Robust Framework of Single-Frame Face Superresolution Across Head Pose, Facial Expression, and Illumination Variations

Xiang Ma, Huansheng Song, and Xueming Qian, *Member, IEEE*

***Abstract*—This paper presents a robust framework to solve the face hallucination problem across multiple factors, i.e., different expressions, head poses, and illuminations. It proposes a redundant transformation with diagonal loading for modeling the mappings among different new face factors, and a local reconstruction with geometry and position constraints for incorporating image details in the new factor spaces. Our proposed redundant and sparse strategies are discussed, and the experiments indicate that it is not necessary to adopt sparse representation in the proposed framework. The experimental results demonstrate that the proposed framework offers robustness when dealing with the inputs that have different expressions, head poses, and illuminations compared with the state-of-the-art methods, can generate high-resolution face images with better image qualities than the hierarchical tensor-based method, and improves the state of the art from single one output to multiple outputs with new factors.**

***Index Terms*—face hallucination, Face superresolution, superresolution.**

I. Introduction

Face images are the only biometric information available in some legacy databases and can be acquired even without the subjects' cooperation. Unlike traditional access control scenarios, where facial images are taken under controlled illumination, head pose, and expression, images in other domains suffer from uncontrolled illumination, large pose variation, a range of facial expressions, make-up, and severe partial occlusions. The goal of superresolution is to recover one or multiple high-resolution (HR) images from low-resolution (LR) image sequences or a single LR one [1].

The problem can be stated as that of recovering an HR image x, from its LR version y. We model the relation between these two by

$$\mathbf{y} = \mathbf{SHx} = \mathbf{Lx} \tag{1}$$

where $\boldsymbol{H}$ is a linear filter that models certain low-pass filtering (blurring, e.g., with a Gaussian kernel), $\boldsymbol{S}$ is a down-sampling operator, and $\boldsymbol{L} = \boldsymbol{S}H$. The dimension of $\mathbf{y}$ is significantly smaller than that of $\mathbf{x}$; thus there are infinitely many possible vectors $\mathbf{x}$ that satisfy the above equation. To obtain a unique and "good" HR image, proper regularization is needed by imposing certain priors on the solution.

The approach to generate an HR image from multiple LR images is called multiple-frame superresolution [2]. The approach to generate an HR image from a single LR observation with a set of training images is called single-frame superresolution [3]. Superresolution can also be classified into general superresolution and domain specific superresolution, for example, face superresolution, according to the type of applied LR images.

Single-frame face superresolution based on training sets, also known as face hallucination, is attractive for numerous applications including visual surveillance and security, social networking websites. Baker *et al.* [4] coined the term "face hallucination" and developed a face hallucination method using a Bayesian formulation. Liu *et al.* [5] presented a two-step approach integrating a global parametric model with Gaussian assumption and a local nonparametric model based on Markov random fields. Inspired by locally linear embedding, a well-known manifold learning method, Chang *et al.* [6] developed neighbor embedding based on the assumption that the LR and HR training images form manifolds with similar local geometry in two distinct feature spaces. Following [5], the authors in [7]–[11] treat face hallucination as a two-step problem. Ma *et al.* [12] proposed a fast local one-step method, in which patch position in the face image is used as well as image features to synthesize an HR face image. Yang *et al.* [27] applied sparse representation to superresolution by training all atoms to construct only a single dictionary. Ma *et al.* [1] classified all atoms to small dictionaries according to the different regions of human face and obtained better results. Zhang and Cham [34] propose a learning-based face hallucination method in the discrete cosine transform. Hu *et al.* [35] used the input LR face and the learned pixel structures as priors to estimate the target HR face. An approach based on similarity constraints is proposed by Li *et al.* [36]. The authors in [7]–[11] and [35] only performed on the inner facial parts. These algorithms fail to synthesize hair, and facial contour lines, not important for face recognition, but important for face superresolution.

Most face hallucination methods are limited to a frontal face without consideration of illumination, head pose, and expression variations. Some methods such as in [12] consider face

Manuscript received December 1, 2013; revised May 7, 2014 and September 12, 2014; accepted November 9, 2014. Date of publication December 11, 2014; date of current version March 12, 2015. This work was supported by the National Natural Science Foundation of China under Grant 61101215 and Grant 61173109, the China Postdoctoral Science Foundation under Grant 2013M542308 and Grant 2014T70895, and the China Fundamental Research Funds for the Central Universities (No. 2014G3242005).

X. Ma and H. Song are with the School of Information Engineering, Chang'an University, Xi'an 710048, China (e-mail: maxiangmail@163.com; hshsong@chd.edu.cn).

