was used in Algorithm 1. Problem (21) indicates that the probe image, which may have pose variations, is directly coded over the given training dataset to gain the representation features. In the second case, the Frobenius norm is adopted to regularize the reconstruction error and the model can be rewritten as

$$\min_{x,Q} \|yQ - A(x)\|_F^2 + \lambda \|d \otimes x\|_2^2, \quad s.t. \quad Q^T Q = I. \tag{22}$$

This variant degrades to the original orthogonal Procrustes problem and can be optimized by updating the orthogonal matrix Q and the representation feature x iteratively. For more details, please refer to [20,32].

For a better illustration, we give one example. We select some LR frontal images (down-sampled from their HR versions) from the CMU PIE database to linearly represent the LR probe images with pose variations from ten classes via LCSOPR and its two variations. The representation features obtained by LCSOPR and its two variations can be seen in Fig. 2, which shows that the features obtained by LCSOPR can reveal the data sample structures better: the between-class affinities are zeros, while the within-class affinities are dense. These compact and discriminative features are helpful for classification.

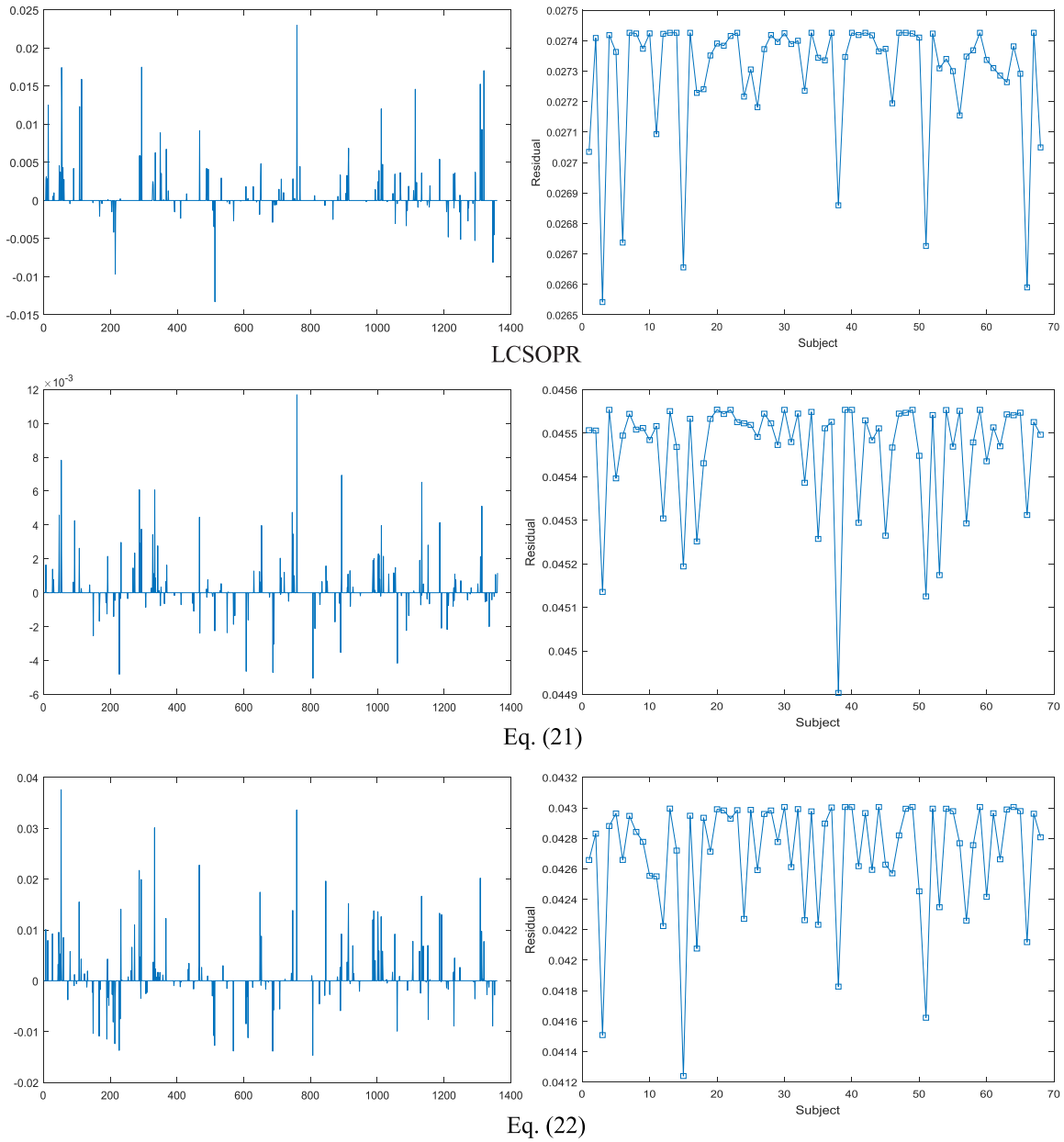

Fig. 3. The residuals of the LR test image from subject 3 obtained by LCSOPR and its two variations.

The corresponding sparse coefficients w and reconstruction residuals $e_i(x_y)$ of the LR test image from subject 3 computed by LCSOPR and its two variations can be seen in Fig. 3, which shows that LCSOPR can achieve the correct recognition result, while its two variations fail.

5. Experiments and discussions

In this section, we will assess the superiority of our method over some state-of-the-art methods by conducting experiments on three publicly available face datasets (CMU PIE, FERET, and CMU Multi-PIE). In the following test, the original HR faces are treated as the gallery, while the down-sampled HR face images are treated as the LR probes.

