



Face image super-resolution with pose via nuclear norm regularized structural orthogonal Procrustes regression

Guangwei Gao^{1,5} · Dong Zhu¹ · Meng Yang² · Huimin Lu³ · Wankou Yang⁴ · Hao Gao¹

Received: 1 August 2018 / Accepted: 15 October 2018 / Published online: 25 October 2018
© The Natural Computing Applications Forum 2018

Abstract

In real applications, the observed low-resolution face images usually have pose variations. Conventional learning-based methods ignore these variations; thus, the hallucinated high-resolution faces are not reasonable for the following recognition task. For recognition purpose, we prefer to obtain near-frontal faces. To this end, we propose a nuclear norm regularized structural orthogonal Procrustes regression (N2SOPR) approach in this work to acquire pose-robust feature representations for face hallucination with pose. The orthogonal Procrustes regression is used to seek an appropriate transformation between two data matrixes. Additionally, the nuclear norm regularization is imposed on the representation residual to preserve image structural property. We also impose a low-rank restraint on the combination weight to automatically cluster each input into the same subspace with the training samples. Both hallucination and recognition experiments conducted on common face databases have verified that our N2SOPR can obtain reasonable performance than some related methods.

Keywords Face hallucination · Pose variations · Nuclear norm · Low-rank constraint

1 Introduction

In the past few decades, great achievements have been made in the community of face recognition [1–16] with the contributions of researchers. However, due to the limitations of network bandwidth, hardware storage, long distance between the electric imaging system and the interest

object, the observed interested face regions usually have low resolution (LR). Due to the low-quality property, the discriminative details extracted from these LR face regions are so limited that the following face recognition performance is unsatisfactory [17]. To provide more facial details for the subsequent recognition procedure, face hallucination technologies have been applied to forecast target high-resolution (HR) faces from captured LR ones. In the past several decades, learning-based approaches have attracted extensive attention for its remarkable achievements in face image super-resolution tasks.

Freeman et al. [18] firstly developed a learning-based face hallucination approach, where the potential correlation between HR and corresponding LR patches is formulated by a Markov network. Wang et al. [19] utilized principal component analysis (PCA) method to code the input as a linear representation over the training images and then reconstructed the desired HR version by applying the same combination coefficients on their HR counterparts. Hu et al. [20] then presented a kernel extension of the PCA-based holistic model by taking into account the higher-order image statistic information. Shi et al. [21] proposed a two-phase face hallucination framework. They first

✉ Guangwei Gao
csgwgao@njupt.edu.cn

¹ Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, People's Republic of China

² School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, People's Republic of China

³ Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Kitakyushu, Japan

⁴ School of Automation, Southeast University, Nanjing, People's Republic of China

⁵ Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing, People's Republic of China

produced a temporary HR version in patch-wise manner and then integrated local sparsity, global consistency, and pixel correlation to hallucinate the final desired HR face image. In order to search a consistency subspace to maximize the relation between the PCA coefficients of LR images and corresponding HR ones, canonical correlation analysis (CCA) is utilized by Huang [22] and An [23]. Gao et al. [24] designed an efficient face hallucination framework, which estimates a mapping between the embedding geometrics in each image spaces via sparse representation [25–28].

These previous global face hallucination models loss some fine individual details of the input face images sometimes. To tackle this problem, local patch-based strategies have been presented to render more facial details. Chang et al. [29] assumed that the local geometric structure of LR image patches is similar to that of HR ones and then proposed a neighbor embedding (NE) scheme to pick the nearest neighbors from the training set. After that, numerous related works have been proposed, such as iterative neighbor embedding [30] and coupled-layer neighbor embedding [31]. Unlike those patch-based methods utilizing all related neighbors for hallucination, Yang et al. [32] proposed to minimize the reconstruction error by using sparse representation technique to automatically choose the most relevant neighbors. Based on an observation that human faces contain highly structured property, many works incorporated the prior position information into super-resolution reconstruction procedure. Ma et al. [33] formulated the face super-resolution process as a constrained least square problem and presented a position-patch-based method. One problem is that the solution of the least square estimation may not be stable if the size of a patch is smaller than that of the referenced set. To handle this issue, Jung et al. [34] adopted sparse regularization term to penalize the coefficients. After that, Wang et al. [35] further designed a weighted adaptive sparse regularization model to improve the performance. To make full use of both sparsity and locality simultaneously, Jiang et al. [36] proposed a locality-constrained representation model. In consideration that close patches naturally have analogical representations, Jiang et al. [37] presented a robust face super-resolution method based on smooth sparse representation. With the help of a group of training pairs, a robust bilayer representation scheme is presented by Liu et al. [38] to perform hallucination and denoising simultaneously. Recently, as a highly effective model, deep learning has achieved remarkable achievements in various visual applications [39–41]. For example, Dong et al. [39] presented to train an end-to-end image super-resolution model by using deep convolutional network.

These above methods have achieved satisfactory hallucination results on the frontal view face images. However,

in real-world surveillance applications, the captured LR face images usually have pose variations. When learning the LR patch representations, the aforementioned algorithms ignore the pose variations; thus, the learned representations are sensitive to these variations. We prefer to hallucinate near-frontal faces for the following recognition procedure. To handle face hallucination with pose variations, we present a nuclear norm regularized structural orthogonal Procrustes regression (N2SOPR) method. Orthogonal Procrustes regression is applied to seek an appropriate transformation between two data matrixes to make the pose of one data adaptive to the other, and imposes the structural constraint on the representation residual to preserve image inherent property. Also, the proposed method tries to cluster the input into the same subspace with the training samples by low-rank constraint. Locality constraint is also incorporated to enable the recovery of local manifold from local patches. Experimental results in terms of hallucination and recognition have shown the efficiency of our method.

We organize the remainder of the paper as follows: In Sect. 2, several relevant position-patch-based face super-resolution approaches are simply introduced. In Sect. 3, we detail our nuclear norm regularized matrix structural orthogonal Procrustes regression-based method, including the optimization details and complexity analysis. Experiments and discussions are shown in Sect. 4. The conclusion and future work are provided in Sect. 5.