X. Qian is with the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: qianxm@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2014.2375329

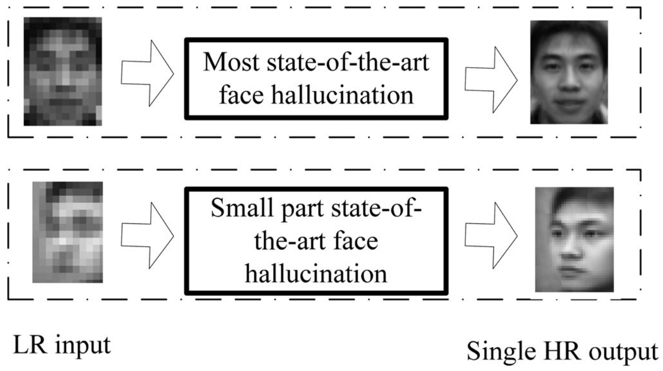

Most state-of-the-art face hallucination
Small part state-of-the-art face hallucination
LR input
Single HR output

Fig. 1. State-of-the-art methods only generate a single output with the same factor of LR input.

variations, but they are limited in that both the HR output and LR input have the same view (see Fig. 1). Thus, they cannot generate an HR front face from a profile face, which is important for identification or recognition. In real-world applications, it is essential to generate HR outputs with new factors from an LR input, e.g., to obtain the HR frontal face with normal illumination from an LR profile face with nonnormal illumination.

The face images in surveillance videos are normally LR and nonfrontal with diverse expressions and illuminations. It is essential to obtain the corresponding HR face from the LR face at multiple factors (different expressions, head poses, and illuminations) in a video. However, most of the state-of-the-art face hallucination methods fail to do it. Our current proposed framework addresses this issue (see Fig. 2).

We next briefly review the methods with consideration of different poses, expressions, and illuminations in face superresolution or face recognition system. Li and Lin [19] used a texture model [37] and Gabor wavelet features to synthesize the corresponding frontal face. Many ideas of generating new face views have been reported in face recognition systems, e.g., the 2-D technique based on active appearance models [13], and the technique based on complicated 3-D face morphable model [14], [15], etc. Tian and Fan [16] adopted a tensor [17], [18] framework with manifold learning to explore the relationship between multiview faces. The authors in [16] adopted tensor [17], [18] framework with manifold learning to explore the relationship between multiview faces. Vetter [39] separated texture and shape of the face and used a 3-D model to produce a new view. Chai *et al.* [20] used locally linear regression for pose-invariance. These face transformation methods require face features or face shape from face image input. Because it is nearly impossible to obtain face features, or an accurate shape in an LR face, e.g., the size of 16 × 12, they cannot work on an LR face image and, thus, cannot be used for superresolution. Other face transformation methods (see, e.g., [13]–[16], and [39]) cannot construct hair, ears, or facial contour lines successfully. While some have addressed the expression and illumination problem [20]–[23], these techniques are also only applied to HR face images and cannot be used for superresolution.

Jia and Gong [24]–[26] presented a generalized approach based on a hierarchical tensor for hallucinating HR face images across multiple factors, achieving generalization to variations in expression, pose, and illuminations. However, a tensor is a general extension of traditional linear methods [26]. A 2-D tensor is a matrix singular value decomposition, which is similar to principal component analysis (PCA). The algorithm of PCA eliminates nonfeature information and keeps feature information. Superresolution aims to incorporate image details and requires information not be lost during the processing. However, the principal component is kept and nonfeature information is abandoned in the hierarchical tensor framework. Nonfeature information is important to superresolution, because facial detail should be recovered. Sparse representation takes much computation time to select only a small number of training atoms for obtaining image details, and most training atoms that contain certain information are not used. Therefore, the tensor and sparse representation models have problems being adopted for superresolution.

Almost all state-of-the-art face hallucination methods fail to generate HR outputs with new factors from an LR input, e.g., fail to obtain the HR frontal face with normal illumination from an LR nonfrontal face with nonnormal illumination. This paper presents a framework that can produce multiple HR faces with new factors from a single given LR input and is robust to LR inputs with multiple illuminations, poses, and expressions.

The rest of this paper is organized as follows. Section II describes the proposed framework. Section II-A presents the methods for face reconstruction in the same resolution space. Section II-B presents the methods for face transformation of multiple factors in the same resolution space. Section II-C presents face transform of multiple factors in LR space. Section II-D presents local patch-based method for incorporating face details. Section III presents an evaluation, and Section IV concludes this paper.

II. Proposed Framework

A. *Redundant Linear Combination for Face Reconstruction in the Same Resolution Space*

Face images can be synthesized from the linear combination of training samples because of structural similarity [7]. We first find that faces of different poses, illumination, and expressions can also be reconstructed from redundant linear combination of other samples with small error (see Fig. 3).

A face image is represented as a column vector of all pixel values. Let I denote a face of a certain factor. We have

$$I \cong w_1 L_1 + w_2 L_2 + \cdots + w_N L_N = \sum_{i=1}^{N} w_i L_i \qquad (2)$$

where $L_1, L_2, \ldots, L_N$ are the training faces at the same factor with I; N is the maximum number of the training faces at a certain factor; and $w_1, w_2, \ldots, w_N$ are the construction coefficients. Every w_i indicates the contribution of L_i to reconstruct the input face I.