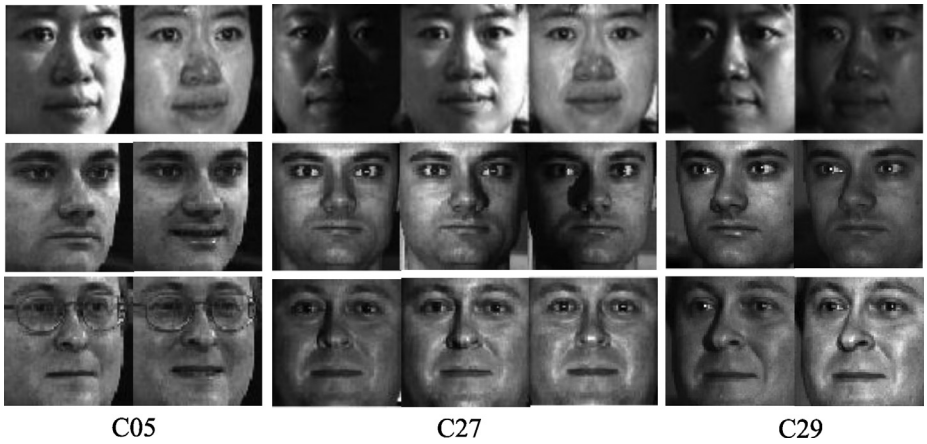

Fig. 4. Sample images of the CMU PIE dataset with various poses.

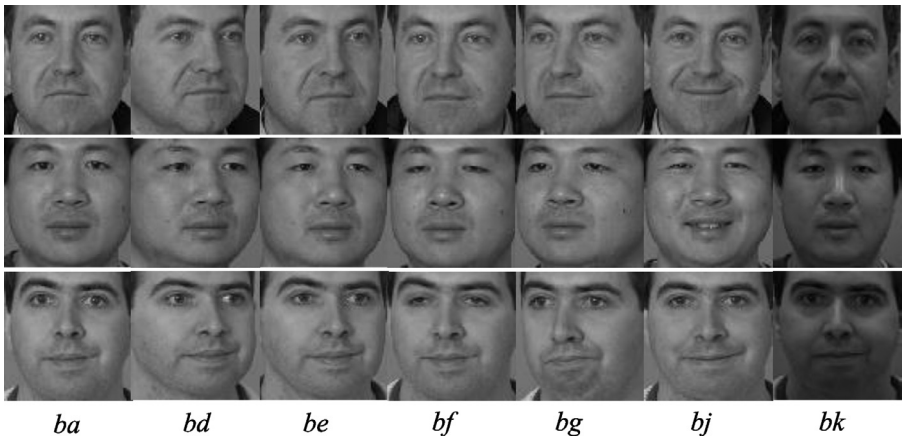

Fig. 5. Sample images of the FERET dataset with various poses.

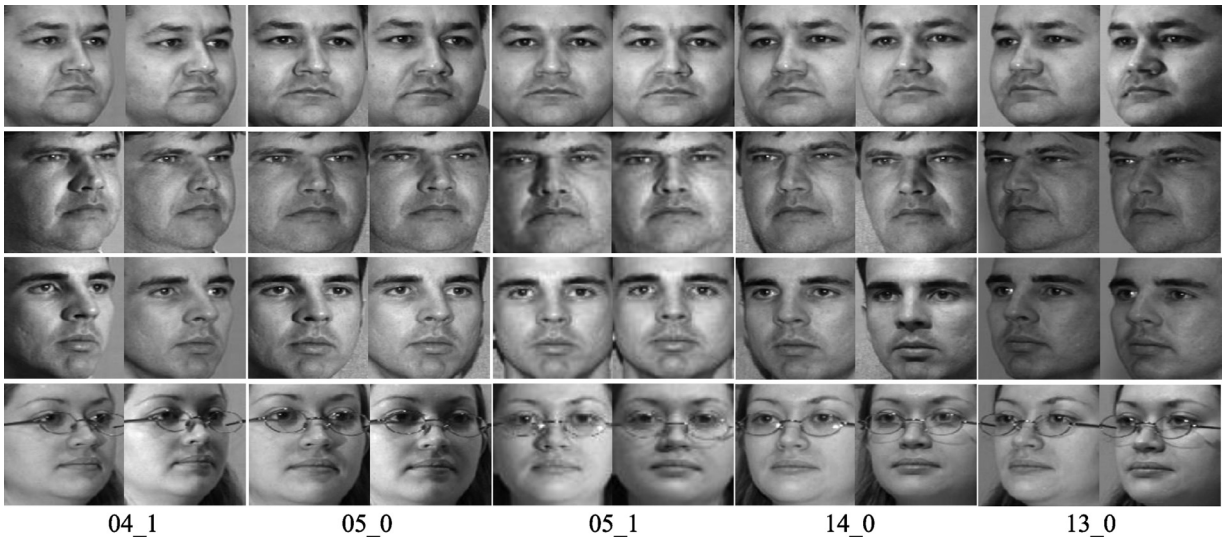

Fig. 6. Sample images of the CMU Multi-PIE dataset with various poses.