2 Related work

In learning-based face hallucination methods, each input image patch extracted from its global version is represented by a given dictionary in the same position with the training samples. Different schemes convert the input into a representation vector to gain the desired HR patch. In this part, we will briefly review three popular face hallucination approaches related to our proposed one.

Denoting A^m ($m = 1, \dots, N$) the training samples (N is the number of the training images) and y the test image, the input testing sample and each training sample are decomposed into some overlapping square patches and denoted as $\{y(i, j) | 1 \leq i \leq R, 1 \leq j \leq C\}$, $\{A^m(i, j) | 1 \leq i \leq R, 1 \leq j \leq C\}$, respectively, where term (i, j) is the position index, C and R represent the patch numbers in every column and row.

2.1 Least square representation

An efficient yet simple least square representation model [33] is presented to collaboratively code each input patch as a linear representation over the same position patches

extracted from all training samples. The optimal representation vector associated with the acquired input $y(i, j)$ can be computed by:

$$x^*(i, j) = \arg \min_{x(i, j)} \left\| y(i, j) - \sum_{m=1}^N A^m(i, j)x_m(i, j) \right\|_2^2 \tag{1}$$

s.t. $\sum_{m=1}^N x_m(i, j) = 1.$

By considering a Gram matrix, the analytical solution of above least square problem can be obtained.

2.2 Sparse representation

One problem of Eq. (1) may be that the solution of such least square estimation may not be stable. By imposing sparse constraint on the combination vector, Jung et al. [34] proposed to adaptively select the most relevant training patches to represent input patch. It converts Eq. (2) to an l_1 -norm minimization problem:

$$\min_x \|x(i, j)\|_1 \quad \text{s.t.} \quad \left\| y(i, j) - \sum_{m=1}^N A^m(i, j)x_m(i, j) \right\|_2^2 \leq \epsilon. \tag{2}$$

Here, l_1 -norm adds the absolute value of a vector. Many convex optimization algorithms could be used to solve Eq. (2), such as l_1 -ls [42]. It should be noted that the learned representation can capture salient properties of the training images due to the sparsity constraint.

2.3 Locality-constrained representation

Aforementioned sparse representation method [34] neglects the locality characteristic of the representation vector. In contrast, locality-constrained representation [36] incorporates a manifold regularization into representation vector to maintain the intrinsic geometry of training samples. The objective is defined as follows:

$$\min_{x(i, j)} \left\| y(i, j) - \sum_{m=1}^N A^m(i, j)x_m(i, j) \right\|_2^2 + \lambda \sum_{m=1}^N [d_m(i, j)x_m(i, j)]^2$$

s.t. $\sum_{m=1}^N x_m(i, j) = 1,$

(3)

where λ denotes regularization parameter; each $d_m(i, j)$ describes the distance between the m th atom $A^m(i, j)$ and input patch $y(i, j)$. The analytical solution of problem (3) can be simply gained by solving a regularized least square problem.

2.4 Orthogonal Procrustes problem

We briefly review the orthogonal Procrustes problem (OPP) in this subsection. OPP stems from factor analysis in psychometrics [43]. OPP aims at finding an appropriate transformation which corrects a data matrix X to match another objective data matrix B .

Generally, denoting the orthogonal matrix as Q and the two-dimensional data matrices as $X, B \in \mathbb{R}^{p \times q}$, the orthogonal Procrustes problem can be denoted as:

$$\min_Q \|XQ - B\|_F^2, \quad \text{s.t.} \quad Q^T Q = I_q. \tag{4}$$

Here, identity matrix I_q has size $q \times q$ and $\|\cdot\|_F$ represents the Frobenius norm. The analytical solution of problem (4) could be gained by utilizing the singular value decomposition (SVD) operation. Let $USV^T = B^T X$; the optimal solution related to problem (4) is $\hat{Q} = VU^T$. Furthermore, we can also multiply the orthogonal transformation matrix Q in the left side of matrix X , which can be written by:

$$\min_Q \|QX - B\|_F^2, \quad \text{s.t.} \quad Q^T Q = I_p, \tag{5}$$

where I_p denotes a $p \times p$ identity matrix. By using the SVD operation: $USV^T = XB^T$, the analytical solution of problem (2) is given by $\hat{Q} = VU^T$. As discussed in [7], problem (1) handles the horizontal direction variations well, while problem (2) handles the vertical direction variations well. Without loss of generality, we mainly deal with the horizontal direction variations in this paper.

3 The proposed N2SOPR

3.1 Problem formulation

The patches of well-aligned two-dimensional training images can be denoted as $A_i \in \mathbb{R}^{p \times q}$ ($i = 1, \dots, N$), where N is the number of samples. Then, the linear mapping can be denoted as

$$A(x) = \sum_{i=1}^N x_i A_i, \tag{6}$$

x_i denotes the weight coefficient here. Generally, if there comes a frontal probe face image patch $y \in \mathbb{R}^{p \times q}$, it could be linearly represented as: $y = A(x) + E$; here matrix $E \in \mathbb{R}^{p \times q}$ represents the error. Sometimes, the pose between the training images and test image may be different. In this case, test image can be modified by yQ , where $Q \in \mathbb{R}^{q \times q}$ represents an orthogonal transformation matrix, aiming to make the pose of the input to adapt that

of the training ones. The modified test image yQ can be linearly approximated as follows:

$$yQ = \sum_{i=1}^N x_i A_i + E. \tag{7}$$

Researchers in [9, 10] have revealed that, to depict the potential structure noise, nuclear norm regularization is a better choice. To benefit from this observation, the nuclear norm regularization is also imposed on our representation residual to preserve image structure property. Then, we can formulate the orthogonal Procrustes regression as follows:

$$\min_{x,Q} \|yQ - A(x)\|_*, \quad \text{s.t. } Q^T Q = I, \tag{8}$$

where symbol $\|\cdot\|_*$ denotes the sum of the singular value of a data matrix. The locality constraint is also incorporated into the representation via a metric between each training sample and the input image to reveal the prior information from nearest neighbors. Our structural orthogonal Procrustes regression model can be formulated as:

$$\min_{x,Q} \|yQ - A(x)\|_* + \eta \|d \otimes x\|_2^2, \quad \text{s.t. } Q^T Q = I, \tag{9}$$

where $d = (d_1, \dots, d_N)^T$ is a distance vector, \otimes is the element-wise product, η is used to balance the contribution of the locality constraint, and $d_i = \|y - A_i\|_F^2$ denotes the locality criterion to measure the distance between each training atom and the input. For better reconstruction, we also impose low-rank restraint on the weights to cluster the input into the same subspace with the training samples. Finally, we formulated our problem as follows:

$$\min_{x,Q} \|yQ - A(x)\|_* + \lambda \|H \text{diag}(x)\|_* + \eta \|d \otimes x\|_2^2, \quad \text{s.t. } Q^T Q = I. \tag{10}$$

where $H = [\text{Vec}(A_1), \text{Vec}(A_2), \dots, \text{Vec}(A_N)]$ and $\text{Vec}(\cdot)$ denotes the vectorization operation of the matrix, λ is the parameter to control the contribution of low-rank term.

3.2 Optimization

The above optimization problem can be reformulated as follows:

$$\min_{x,Q,S,T} \|E\|_* + \lambda \|T\|_* + \eta \|d \otimes x\|_2^2 \tag{11}$$

s.t. $E = yQ - A(x), T = H \text{diag}(x), Q^T Q = I$

The above nuclear norm optimization problem could be optimized by augmented Lagrange multipliers (ALM) method or the alternating direction method of multipliers

(ADMMs) [9, 10, 44, 45], using the next augmented Lagrange function:

$$L_\mu(x, Q, E, T) = \|E\|_* + \lambda \|T\|_* + \eta \|d \otimes x\|_2^2 + \text{Tr}(Z_1^T (yQ - A(x) - E)) + \text{Tr}(Z_2^T (H \text{diag}(x) - T)) + \frac{\mu}{2} (\|yQ - A(x) - E\|_F^2 + \|H \text{diag}(x) - T\|_F^2), \tag{12}$$

where $\text{Tr}(\cdot)$ represents the trace operator, $\mu > 0$ represents a penalty parameter, Z_1 and Z_2 represent the Lagrange multipliers. The above unconstrained problem can be solved by optimizing the variables one by one, until some convergence conditions are achieved.

(1) Updating Q : Given x, S and T , the optimization problem can be reformulated as:

$$\min_Q \frac{\mu}{2} \left\| yQ - A(x) - E + \frac{Z_1}{\mu} \right\|_F^2. \tag{13}$$

We set $P = A(x) + E - Z_1/\mu$. Following some simple algebraic steps, we have

$$\begin{aligned} \|yQ - P\|_F^2 &= \text{Tr}((yQ - P)^T (yQ - P)) \\ &= \text{Tr}(Q^T y^T y Q) - 2\text{tr}(y Q P^T) + \text{tr}(P^T P) \\ &= \|y\|_F^2 - 2\text{tr}(y Q P^T) + \|P\|_F^2. \end{aligned} \tag{14}$$

By using the SVD: $USV^T = P^T y$, the solution of optimal Q is given by $Q^{k+1} = VU^T$.

(2) Updating E : Fix x, Q and T , we can rewrite the objective function as:

$$\min_E \frac{1}{\mu} \|E\|_* + \frac{1}{2} \left\| E - \left(yQ - A(x) + \frac{1}{\mu} Z_1 \right) \right\|_F^2 \tag{15}$$

Its solution is given by the singular value shrinkage operator [46]:

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

where $(U, S, V^T) = \text{svd}(yQ - A(x) + Z_1/\mu)$; function svd represents the singular value decomposition scheme.

Symbol $T(\cdot)$ denotes the singular value thresholding factor, which is defined as

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

Here, r denotes the rank of matrix S .

(3) Updating T : Fix x, Q and E , our objective function can be rewritten as:

$$\min_T \frac{\lambda}{\mu} \|T\|_* + \frac{1}{2} \left\| T - \left(H \text{diag}(x) + \frac{1}{\mu} Z_2 \right) \right\|_F^2 \quad (18)$$

From the above discussion, the optimal solution of problem (18) is

$$T^{k+1} = UT_{\frac{\lambda}{\mu}}[S]V, \quad (19)$$

where $(U, S, V^T) = \text{svd}(H \text{diag}(x) + Z_2/\mu)$.

(4) **Updating x:** Fix Q, E and T , we can reformulate our optimization problem as:

$$\begin{aligned} \min_x \quad & \eta \|d \otimes x\|_2^2 \\ & + \frac{\mu}{2} \left(\left\| yQ - A(x) - E + \frac{1}{\mu} Z_1 \right\|_F^2 + \left\| H \text{diag}(x) - T + \frac{1}{\mu} Z_2 \right\|_F^2 \right). \end{aligned} \quad (20)$$

This is a quadratic problem related to variable w . Following some algebraic steps, the analytical solution of the above regularized optimization problem can be derived by:

$$\begin{aligned} x &= (b + \text{diag}(b_1)) \setminus b_2 \\ b &= \mu H^T H + 2\eta \text{diag}(d) \otimes \text{diag}(d) \\ b_1 &= \mu(H \otimes H)^T \mathbf{1} \\ b_2 &= \mu H^T b_3 + \mu(T \otimes H)^T \mathbf{1} - (Z_2 \otimes H)^T \mathbf{1} \\ b_3 &= \text{Vec}(yQ - E + Z_1/\mu). \end{aligned} \quad (21)$$

Here, column vector $\mathbf{1}$ has $N \times 1$ entries with all ones, and “ \setminus ” represents the left matrix division operation.

Algorithm 1. ADMM algorithm for solving N2SOPR

Input: A set of offline training patches $A_1, \dots, A_N \in \mathbb{R}^{p \times q}$ and probe image patch $y \in \mathbb{R}^{p \times q}$, parameters λ, μ , the termination condition parameter ε .