Equation (2) can be rewritten as

$$I \cong W\mathbf{L} = \tilde{I}. \qquad (3)$$

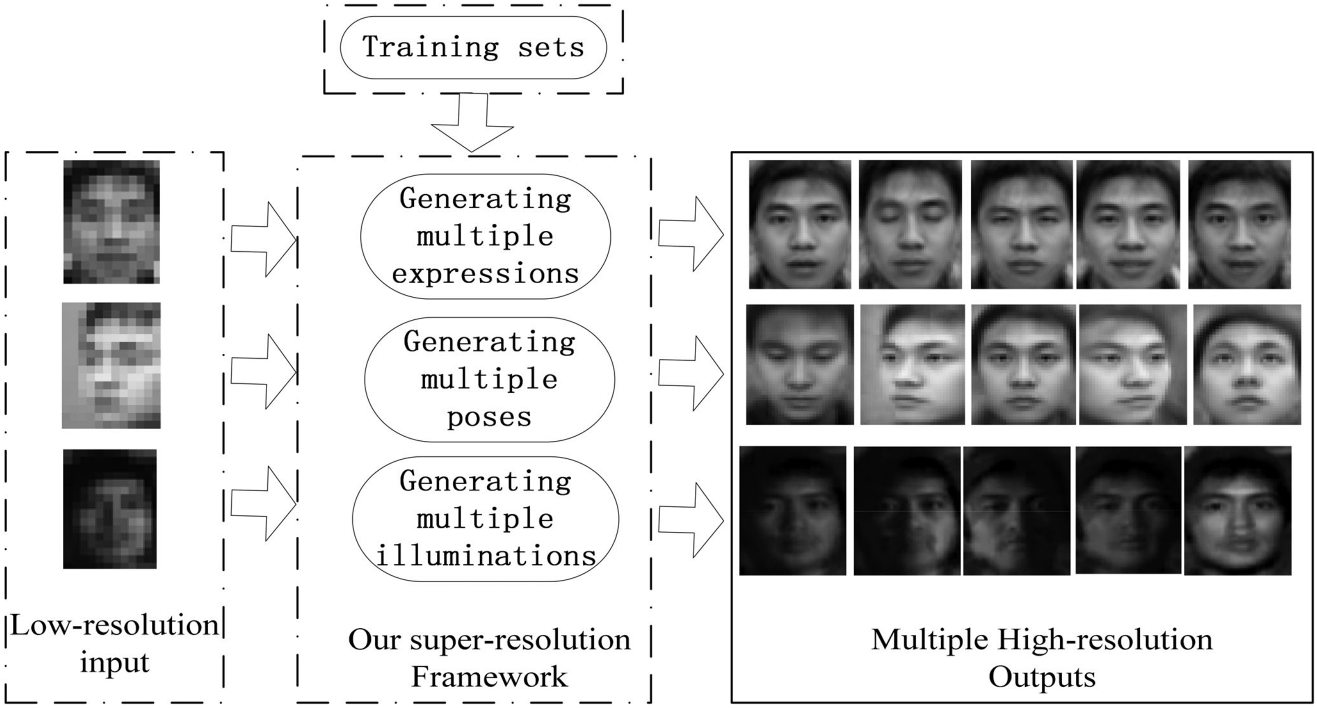

Training sets
Generating multiple expressions
Generating multiple poses
Generating multiple illuminations
Low-resolution input
Our super-resolution Framework
Multiple High-resolution Outputs

Fig. 2. Proposed framework can generate multiple outputs with new factors

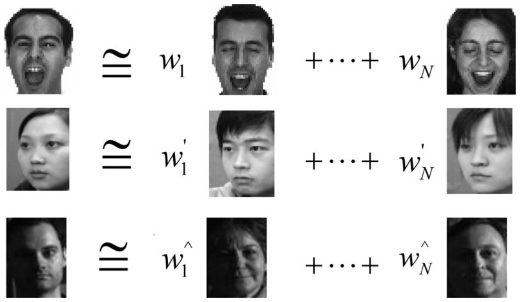


Fig. 3. Redundant linear combination of faces at different factors.

The reconstruction error θ is measured between the face input and the redundant linear combination of training samples:

$$\theta = \|I - W\mathbf{L}\|^2 = \left\|I - \tilde{I}\right\|^2. \tag{4}$$

We assume that if the reconstruction is successful, the distance is very small. The optimal weights W are obtained with the constraint of minimization of θ. Let all coefficients w sum to one. We have

$$W = \underset{w_1, w_2, \ldots, w_N}{\arg\min}\ \theta. \tag{5}$$

It is a constrained least squares problem that can be obtained using the steps in [28]

$$w_n = \left(\sum_{k=1}^{N} \mathbf{C}_{nk}^{-1}\right) \Bigg/ \left(\sum_{l=1}^{N}\sum_{m=1}^{N} \mathbf{C}_{lm}^{-1}\right) \tag{6}$$

where l and m are integers, $\mathbf{C}$ is the local covariance matrix, and $C_{ik} = (I - L_i)^T(I - L_k)$.

For a given face at a certain factor, it can be reconstructed using the coefficients obtained above in the same resolution space with some acceptable errors. Fig. 4 shows that the image I can be generated as $\tilde{I}$ using redundant reconstruction. A set of redundant reconstruction weights is also shown in Fig. 4(c).