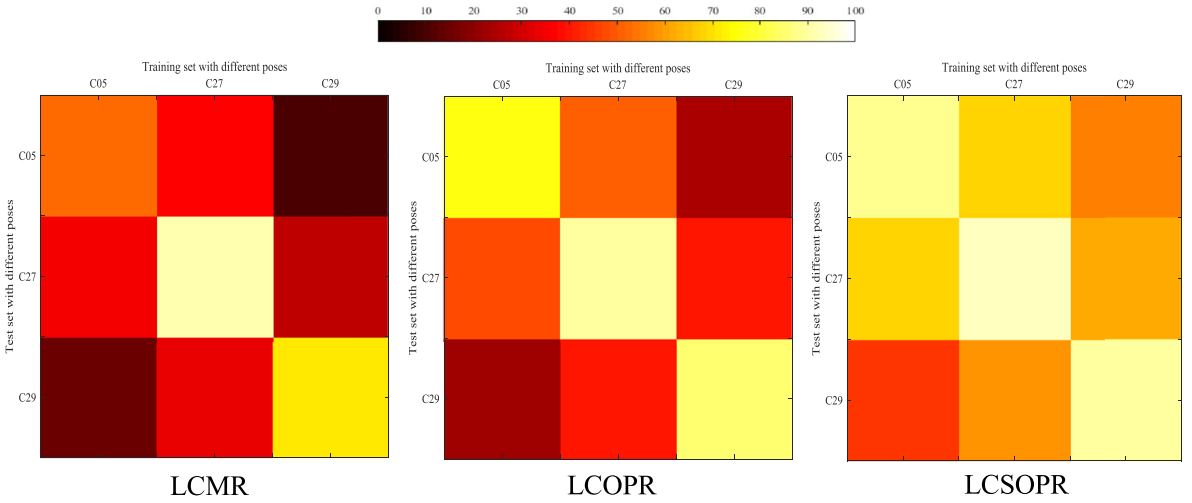

Fig. 7. The recognition matrices obtained by various methods on the CMU PIE dataset.

Fig. 8. Hallucination performance of respective approaches on the CMU PIE dataset. The first five columns denote the acquired LR images and the hallucinated results of BIC, LINE [16], SSR [18], and LcBR [23]. The last two columns are the original HR probe and corresponding frontal images for comparison.

Fig. 9. Hallucination performance of respective approaches on the FERET dataset. The first five columns denote the acquired LR images and the hallucinated results of BIC, LINE [16], SSR [18], and LcBR [23]. The last two columns are the original HR probe and corresponding frontal images for comparison.

5.1. Datasets

In the CMU PIE face dataset [3], samples from 68 subjects in 13 various poses are collected with yaw- and pitch-angle differences and expression and illumination variations. In this experiment, we select three different poses, C05, C27 and C29. Some sample images are shown in Fig. 4.

In the FERET face dataset [31], samples from 200 persons are collected. In this dataset, each subject was captured at nine viewing angles of 0, 60, 40, 25, 15, −15, −25, −40 and −60, which roughly correspond to nine view-points, *ba, bb, bc, bd, be, bf, bg, bh* and *bi*, respectively. Frontal images (denoted as *bk*), which have the same viewpoint as *ba*, are also collected using different lighting conditions. Fig. 5 shows some example images from this dataset.

In the CMU Multi-PIE face dataset [11], samples from 337 persons with expression, pose and illumination variations are collected in four sessions. In each session, for each person, there are 20 illuminations indicated from 0 to 19 per expression per pose. We choose five various poses {04_1, 05_0, 05_1, 14_0, 13_0} for this experiment. Some sample images from this dataset are depicted in Fig. 6.

5.2. Experiments using various poses as the gallery set

To investigate the robustness of the LCSOPR approach, we conduct tests on the CMU PIE dataset in case the samples in the gallery set also have pose variations. We select images of persons with neutral expressions in three different poses (C05, C27, and C29) as the gallery set. In addition, the probes set also has different poses. For each person, 20 images from pose C27 (frontal) are randomly gathered to form training sets A_h and A_l. The HR face samples have a size of 32×32, while their corresponding LR versions have a size of 8×8. The results of our model compared with those of models (21) and (22) (denoted as LCMR and LCOPR, respectively) are shown in Fig. 7. The bright block indicates a higher performance, while the dark block indicates a lower performance. Compared with LCMR and LCOPR, due to the ability of the OPP to handle re-

Fig. 10. Hallucination performance of respective approaches on the CMU Multi-PIE dataset. The first five columns denote the acquired LR images and the hallucinated results of BIC, LINE [16], SSR [18], and LcBR [23]. The last two columns are the original HR probe and corresponding frontal images.

flection and the ability of the nuclear norm constraint to reveal structure information, our LCSOPR is relatively less sensitive to whether the gallery has frontal images or not.

5.3. Comparison results

We compare our proposed approach with several state-of-the-art methods in this section. The compared cross-resolution face recognition methodologies are summarized into two varieties: (i) super-resolution based methods using super-resolved HR face images as the probe, including the bicubic interpolation (BIC) method, the locality-constrained iterative neighbor embedding (LINE) model [16], the smooth sparse representation (SSR) model [18] and the locality-constrained bi-layer representation (LcBR) model [23]; and (ii) typical resolution-robust approaches using just LR images as the probe, such as MDS [28], DCA [12] and LR-GAN [2]. It should be noted that in all experiments, we use the same training dataset for super-resolution based methods and the same gallery and probe dataset for resolution-robust feature extraction based methods.

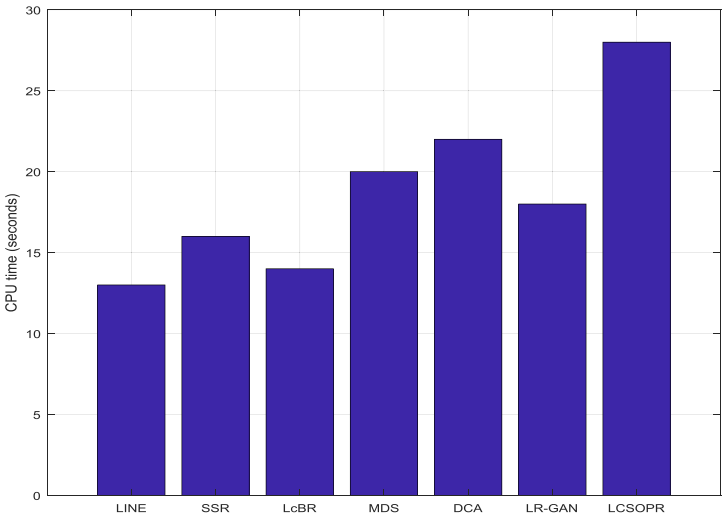

Fig. 11. The average CPU time of the respective methods on the CMU PIE face dataset.