Initialize: $x = 0, Q = 0, E = 0, T = 0, \varepsilon = 10^{-6}$.

1: Fix others and set $P = A(x^{k+1}) + E^k - Z_1^k/\mu$. By using the SVD: $USV^T = P^T y$, we can update Q by $Q^{k+1} = VU^T$;

2: Fix others and update E via

$$\min_E \frac{1}{\mu} \|E\|_* + \frac{1}{2} \|E - (yQ - A(x) + Z_1/\mu)\|_F^2;$$

3: Fix others and update Z via

$$\min_T \frac{\lambda}{\mu} \|T\|_* + \frac{1}{2} \|T - (H \text{diag}(x) + Z_2/\mu)\|_F^2;$$

5: Fix others and update x according to (21);

6: Update the multiplies
 $Z_1 = Z_1 + \mu(yQ - A(x) - E)$;
 $Z_2 = Z_2 + \mu(H \text{diag}(x) - T)$;

7. If achieve termination condition (20), go to 8; otherwise go to 1.

8. **Output:** Optimal coding vector x^{k+1}

Finally, the next termination conditions are utilized:

$$\left(\|yQ - A(x) - E\|_\infty \leq \varepsilon \text{ and } \|H \text{diag}(x) - T\|_\infty \leq \varepsilon, \right) \quad (20)$$

where ε is a given tolerance. Algorithm 1 summarizes the whole optimization procedure.

3.3 Face hallucination via N2SOPR

Denoting A_H^m the HR training face images, and A_L^m ($m = 1, \dots, N$) its LR counterparts, our target is desired to restore the target HR version Y from its observed LR input y .

Firstly, as in [36], the input LR face image and each training one are decomposed into many overlapped matrix patches and represented as $y(i, j)$, $A_L^m(i, j)$ and $A_H^m(i, j)$, respectively. By using N2SOPR, each position patch $y(i, j)$ is denoted as a linear representation on the same position training patches $A_L^m(i, j)$ ($m = 1, \dots, N$). Then, we obtain the desired HR version $Y(i, j)$ by utilizing the same combination weights on the relevant HR training patch matrices. Last but not least, we could obtain the target HR face by means of integrating all the hallucinated HR patches and meanwhile averaging values in the superposed regions according to their corresponding positions. Algorithm 2 summarizes the whole face super-resolution algorithm in detail.

Algorithm 2. Using N2SOPR for Face hallucination

Input: HR training set $A_H = \{A_H^1, \dots, A_H^N\}$, corresponding LR counterpart $A_L = \{A_L^1, \dots, A_L^N\}$, observed LR image y .

1: Decompose the input LR test and each LR training image into M superposed patches;

2: **For** each divided patch:

a) Obtain the similarity d as follows;

$$d_m(i, j) = \|y(i, j) - A_L^m(i, j)\|_2, m = 1, \dots, N;$$

b) Obtain the desired linear representation $x^*(i, j)$ for the input LR patch $y(i, j)$ over the training dictionary A_L Using N2SOPR;

c) Obtain the HR patch via

$$Y(i, j) = \sum_{m=1}^N A_H^m(i, j) x_m^*(i, j);$$

4: **End for**

6: Reconstruct desired HR image Y by merging all the hallucinated HR patch matrices.

Output: Desired reconstructed HR face Y .

3.4 Complexity analysis

The computational cost of the proposed approach is discussed in this part. Because the super-resolution of HR face images involves steps of matrix multiplications and additions, we only access the time complexity of the representation learning in the proposed method. The major computation of Algorithm 1 depends on two parts: (a) the SVD computing and (b) combination weights learning. It can be found that there are mainly four parts affecting the time cost: the patch number M in one image, the training samples number N , maximum iteration times maxIter and the dimension pq of one patch.

In steps 1, 2 and 3, the time complexity of SVD operation is $O(q^3)$, $O(pq^2)$ and $O(pqN^2)$, respectively. Following [30], it takes $O(pqN^3)$ for the representation learning

phase. Thus, it costs $O(q^3 + pq^2 + pqN^2 + pqN^3)$ for each patch in Algorithm 1. By taking into account the maximum iteration times and patch number, the computational cost of our proposed N2SOPR algorithm is about $O(\maxIter(q^3 + pq^2 + pqN^2 + pqN^3)M)$. For comparison, the cost of LCR and SSR is about $O(pqN^3M)$ and $O(\maxIter^2pqN^3M)$, respectively.

4 Experiments and discussion

4.1 Database description

To access the efficiency of our proposed method, we conduct tests on the FERET face database [47]. This database includes samples from 200 persons; each person was captured at nine viewpoints $ba, bb, bc, bd, be, bf, bg, bh, bi$, respectively. The database also contains frontal images (denoted as bk) corresponding to viewpoint ba , but

with different lighting conditions. We select 180 frontal images from 180 subjects as training samples (some samples are listed in Fig. 1), and other 20 face images from 20 subjects with pose variations as the testing set (some samples are listed in Fig. 2). All the faces are manually cropped and resized to the size of 80×80 . By smoothing (filter size is 4×4) and down-sampling (down-sampling factor is 4) the HR faces, we can form the corresponding LR ones, whose size is 20×20 .

4.2 Parameter analysis

To achieve the best performance, we adjust the parameters for our algorithm. In our method, the first parameter λ is applied to balance the low-rank term, and the second parameter η is used to control the locality constraint. We adjust our parameters to find the optimal value by means of the peak signal-to-noise ratio (PSNR) and Structural SIMilarity (SSIM) [48] scores.