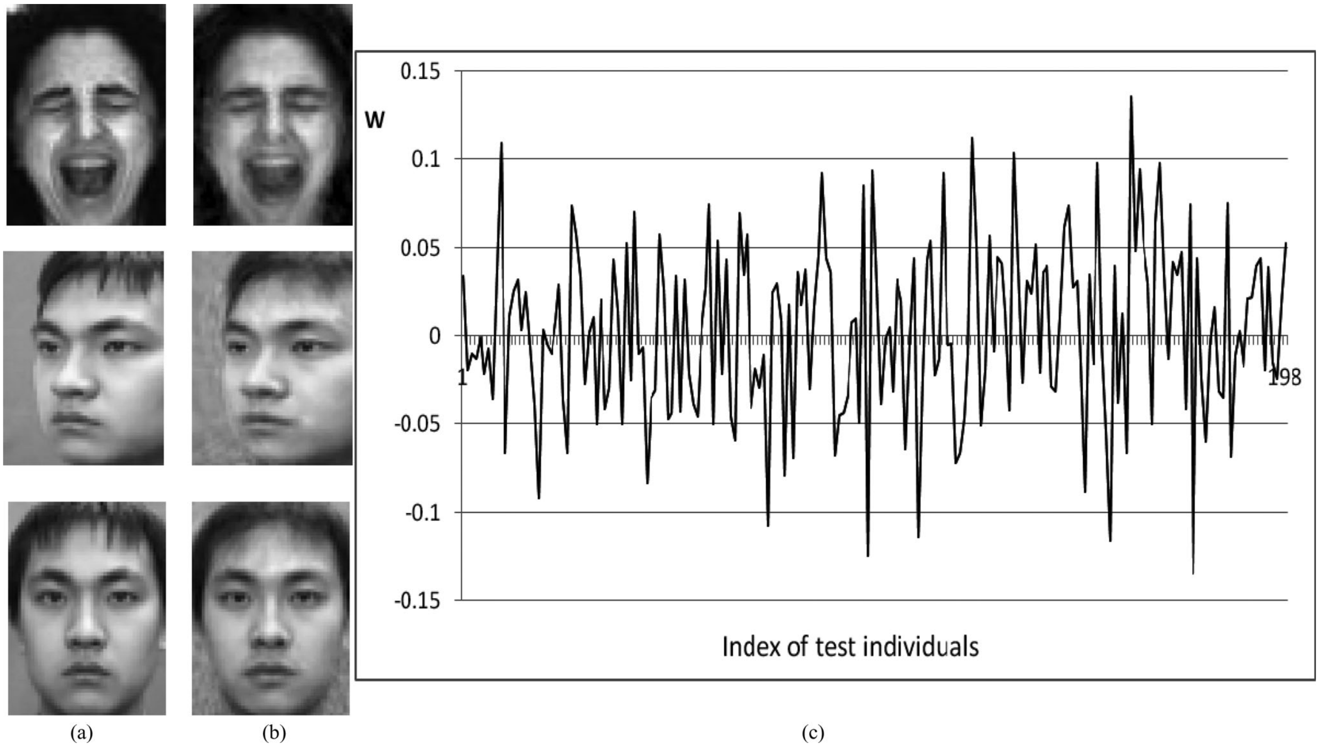
A given face at a certain factor can be reconstructed using the coefficients obtained above in the same resolution space with some acceptable errors. Some results are given in Fig. 4, which show that the image$I$ can be generated as $\tilde{I}$ using redundant reconstruction. A set of redundant reconstruction weights is given in Fig. 4(c).

In practice, the solution to obtain *W* may not be unique, and one approach is to impose several regularization terms. Sparse representation theory can be used to obtain *W*. It is converted to a standard sparse representation problem:

$$\min_{\mathrm{w}} \|W\|_1 \ \text{ subject to } \|I - L \cdot W\|_2^2 \leq \theta \tag{7}$$

where $|| \bullet ||_1$ denotes the ℓ_1-norm. This sparsity constraint can ensure that the underdetermined equation has an exact solution to obtain *W*.

Some representative results are given in Fig. 5. The proposed redundant representation has better image quality over the proposed sparse representation. The solution to (7) requires large computational resources. Even if both methods have the same image qualities, redundant representation is still superior. Therefore, the sparse representation strategy is unnecessary.



(a) (b) (c)

Fig. 4. Face redundant reconstruction. (a) Input. (b) Reconstruction result. (c) Redundant reconstruction weights.

B. Generating New Factors in the Same Resolution Space

Most of the methods for multiple factors transformation are limited to HR images with a single factor (see, e.g., [20] and [39]). These methods require obtaining face features, or face shape from face image input. Because it is nearly impossible to obtain face features, or an accurate shape in an LR face (e.g., the size of 16 × 12 pixels) they cannot be used for superresolution. Other face transformation methods (see, e.g. [13]–[16], and [39]) cannot address hair, ears, or facial contour lines. Some methods addresses the expression and illumination problem [21]–[23], but these techniques are also only applied to HR face images and cannot be used for superresolution. We propose a new method which is applicable to LR images.

Let the term o denote multiple factors. Suppose that I_p is an input face of factor p, and I_o is its corresponding faces of factor o. The following produces I_o from I_p in the same resolution space. The training faces of factor p are represented as L_p^1, $L_p^2, \ldots, L_p^N$, whose same resolution correspondences at factor o are $L_o^1, L_o^2, \ldots, L_o^N$.