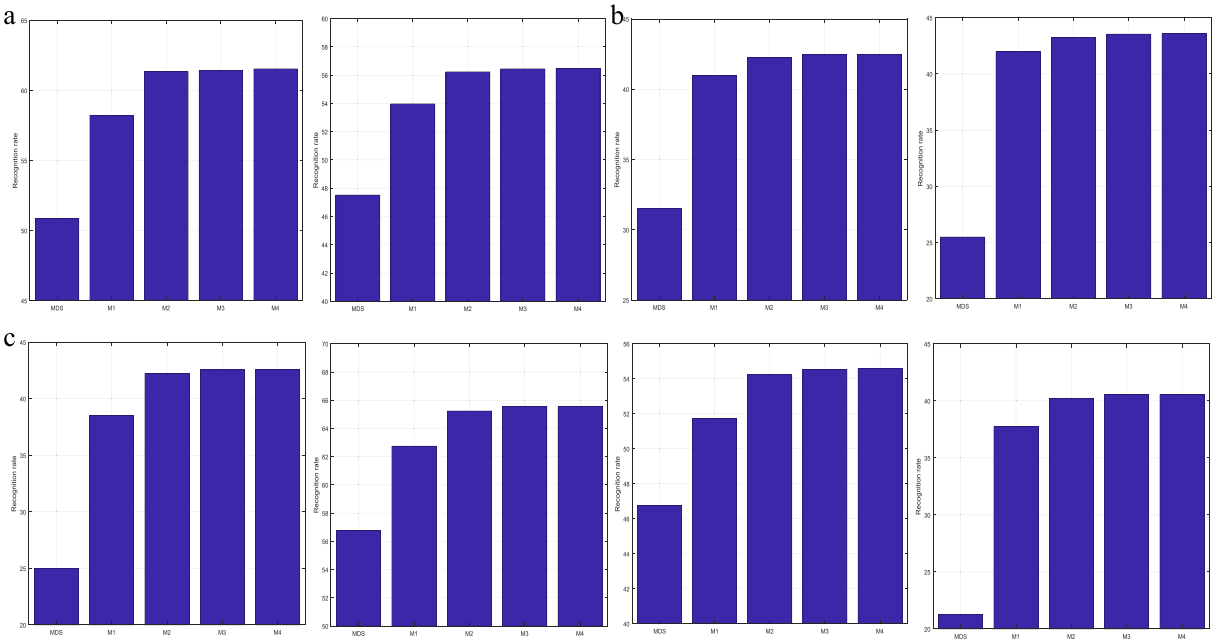

Fig. 12. The recognition results (%) of MLCSOPR on different datasets using different layer numbers. (a) CMU PIE dataset, (b) FERET dataset, and (c) CMU Multi-PIE dataset.

For a fair comparison, we use pose-robust orthogonal Procrustes regression-based classifiers [32] for super-resolution based methods.

For the CMU PIE dataset, we follow the experimental configuration where the probe poses are C05 and C29 while the gallery pose is C27 (frontal). Concerning pose C27, we randomly pick 20 samples from each subject to establish the training sets A_h and A_l, and another 10 samples to establish the HR gallery set X_h. Concerning pose C05 and C29, we randomly pick 10 samples from each subject to establish the LR probe set Y_l. All HR face images have a size of 32×32 and the down-sampled LR images have a size of 8×8 (by scaling factor of 4). Table 2 shows the comparison results of respective approaches.

For the FERET dataset, view-points *ba, bd, be, bf, bg, bj*, and *bk* are selected in our experiment. All the images are manually cropped and resized to 32×32. For each person, we gather one near-frontal face image from view-point *ba* to construct the HR gallery set X_h. For view-point *bd* or *bg*, we gather one non-frontal face image to construct the LR probe set Y_l. We