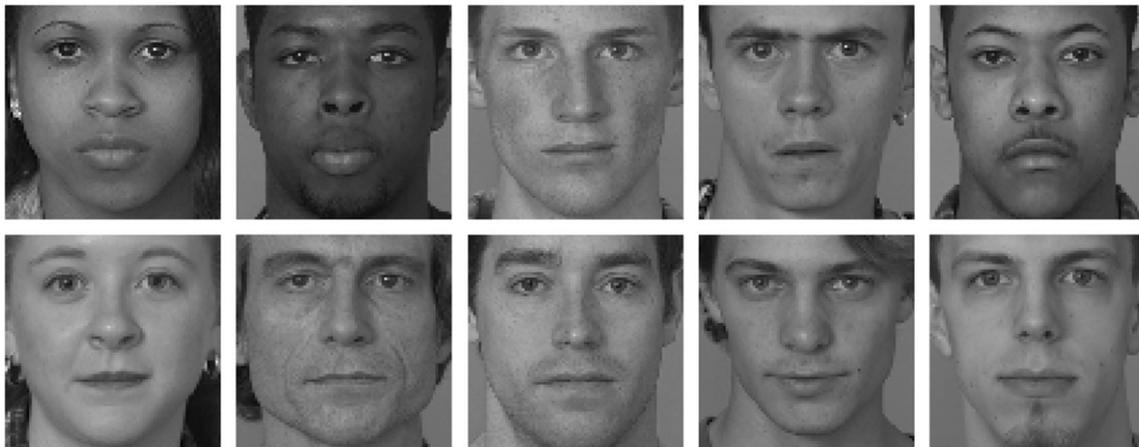


Fig. 1 Some training sample images

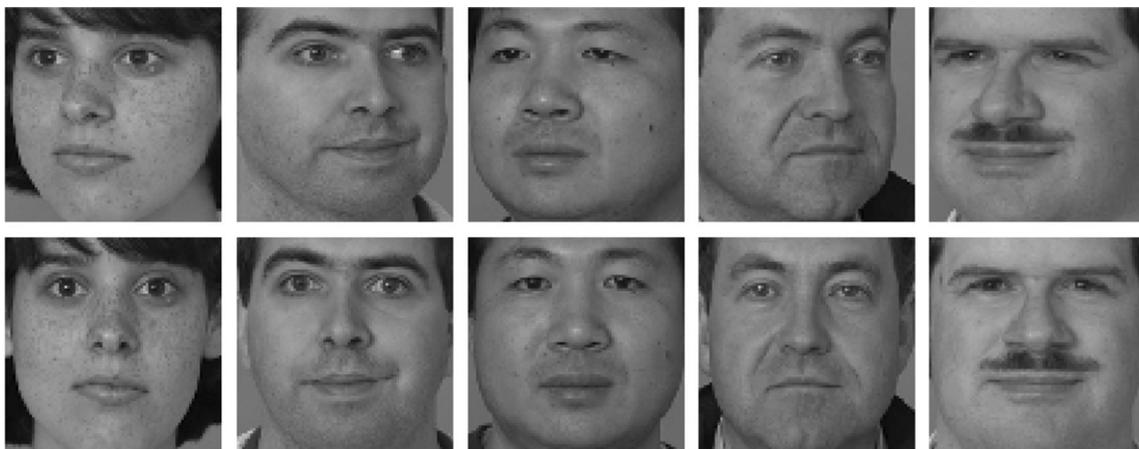


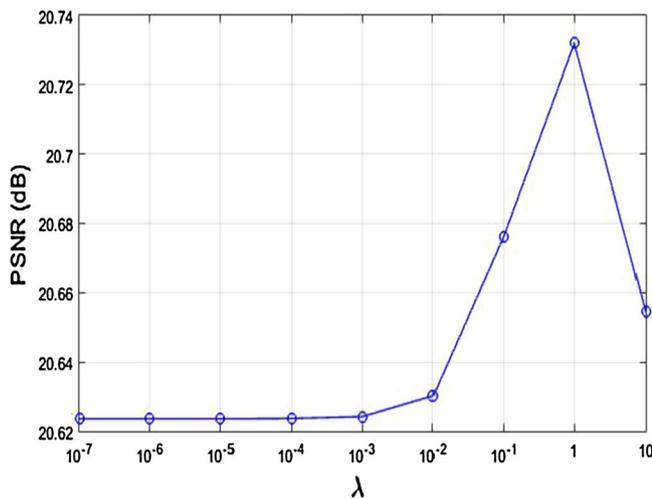
Fig. 2 Some testing images with pose variations and their corresponding frontal counterparts

4.2.1 Influence of parameter λ

We first fix the value of η and set the range of λ as 10^{-7} –10. The performance with different λ by means of PSNR and SSIM scores is shown in Fig. 3, from which we can survey that the PSNR and SSIM scores grow with the increase in λ . When λ is set to 1, our method can obtain the best performance. When λ is larger than 1, the PSNR and SSIM scores begin to decline. So we can select the value of the parameter λ as 1. In Fig. 4, we can also find that, when $\lambda = 1$, the recovered image is clearest and has many facial details. But when $\lambda = 10$, the recovered image is blurred again. Because in the process of synthesis, the low-rank property is over-strengthened, resulting in increased reconstruction error.

4.2.2 Influence of parameter η

We then fix the value of λ and set the range of η as 10^{-7} –10. Figure 5 shows the performance with different η in



terms of PSNR and SSIM scores. Figure 6 displays some super-resolved faces with different η . From the above figures, we can intuitively see that with η increases, the average PSNR and SSIM values increase. When η is set to 10^{-4} , the average PSNR reaches the peak and then gradually decreases and tends to be stable with larger η . However, when the η is set to 10^{-5} , the average SSIM reaches the peak. The change in terms of PSNR and SSIM scores may be slightly different because the locality parameter η ensures the quality of the face image synthesis in a certain degree. As the SSIM value increases, the proportion of errors in the synthesized face image decreases, and the face structure is affected. By taking both PSNR and SSIM values into account, we set η to 10^{-4} .