From (2)–(6), we have

$$I_p \cong W_p L_p^n \tag{8}$$

$$W_p = \underset{w^n}{\arg\min} \left\| I_p - W_p L_p^n \right\|^2 \text{ s.t.} \sum_{n=1}^{N} w_P^n = 1 \tag{9}$$

$$I_o \cong W_o L_o^n \tag{10}$$

$$W_o = \underset{w^n}{\arg\min} \| I_o - W_o L_o^n \|^2 \text{ s.t.} \sum_{n=1}^{N} w_o^n = 1. \tag{11}$$

In (9), W_p can be obtained using (6) because I_p and L_p^n are known:

$$w_P^n = \left(\sum_{k=1}^{N} (\mathbf{C}_{nk})^{-1} \right) \Bigg/ \left(\sum_{l=1}^{N} \sum_{m=1}^{N} (\mathbf{C}_{lm})^{-1} \right) \tag{12}$$

where l and m are integers, $\mathbf{C}$ is the local covariance matrix, and $C_{ik} = (I_P - L_p^i)^T (I_P - L_p^k)$. If W_o are determined, I_o can be obtained. However, I_o and W_o are unknown in (10). We assume that there exists an approximate linear mapping between redundant linear combination of faces at factors o and p. We use weights W_p of factor p to generate $\tilde{I}_o$ as follows:

$$\tilde{I}_o = W_p L_o^n \tag{13}$$

where $\tilde{I}_o$ is close to I_o. Equation (13) shows that $\tilde{I}_o$ is the linear combination of the training face images of factor o; therefore, it should be face-like of factor o (see Fig. 3). Therefore, face I_p is transformed from one factor$p$ to multiple factors o.

It is not known how humans identify the relation between two different factors of images [29]. In order to reduce the error, we improve upon (12). W_p will be adjusted to make its variance small so that $\tilde{I}_o$ has more general characteristics of factor o. A compensation matrix is diagonal loaded to the local covariance matrix $\mathbf{C}$:

$$\mathbf{C} = \mathbf{C} + \lambda \mathbf{D} \tag{14}$$

where λ is a constant, and $\boldsymbol{D}$ is an $N \times N$ identity and diagonal matrix. The value of λ is determined empirically.

Since the resolution of 16 × 12 pixels is small, we use a resolution of 64 × 48 to illustrate the superiority of the