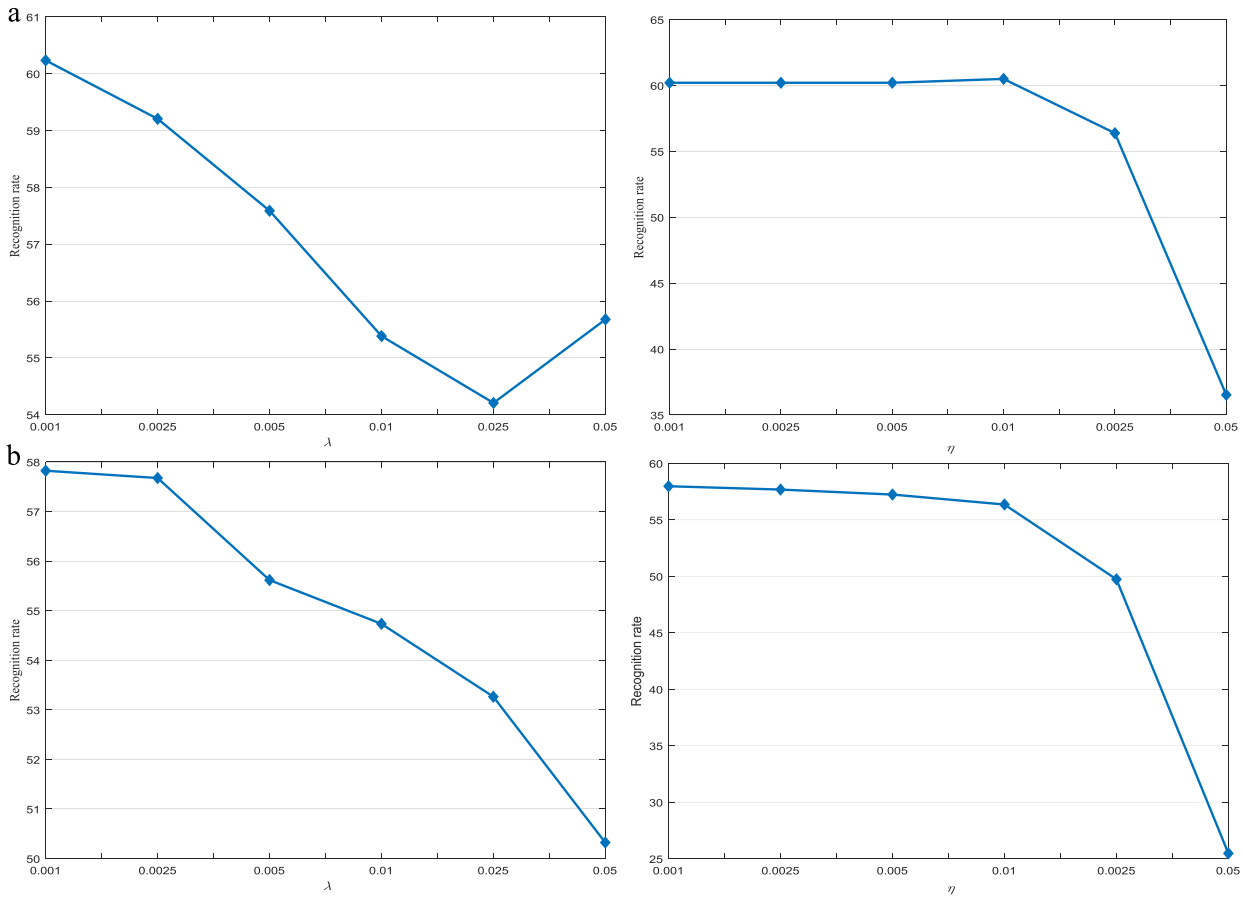

Fig. 13. The recognition results (%) of MLCSOPR with different parameter configurations on the CMU PIE dataset based on different input poses. (a) Pose C05, and (b) pose C29.

use the remainders to construct the training sets A_h and A_l. The LR images have size 8×8. The results of the respective approaches are listed in Table 3.

For the CMU Multi-PIE face dataset, we select 80 subjects in this experiment. For each subject, 15 frontal images from pose 05_1 are chosen to form the training sets A_h and A_l, and another 5 images are chosen to construct the gallery set X_h. For poses {04_1, 05_0, 14_0, 13_0}, we choose 5 samples of each subject to form the LR probe Y_l. The HR faces have a size of 32×32, and the down-sampled LR faces have a size of 8×8 (by a scaling factor of 4). The recognition results of respective approaches are tabulated in Table 4.

From the above results, we can make the following observations: (1) The recognition results of hallucination-based approaches are unsatisfactory. Because the observed LR images have pose variations, the vital representation coefficients cannot be learned well in conventional learning-based SR approaches. Thus, the detailed distinctive features are lost in the hallucinated faces. (2) For a better illustration, some SR results are given in Figs. 8–10. As seen in Figs. 8–10, some "ghosting" artefacts have appeared around the mouth, eye and face contours. Furthermore, these hallucinated faces seem to have a non-frontal pose, to some extent, which leads additional challenges for the subsequent recognition stage. (3) Conventional resolution-robust cross-resolution face recognition technologies (e.g., MDS, DCA, and LR-GAN) have better performance than hallucination-based ones. The introduction of discriminative feature extraction might contribute to this observation. (4) Compared with the resolution-robust feature-based methods and hallucination-based methods, our proposed approaches significantly improve performance. These achievements demonstrate that by incorporating the OPP and nuclear norm regularization, our methods can learn discriminative resolution-robust representation features to improve performance.

5.4. Running time

We compare the time cost of each approach in this subsection. The experiments were implemented with this hardware configuration: Intel Core i7-6700 CPU @ 3.4 GHz, 24 GBytes RAM. For the demonstration, we only report the results on the CMU PIE face dataset. The average running time of our method and other methods are listed in Fig. 11. Due to the iterative