4.3 Comparisons with other approaches

To access the efficiency of our N2SOPR approach, we compare it with other state-of-the-art approaches in this part: LSR method [33], SR method [34], LCR method [36],

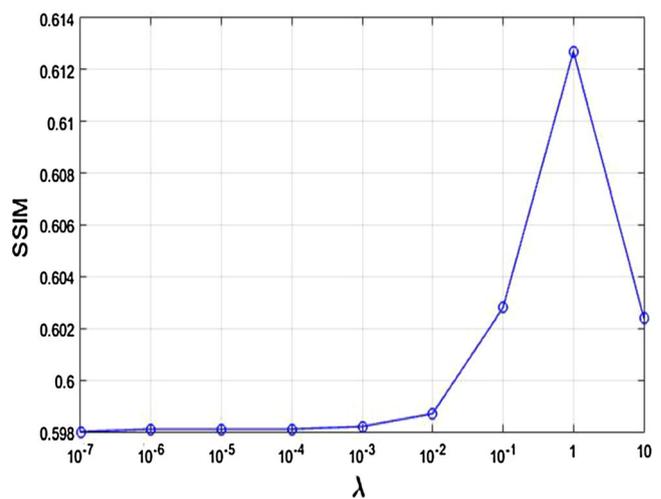


Fig. 3 PSNR and SSIM values of our method with different λ

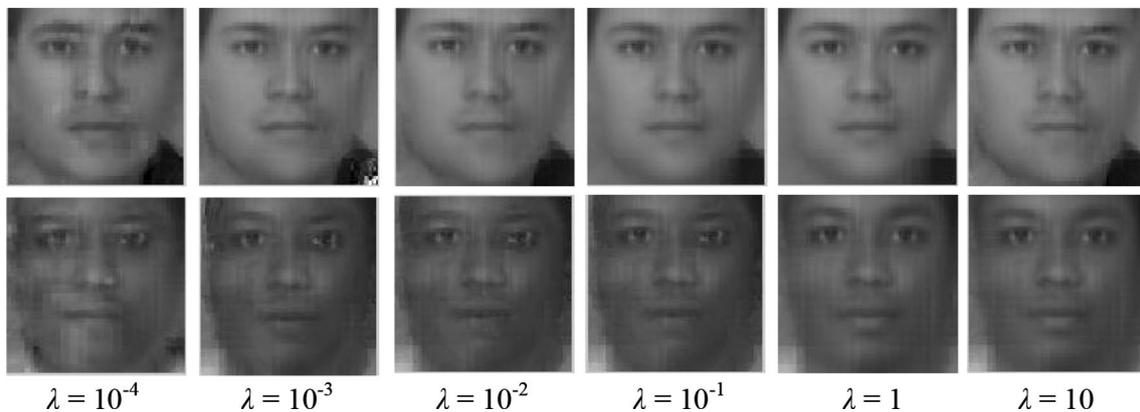


Fig. 4 Hallucinated faces with different λ values

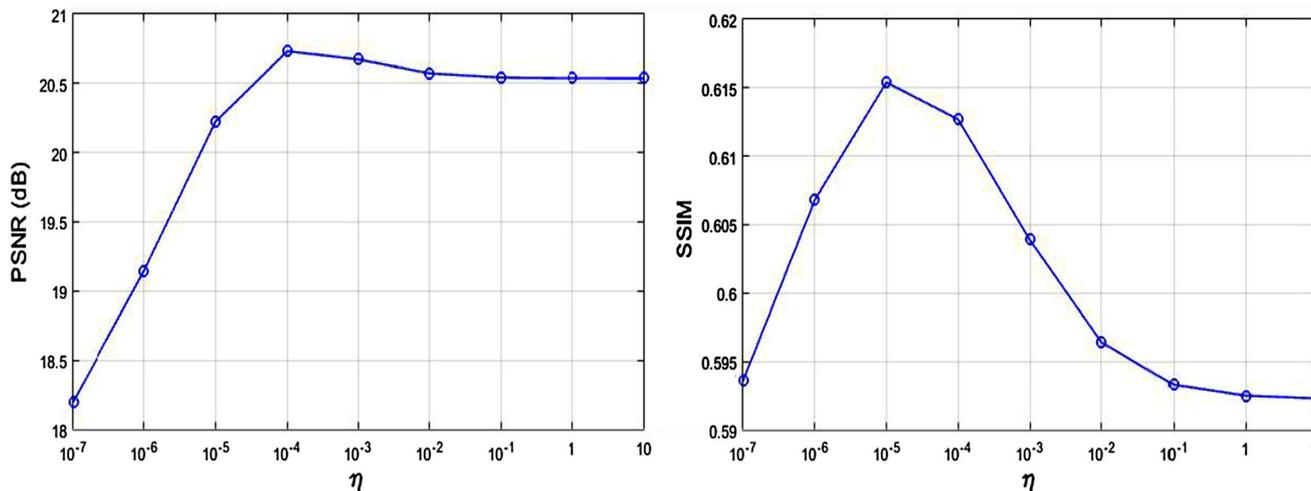


Fig. 5 PSNR and SSIM values of our method with different η

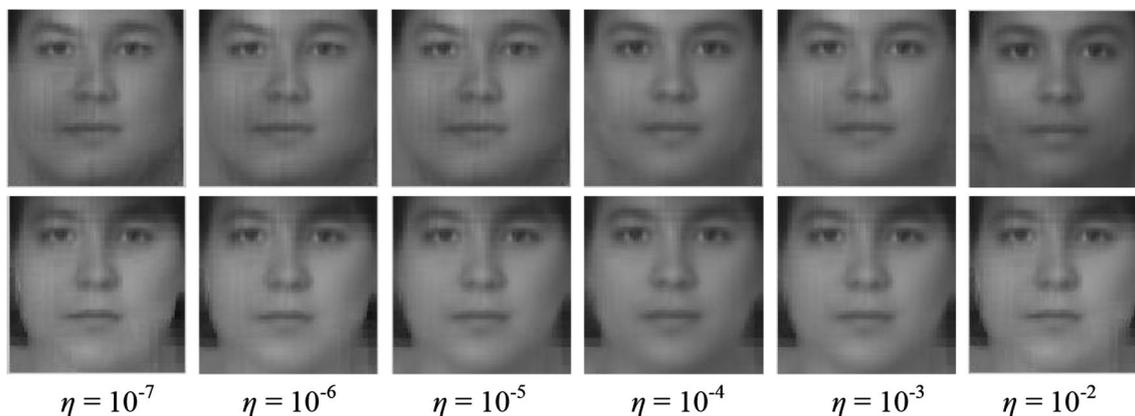


Fig. 6 Hallucinated faces with different η values

SRCNN method [39] and SSR method [37]. As for those local approaches, we suggest using the size of 12 × 12 pixels for HR patch, and there is 12 × 4 pixels overlapped with neighbor patches. Thus, the relevant LR patch has size of 3 × 3 pixels, so there is 3 × 1 pixel overlapped with its neighbors. For reasonable comparison, the parameters of all comparative approaches are tuned to obtain their best possible performance. In addition to the visual comparisons of different methods, the performance of each method is also quantified by means of PSNR and SSIM [48] values between the reconstructed images and the ground truth ones.