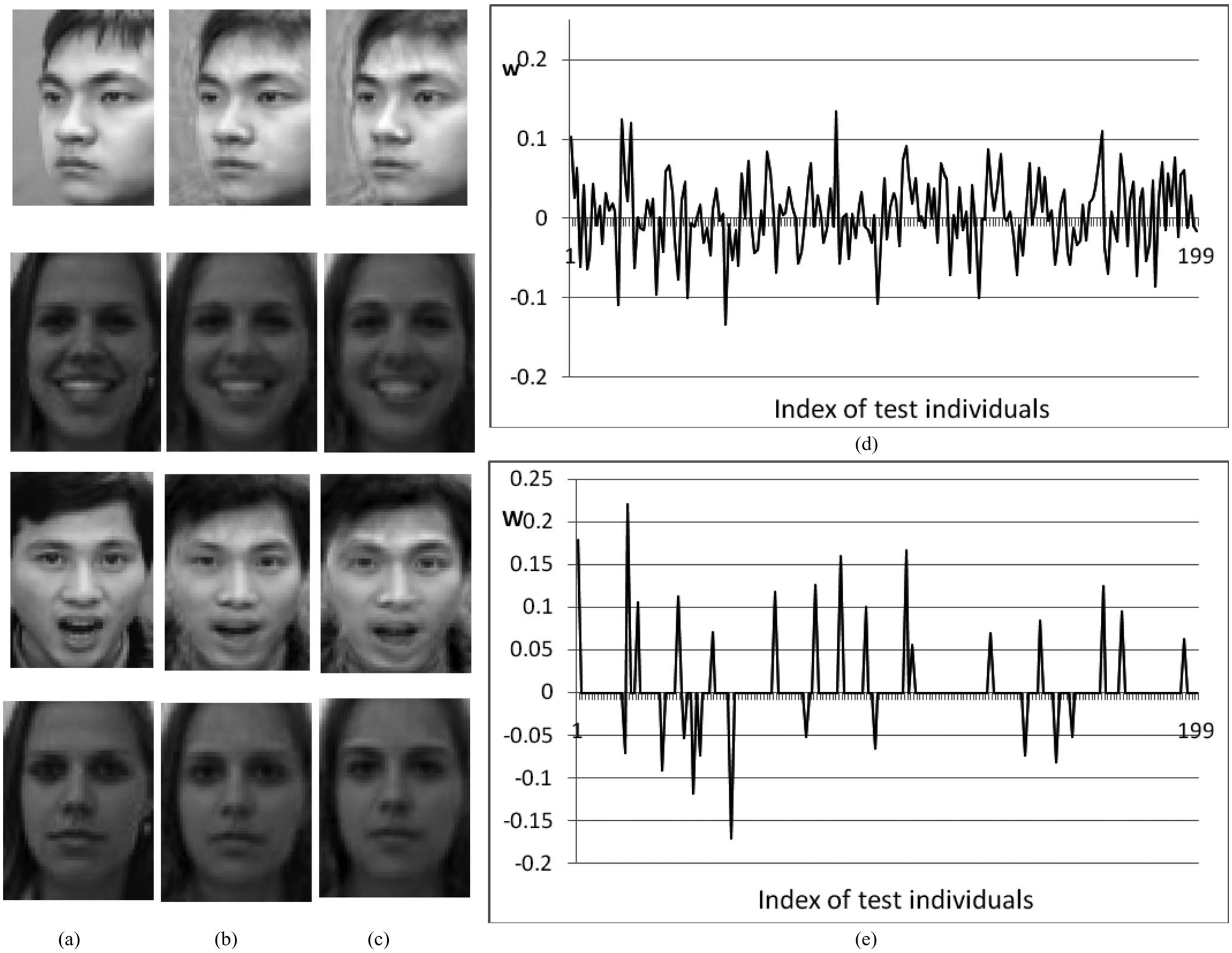


Fig. 5. Face reconstruction in the same resolution space. (a) Input. (b) Proposed redundant linear combination. (c) Proposed sparse linear combination. (d) Redundant weights. (e) Sparse weights.

loading diagonal matrix. The effect of diagonal loading is shown in Fig. 6. The results with and without the loading diagonal matrix appear in Fig. 6(c); the results have been improved with diagonal loading and the image quality is improved when the value of λ increases. However, when λ exceeds a certain value, the individual ingredient of the reconstructed face decreases, and the universal ingredient increases to the average face of factor o. We hope that the results maintain both the general characteristics of factor o and specific characteristics of input and we must balance them. Fig. 6(k) provides the relationship between λ and PNSR values of the results. The value of λ between 800 000 and 1 600 000 are optimal. When the resolution of LR face input is 16×12, the best values of λ are between 50 000 to 100 000.

We compared our method with the redundant and sparse strategies [W_p in (13) is calculated using (7)]. Representative results of HR images appear in Fig. 7, which illustrate that face transformations using the redundant strategy have better image qualities than from the sparse method. Furthermore, the proposed strategy of sparsity takes more computation time. Using compressed sense theory, the exact solution to (7) (for sparsity) is NP hard due to its nature as a combinatorial optimization problem. Suboptimal solutions to this problem can be found by iterative methods or for the design of dictionaries. Because the iterative and dictionary generating steps are involved, it takes much computation time to obtain sparse coefficients.

The computation of the sparse coefficients takes at least several minutes using a PC with four cores 1.7-G CPU, and the redundant coefficients method only takes several seconds. After the coefficients are determined, the rest of computation takes less than a minute despite all nonzero weights or sparse weights

Therefore, we choose the redundant strategy in our proposed framework.

C. Step 1: Generating Multiple Factors in Low-Resolution Space

We next present the proposed framework, which can produce HR faces with new face factors. Our proposed robust framework includes two steps: a global transformation with diagonal loading for modeling the mappings among different new facial factors, and a local position-patch based method with weights compensation for incorporating image details. Because the