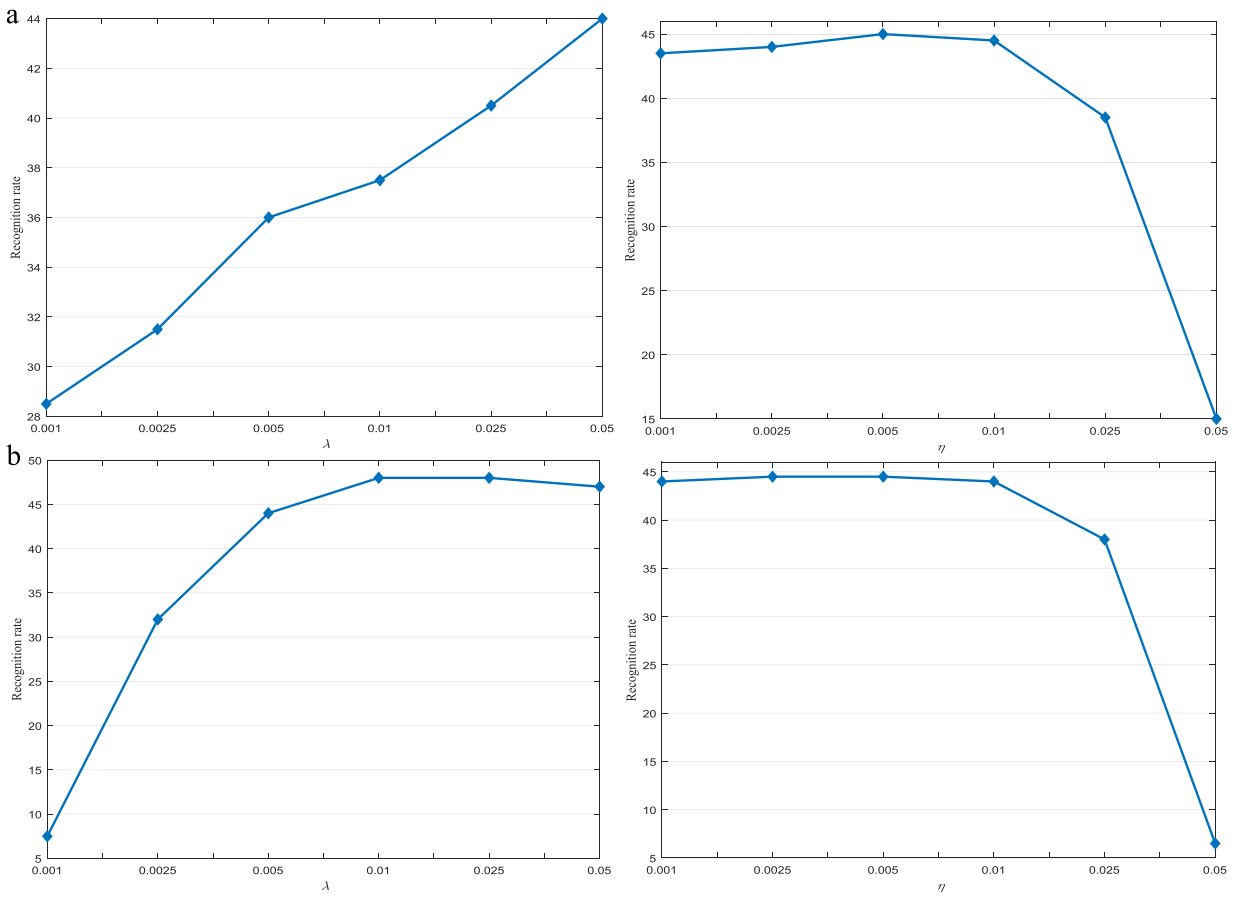

Fig. 14. The recognition results (%) of MLCSOPR with different parameters on the FERET dataset based on different input poses. (a) View-point *bd* and (b) view-point *bg*.

scheme in the feature learning of LCSOPR, our method requires more time than the other methods. LINE, SSR and LcBR are faster because they only require a few operations for matrix multiplications and additions. Suffering from the feature construction operation, MDS and DCA spend much time handling one probe image.

5.5. Parameter discussion

We mainly discuss the effect of layer number B, parameter λ and η on the performance of our approach in this section. The experiments are also conducted on three standard face datasets (CMU PIE, FERET and CMU Multi-PIE), and the same experimental configurations are set as the above experiments. With the help of the training set, we utilize the *leave-one-out* strategy to tune the parameters: for each subject, one image is selected as the probe, another image is selected as the gallery, and the rest are used as the dictionary.

(1) The influence of layer number

Fig. 12 depicts the recognition results using different layer numbers. We take the MDS [28] method as the baseline for comparison. The single-layer LCSOPR method is denoted as “M1”, the two-layer LCSOPR method is denoted as “M2”, and so on. As the layer number increases, the superiority of the presented method over SLR becomes remarkable. Additionally, the improvements of MLCSOPR over single-layer LCSOPR are distinct. It should be noted that MLCSOPR will achieve a relatively stable state when layer number B is larger than 3. We set $B=3$ in all experiments.

(2) The influence of parameter λ and η

Fig. 13 depicts the recognition results of MLCSOPR with different parameter configurations on the CMU PIE face dataset using different poses as input. Fig. 13(b) shows that the performance of MLCSOPR degrades as the value of parameters increases. However, MLCSOPR can achieve the best results when η is 0.01 for pose C05.