Some typical super-resolved faces via different approaches are listed in Fig. 7. We can see that the images synthesized by LSR and SR methods are seriously distorted, especially in the contour of eyes. Compared with these three methods, the results of LCR method are improved. But the synthesized face images are still blurred. The face image synthesized by SRCNN method has richer face details, but there is still blurring. In addition to the

distortion, we can also find that the pose variations are still existed in the synthesized images, which will reduce the performance of following recognition system. Although the face images synthesized by the SSR method still have a little noise in the faces, the details of nose, eyes and mouth are basically recovered. By introducing an orthogonal matrix low-rank constraint, our N2SOPR method can obtain more clear images, and the recovered faces are closed to the frontal HR face images. The objective comparisons by means of PSNR and SSIM scores are also given in Table 1. All the above results show a considerable superior of our method over some conventional learning-based methods and recently proposed deep learning-based method.

4.4 Recognition experiments

For face recognition task, we prefer to obtain near-frontal faces in some applications. To this end, a face recognition test is conducted in this part, using the hallucinated HR



Fig. 7 Hallucinated faces on the FERET database by different methods. From top to bottom: LR input face images, the hallucinated face images by LSR [33], SR [34], LCR [36], SRCNN [39], SSR [37], our N2SOPR method and the original HR frontal face images

Table 1 The average PSNR and SSIM values of different methods on the FERET database

Methods	LSR [33]	SR [34]	LCR [36]	SRCNN [39]	SSR [37]	N2SOPR
PSNR (dB)	18.3648	18.5988	19.4243	19.4464	19.7119	20.7319
SSIM	0.3899	0.4171	0.5045	0.5156	0.5367	0.6127

faces of our proposed method. We select 400 frontal face images from FERET for training, with two faces for each subject. One non-frontal face per subject is selected for testing. All the test faces are smoothed and down-sampled to 40×40 pixels. It should be noted that each LR input is

hallucinated by using “leave-one-out” scheme, and the faces belong to the testing individual are removed from the training samples. For the sake of simplicity, the sparse representation-based classifier (SRC) [25] is applied to recognize the hallucinated faces.

Table 2 Face recognition rates (%) associated with different hallucinated methods

Methods	LSR [33]	SR [34]	LCR [36]	SRCNN [39]	SSR [37]	N2SOPR
RA	25.50	25.50	26.00	26.50	28.50	32.50

Table 2 tabulates the recognition rates (RAs) associated with comparable hallucinated approaches. In Table 2, we can observe that the hallucinated HR faces by our N2SOPR method could obtain a better performance than other approaches. Previous SR methods ignore the pose variations when we learn the representation weights, resulting in the hallucinated faces sensitive to these variations. We mainly attribute the superiority of N2SOPR to its ability to take into account these variations in the hallucinated faces.

5 Conclusion and future work

This study presents a nuclear norm regularized structural orthogonal Procrustes regression (N2SOPR) method for face hallucination with pose variations. The orthogonal Procrustes regression is applied to seek an appropriate transformation between two data matrixes to make the pose of one data to adapt that of the other, and we also impose the nuclear norm regularization on the representation residual to preserve image inherent structural property. In addition, the low-rank property and locality constraint are taken into consideration on the representations, desiring to adaptively cluster the input into the same subspace as the training samples. Both hallucination and recognition experiments on the FERET face database have demonstrated that our approach could obtain better result than some related methods.

We mainly discuss the effect of pose in this paper. However, noise, expression and illumination variations also exist in real-world applications. Also, how to take into account the facial structure prior in deep learning-based models needs further investigation in the future work.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China under Grant Nos. 61502245, 61772568, 61571236, 61473086, the Natural Science Foundation of Jiangsu Province under Grant No. BK20150849, the Fundamental Research Funds for the Central Universities of China (No. 18lgzd15), Open Fund Project of Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education (Nanjing University of Science and Technology) (No. JYB201709).

References

- Gao G, Yang J, Wu S et al (2015) Bayesian sample steered discriminative regression for biometric image classification. *Appl Soft Comput* 37:48–59