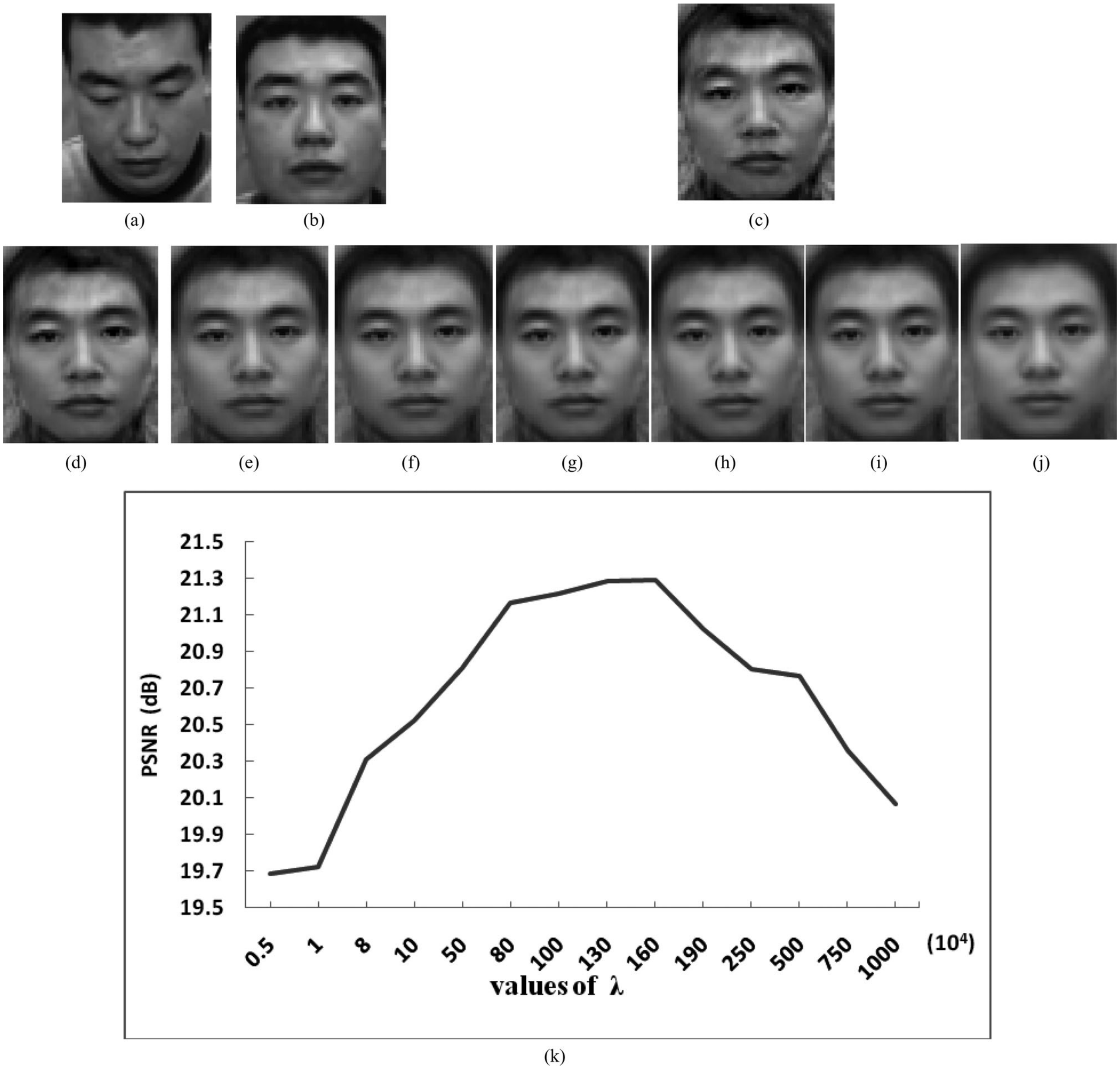


Fig. 6. Effect of loading constant. (a) Input. (b) Ground truth image. (c) Without loading. (d) With loading λ = 5000. (e) λ = 800 000. (f) λ = 1 000 000. (g) λ = 1 500 000. (h) λ = 1 600 000. (i) λ = 2 500 000. (j) λ = 10 000 000. (k) Quantitative data.

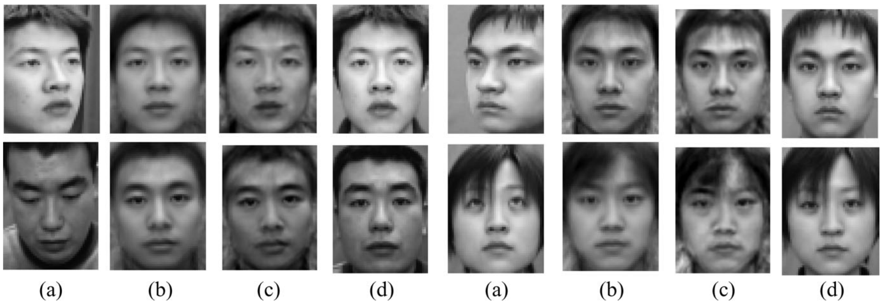


Fig. 7. Face transformation. (a) Input. (b) Redundant. (c) Sparse. (d) Ground truth images.

proposed framework can be used to different expressions, poses, and illuminations, we call it a robust framework.

The proposed method of Section II-B is applied to the LR space. Suppose that I_p is an LR face input of a single factor p. We can use the method of Section II-B to generate corresponding LR faces $\tilde{I}_o$ in LR space.

D. Step 2: Local Patch-Based Method for Incorporating Face Details

The LR face images $\tilde{I}_o$ require a second step to obtain HR correspondences of factor o. In this section, [12] is improved to incorporate image details in the new factor spaces. All training HR face images H_o^n and the corresponding LR training image L_o^n at factor o are respectively divided into $\{L_o^n(i,j)\}$, $\{H_o^n(i,j)\}$. The term (i,j) denotes the position information of each patch. We improve upon [12] by defining the new geometry constraint as follows:

$$\begin{aligned} W_o(i,j) = \ & \underset{W_o^n(i,j)}{\arg\min} \{ \, \| \tilde{I}_o(i,j) - \sum_{n=1}^{N} w_o^n(i,j) L_o^n(i,j) \, \|^2 \\ & + \lambda \sum_{n=1}^{N} \| \, w_o^n(i,j) D_{nn}(i,j) \, \|^2 \ \}, \\ & \text{s.t.} \sum_{n=1}^{N} w_o^n(i,j) = 1 \end{aligned} \tag{15}$$

where $W_o(i,j)$ is an N-dimensional weight vector of each reconstruction weight $W_o^n(i,j)$, for $n = 1, 2, \ldots, N$; λ is a regularization parameter balancing the contribution of the reconstruction error and locality of the solution; and D is an $N \times N$ diagonal matrix. The geometry difference $D_{nn}(i,j)$ of each vector element penalizes the distance between $\tilde{I}_{\text{o}}(i,j)$ and the same position training patches. The geometry difference is determined by the squared Euclidean distance:

$$D_{nn}(i,j) = \left\| \tilde{I}_o(i,j) - L_o^n(i,j) \right\|^2, \quad 1 \le n \le N. \tag{16}$$