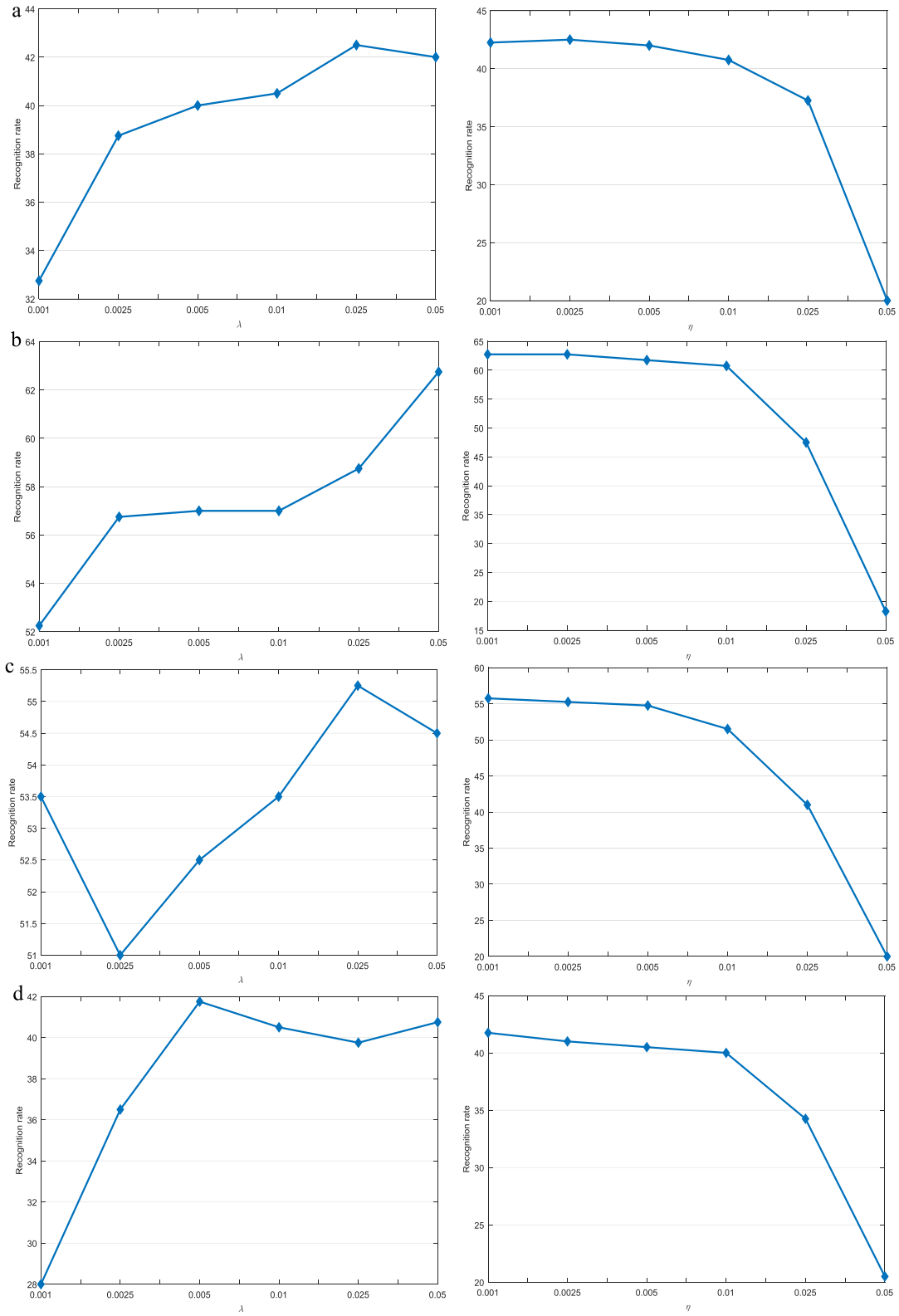

Fig. 15. The recognition results (%) of MLCSOPR with different parameters on the CMU Multi-PIE dataset based on different input poses. (a) Pose 13_0, (b) pose 14_0, (c) pose 05_0 and (d) pose 04_1.

Fig. 14 lists the recognition results of MLCSOPR with different parameter configurations on the FERET dataset using different view-points as input. Fig. 14(a) shows that the performance of MLCSOPR increases as the values of parameter λ increases, and the best performance can be obtained when η is 0.005. In addition, MLCSOPR obtains the best result when λ is 0.01 and η is 0.0025 when using view-point *bg* as input.

The recognition results of MLCSOPR with different parameter configurations on the CMU Multi-PIE face dataset using different poses as input is plotted in Fig. 15, which shows that MLCSOPR obtains better performance when η is lower than 0.005 and obtains the best results when λ is 0.025 for pose 13_0 (or pose 05_0) and 0.005 for pose 04_1. However, MLCSOPR achieves the worse performance when λ is lower than 0.01 for pose 14_0.

6. Conclusions

In this paper, for cross-resolution face matching with pose variations, we presented an approach named multilayer locality-constrained structural orthogonal Procrustes regression (MLCSOPR). The proposed MLCSOPR not only learns the pose-robust discriminative representation features to reduce the resolution gap between the LR image space and the HR one but also strengthens the consistency between the LR and HR image space. Experiments conducted on CMU PIE, FERET and CMU Multi-PIE face datasets verified the superiority of our proposed approaches.

Our current work only discusses the horizontal pose variations. However, in real-world applications, pose variations exist in both the horizontal and vertical directions. Thus, we can modify our model to address multiple pose variations in further work. Additionally, we will combine our learning strategy with deep neural networks to further advance cross-resolution face recognition performance. Furthermore, we will investigate how to extend our MLCSOPR to handle uncontrolled face images.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant nos. 61502245, 61503195, 61772568 and 61806098; the Guangdong Medical Scientific and Technological Research Funding under Grant no. A2017251; the Six Talent Peaks Project in Jiangsu Province under Grant no. RJFW-011; the Natural Science Foundation of Jiangsu Province under Grant no. BK20150849; and the Open Fund Project of Provincial Key Laboratory for Computer Information Processing Technology (Soochow University) (no. KJS1840).