- Huang P, Gao G (2016) Parameterless reconstructive discriminant analysis for feature extraction. *Neurocomputing* 190:50–59
- Jing X-Y, Wu F, Zhu X et al (2016) Multi-spectral low-rank structured dictionary learning for face recognition. *Pattern Recogn* 59:14–25
- Lai Z, Wong WK, Xu Y et al (2016) Approximate orthogonal sparse embedding for dimensionality reduction. *IEEE Trans Neural Netw Learn Syst* 27(4):723–735
- Mudunuri SP, Biswas S (2016) Low resolution face recognition across variations in pose and illumination. *IEEE Trans Pattern Anal Mach Intell* 38(5):1034–1040
- Shen F, Shen C, Zhou X et al (2016) Face image classification by pooling raw features. *Pattern Recogn* 54:94–103
- Tai Y, Yang J, Zhang Y et al (2016) Face recognition with pose variations and misalignment via orthogonal Procrustes regression. *IEEE Trans Image Process* 25(6):2673–2683
- Deng W, Hu J, Wu Z et al (2017) From one to many: pose-aware metric learning for single-sample face recognition. *Pattern Recogn* 77:426–437
- Gao G, Yang J, Jing X-Y et al (2017) Learning robust and discriminative low-rank representations for face recognition with occlusion. *Pattern Recogn* 66:129–143
- Yang J, Luo L, Qian J et al (2017) Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Trans Pattern Anal Mach Intell* 39(1):156–171
- Yang M, Wang X, Zeng G et al (2017) Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person. *Pattern Recogn* 66:117–128
- Hamedani K, Seyyedsalehi SA, Ahamdi R (2016) Video-based face recognition and image synthesis from rotating head frames using nonlinear manifold learning by neural networks. *Neural Comput Appl* 27(6):1761–1769
- Han B, He B, Sun T et al (2016) HSR: L 1/2-regularized sparse representation for fast face recognition using hierarchical feature selection. *Neural Comput Appl* 27(2):305–320
- Zhu Y, Xue J (2017) Face recognition based on random subspace method and tensor subspace analysis. *Neural Comput Appl* 28(2):233–244
- Wu F, Jing X-Y, Liu Q et al (2017) Large-scale image recognition based on parallel kernel supervised and semi-supervised subspace learning. *Neural Comput Appl* 28(3):483–498
- Lan R, Zhou Y, Tang YY (2017) Quaternionic weber local descriptor of color images. *IEEE Trans Circuits Syst Video Technol* 27(2):261–274
- Zou WW, Yuen PC (2012) Very low resolution face recognition problem. *IEEE Trans Image Process* 21(1):327–340
- Freeman WT, Pasztor EC, Carmichael OT (2000) Learning low-level vision. *Int J Comput Vis* 40(1):25–47
- Wang X, Tang X (2005) Hallucinating face by eigen transformation. *IEEE Trans Syst Man Cybern Part C Appl Rev* 35(3):425–434
- Hu Y, Lam KM, Shen T et al (2011) A novel kernel-based framework for facial-image hallucination. *Image Vis Comput* 29(4):219–229
- Shi J, Liu X, Qi C (2014) Global consistency, local sparsity and pixel correlation: a unified framework for face hallucination. *Pattern Recogn* 47(11):3520–3534

22. Huang H, He H, Fan X et al (2010) Super-resolution of human face image using canonical correlation analysis. *Pattern Recogn* 43(7):2532–2543
23. An L, Bhanu B (2014) Face image super-resolution using 2D CCA. *Signal Process* 103:184–194
24. Gao G, Yang J (2014) A novel sparse representation based framework for face image super-resolution. *Neurocomputing* 134:92–99
25. Wright J, Yang AY, Ganesh A et al (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 31(2):210–227
26. Bo C, Wang D (2015) A registration-based tracking algorithm based on noise separation. *Optik-Int J Light Electron Opt* 126(24):5806–5811
27. Li F, Lu H, Wang D et al (2016) Dual group structured tracking. *IEEE Trans Circuits Syst Video Technol* 26(9):1697–1708
28. Zhao W, Lu H, Wang D (2018) Multisensor image fusion and enhancement in spectral total variation domain. *IEEE Trans Multimed* 20(4):866–879
29. Chang H, Yeung D-Y, Xiong Y (2004) Super-resolution through neighbor embedding. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, pp 1275–1282
30. Jiang J, Hu R, Wang Z et al (2014) Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning. *IEEE Trans Image Process* 23(10):4220–4231
31. Jiang J, Hu R, Wang Z et al (2016) Facial image hallucination through coupled-layer neighbor embedding. *IEEE Trans Circuits Syst Video Technol* 26(9):1674–1684
32. Yang J, Wright J, Huang TS et al (2010) Image super-resolution via sparse representation. *IEEE Trans Image Process* 19(11):2861–2873
33. Ma X, Zhang J, Qi C (2010) Hallucinating face by position-patch. *Pattern Recogn* 43(6):2224–2236
34. Jung C, Jiao L, Liu B et al (2011) Position-patch based face hallucination using convex optimization. *IEEE Signal Process Lett* 18(6):367–370
35. Wang Z, Hu R, Wang S et al (2014) Face hallucination via weighted adaptive sparse regularization. *IEEE Trans Circuits Syst Video Technol* 24(5):802–813
36. Jiang J, Hu R, Wang Z et al (2014) Noise robust face hallucination via locality-constrained representation. *IEEE Trans Multimed* 16(5):1268–1281
37. Jiang J, Ma J, Chen C et al (2017) Noise robust face image super-resolution through smooth sparse representation. *IEEE Trans Cybern* 47(11):3991–4002
38. Liu L, Chen CP, Li S et al (2018) Robust face hallucination via locality-constrained bi-layer representation. *IEEE Trans Cybern* 48(4):1189–1201
39. Dong C, Loy CC, He K et al (2016) Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell* 38(2):295–307
40. Zhang Y, Li K, Li K et al (2018) Image super-resolution using very deep residual channel attention networks. *arXiv preprint arXiv:1807.02758*
41. Zhang Y, Tian Y, Kong Y et al (2018) Residual dense network for image super-resolution. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, pp 2472–2481
42. Kim S-J, Koh K, Lustig M et al (2007) An interior-point method for large-scale ℓ_1 -regularized least squares. *IEEE J Sel Top Signal Process* 1(4):606–617
43. Hurley JR, Cattell RB (1962) The Procrustes program: producing direct rotation to test a hypothesized factor structure. *Behav Sci* 7(2):258–262
44. Zhang F, Yang J, Tai Y et al (2015) Double nuclear norm-based matrix decomposition for occluded image recovery and background modeling. *IEEE Trans Image Process* 24(6):1956–1966
45. Chen J, Yang J, Luo L et al (2015) Matrix variate distribution-induced sparse representation for robust image classification. *IEEE Trans Neural Netw Learn Syst* 26(10):2291–2300
46. Cai JF, Candes EJ, Shen ZW (2010) A singular value thresholding algorithm for matrix completion. *SIAM J Optim* 20(4):1956–1982
47. Phillips PJ, Wechsler H, Huang J et al (1998) The FERET database and evaluation procedure for face-recognition algorithms. *Image Vis Comput* 16(5):295–306
48. Wang Z, Bovik AC, Sheikh HR et al (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612