References

[1] H.S. Bhatt, R. Singh, M. Vatsa, N.K. Ratha, Improving cross-resolution face matching using ensemble-based co-transfer learning, IEEE Trans. Image Process. 23 (2014) 5654–5669 Oct.
[2] S. Banerjee, S. Das, LR-GAN for degraded face recognition, Pattern Recognit. Lett. 116 (2018) 246–253 Dec..
[3] S. Baker, T. Sim, M. Bsat, The CMU pose, illumination, and expression database, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 1615–1618 Dec.
[4] S. Baker, T. Kanade, Limits on super-resolution and how to break them, IEEE Trans. Pattern Ana.l Mach. Intel.l 24 (2002) 1167–1183 Sep.
[5] J.F. Cai, E.J. Candes, Z.W. Shen, A singular value thresholding algorithm for matrix completion, Siam J. Optim. 20 (2010) 1956–1982.
[6] Z. Fan, D. Zhang, X. Wang, Q. Zhu, Y. Wang, Virtual dictionary based kernel sparse representation for face recognition, Pattern Recognit. 76 (2018) 1–13 April.
[7] G. Gao, J. Yang, X.Y. Jing, F. Shen, W. Yang, D. Yue, Learning robust and discriminative low-rank representations for face recognition with occlusion, Pattern Recognit. 66 (2017) 129–143 Jun.
[8] G. Gao, Z. Hu, P. Huang, M. Yang, Q. Zhou, S. Wu, D. Yue, Robust low-resolution face recognition via low-rank representation and locality-constrained regression, Comput. Electr. Eng. 70 (2018) 968–977 Aug.
[9] G. Gao, P. Huang, D. Yue, W. Yang, Locality-constrained structral orthogonal procrustes regression for low-resolution face recognition with pose variations, in: Proceedings of IAPR Asian Conference on Pattern Recognition, 2017, pp. 554–558.
[10] B.F. Green, The orthogonal approximation of an oblique structure in factor analysis, Psychometrika 17 (1952) 429–440.
[11] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, Image Vis. Comput. 28 (2010) 807–813 May.
[12] M. Haghighat, M. Abdel-Mottaleb, Low resolution face recognition in surveillance systems using discriminant correlation analysis, in: Proceedings of IEEE International Conference on Automatic Face & Gesture Recognition, 2017, pp. 912–917.
[13] J.R. Hurley, R.B. Cattell, The procrustes program: producing direct rotation to test a hypothesized factor structure, Syst. Res. Behav. Sci. 7 (1962) 258–262.
[14] M. Jian, K.M. Lam, J. Dong, A novel face-hallucination scheme based on singular value decomposition, Pattern Recognit. 46 (2013) 3091–3102.
[15] C. Jung, L. Jiao, B. Liu, M. Gong, Position-patch based face hallucination using convex optimization, IEEE Signal Process. Lett. 18 (2011) 367–370 Jun.
[16] 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 (Oct 2014) 4220–4231.
[17] J. Junjun, Y. Yi, T. Suhua, M. Jiayi, Q. Guo-Jun, A. Akiko, Context-patch based face hallucination via thresholding locality-constrained representation and reproducing learning, IEEE Trans. Cybern. (2019), doi:10.1109/TCYB.2018.2868891.
[18] J. Jiang, J. Ma, C. Chen, X. Jiang, Z. Wang, Noise robust face image super-resolution through smooth sparse representation, IEEE Trans. Cybern. 47 (2017) 3991–4002 Nov..
[19] J. Jiang, R. Hu, Z. Wang, Z. Cai, CDMMA: coupled discriminant multi-manifold analysis for matching low-resolution face images, Signal Process. 124 (2016) 162–172 Jul.
[20] J. Jiang, R. Hu, Z. Wang, Z. Han, Noise robust face hallucination via locality-constrained representation, IEEE Trans. Multimed. 16 (2014) 1268–1281 Aug.

[21] Y. Li, C. Cai, G. Qiu, K.-.M. Lam, Face hallucination based on sparse local-pixel structure, Pattern Recognit. 47 (2014) 1261–1270.
[22] Y. Liang, J.-.H. Lai, P.C. Yuen, W.W. Zou, Z. Cai, Face hallucination with imprecise-alignment using iterative sparse representation, Pattern Recognit. 47 (2014) 3327–3342 Oct.
[23] L. Liu, C.P. Chen, S. Li, Y.Y. Tang, L. Chen, Robust face hallucination via locality-constrained bi-layer representation, IEEE Trans. Cybern. 48 (2018) 1189–1201 Apr..
[24] T. Lu, W. Yang, Y. Zhang, X. Li, Z. Xiong, Very low-resolution face recognition via semi-coupled locality-constrained representation, in: IEEE International Conference on Parallel and Distributed Systems (ICPADS), 2016, pp. 362–367.
[25] X. Ma, J. Zhang, C. Qi, Hallucinating face by position-patch, Pattern Recognit. 43 (2010) 2224–2236 Jun.
[26] D.M. Malioutov, M. Cetin, A.S. Willsky, Homotopy continuation for sparse signal representation, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005, pp. 733–736.
[27] P. Moutafis, I.A. Kakadiaris, Semi-coupled basis and distance metric learning for cross-domain matching: application to low-resolution face recognition, in: IEEE International Joint Conference on Biometrics (IJCB), 2014, pp. 1–8.
[28] S.P. Mudunuri, S. Biswas, Low resolution face recognition across variations in pose and illumination, IEEE Trans. Pattern. Anal. Mach. Intell. 38 (2016) 1034–1040 Aug.
[29] C. Peng, X. Gao, N. Wang, J. Li, Graphical representation for heterogeneous face recognition, IEEE Trans. Pattern. Anal. Mach. Intell. 39 (2017) 301–312 Feb.
[30] C. Peng, N. Wang, J. Li, X. Gao, Re-ranking high-dimensional deep local representation for NIR-VIS face recognition, IEEE Trans. Image Process. (2019), doi:10.1109/TIP.2019.2912360.
[31] P.J. Phillips, H. Wechsler, J. Huang, P.J. Rauss, The FERET database and evaluation procedure for face-recognition algorithms, Image Vis. Comput. 16 (1998) 295–306 Apr..
[32] Y. Tai, J. Yang, Y. Zhang, L. Luo, J. Qian, Y. Chen, Face recognition with pose variations and misalignment via orthogonal procrustes regression, IEEE Trans. Image Process. 25 (2016) 2673–2683 Jun.
[33] T. Viklands, Algorithms for the Weighted Orthogonal Procrustes Problem and Other Least Squares Problems, Datavetenskap, 2006.
[34] N. Wang, X. Gao, L. Sun, J. Li, Bayesian face sketch synthesis, IEEE Trans. Image Process. 26 (2017) 1264–1274 Mar.
[35] N. Wang, X. Gao, J. Li, Random sampling for fast face sketch synthesis, Pattern Recognit. 76 (Apr 2018) 215–227.
[36] Z. Wang, Z. Miao, Y. Wan, Z. Tang, Kernel coupled cross-regression for low-resolution face recognition, Math. Probl. Eng. 2013 (2013).
[37] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 210–227 Feb.
[38] X. Yin, X. Liu, Multi-task convolutional neural network for pose-invariant face recognition, IEEE Trans. Image Process. 27 (2018) 964–975 Feb.
[39] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, Y. Xu, Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 156–171 Jan.
[40] M.C. Yang, C.P. Wei, Y.R. Yeh, Y.C.F. Wang, Recognition at a long distance: very low resolution face recognition and hallucination, in: International Conference on Biometrics, 2015, pp. 237–242.