$W_o(i,j)$ can be solved by the following formulation [30]:

$$W_{\text{o}}(i,j) = 1/(C(i,j) + \lambda D). \tag{17}$$

Let

$$S = \tilde{I}_{\text{o}}(i,j) \cdot G^T - L_o \tag{18}$$

where G is a column vector of ones, and L_o is a matrix with its columns being the training patches $L_o^n(i,j)$. The local covariance matrix $\mathbf{C}$ can be obtained by

$$C = S^T S. \tag{19}$$

Once the reconstruction weights $W_O(i,j)$ are obtained, the HR image patches $\tilde{H}_o(i,j)$ are generated as follows:

$$\tilde{H}_o(i,j) \simeq \sum_{n=1}^{N} H_o^n(i,j) W_o^n(i,j). \tag{20}$$

All HR patches $\tilde{H}_o(i,j)$ are integrated to form the final global HR image $\tilde{H}_o$ according to their original positions. Through the above steps, the multiple HR faces $\tilde{H}_o$ are generated from a single LR face I_p.

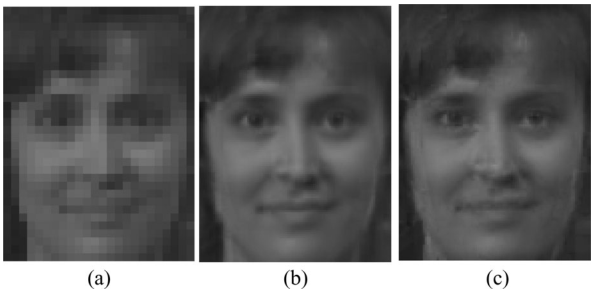

(a) (b) (c)

Fig. 8. Comparison of generating image details. (a) Input LR image 32 × 24. (b) Using redundant weights 128 × 96 [12]. (c) Using sparse weights 128 × 96 [1].

Redundant representation modeling of data provides nonzero weights for all training atoms. Sparse representation modeling of data describes signals as linear combinations of only a few atoms from a predefined dictionary. Thus, some information is lost in sparse signal representation because the signal is approximated with a smaller subset from the dictionary. In addition, when selecting only a small number of training atoms for obtaining image details, most of training atoms that contain certain information are not used. We propose that redundant reconstruction has superiority over sparse reconstruction in superresolution, because the prior information from image training set should be used as much as possible to hallucinate facial details.

Redundant weights are used in [12] to reconstruct the HR patch and sparse weights in [1]. Fig. 8 illustrates that the method using redundant weights yields higher quality and more details than sparse weights. Because the solution for sparse weights has a large computation cost, redundant representation is superior.

Jung *et al.* [31] replaced the redundant weights with the sparse weights in [12]. In this paper, the same experiments under the same experimental conditions were performed on the same database. The experimental conditions from [31] were replicated. The results are shown in Fig. 9.

Jung *et al.* [31] claimed that their method was more effective in preserving the edge and image details in the nose and mouth areas than [12]. Fig. 9 shows that the two results are very similar. The peak signal-to-noise ratio (PSNR) values of face hallucination results appear in Fig. 10. Most PSNR values of face hallucination results of [12] are higher than in [31]. The experimental results [31] need to be questioned.

Therefore, it is not necessary to take the additional step for sparse representation in step two of our framework.

III. Evaluation

We evaluated our framework on public face databases: the CAS-PEAL-R1 Face Database [38] for multiview simulated experiments, the benchmark AR face database [32] for multiple expressions, and the CMU PIE database [33] for multiple illuminations. All experimental face images should be aligned using an automatic tool or manually.

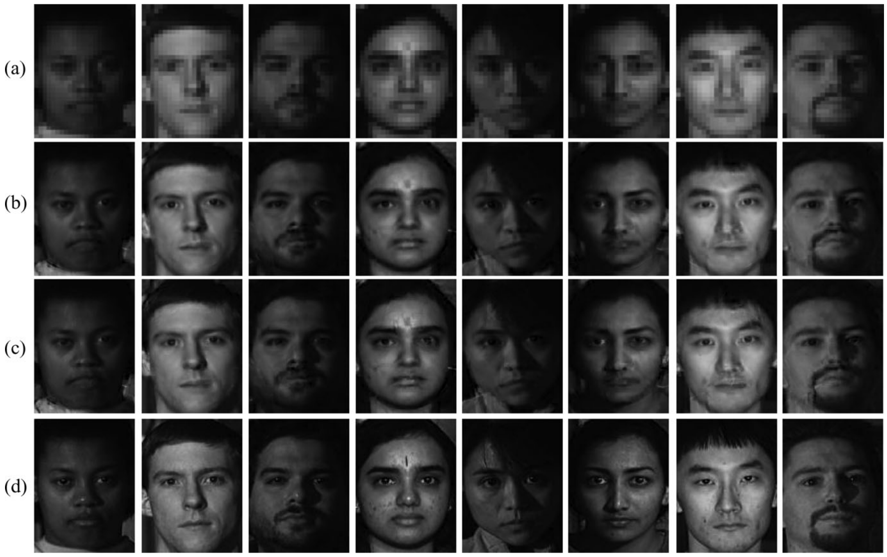


Fig. 9. Face hallucination results. (a) Input 32 × 24 LR faces. (b) Results of [12]. (c) Results of [31]. (d) Original 128 × 96 faces.

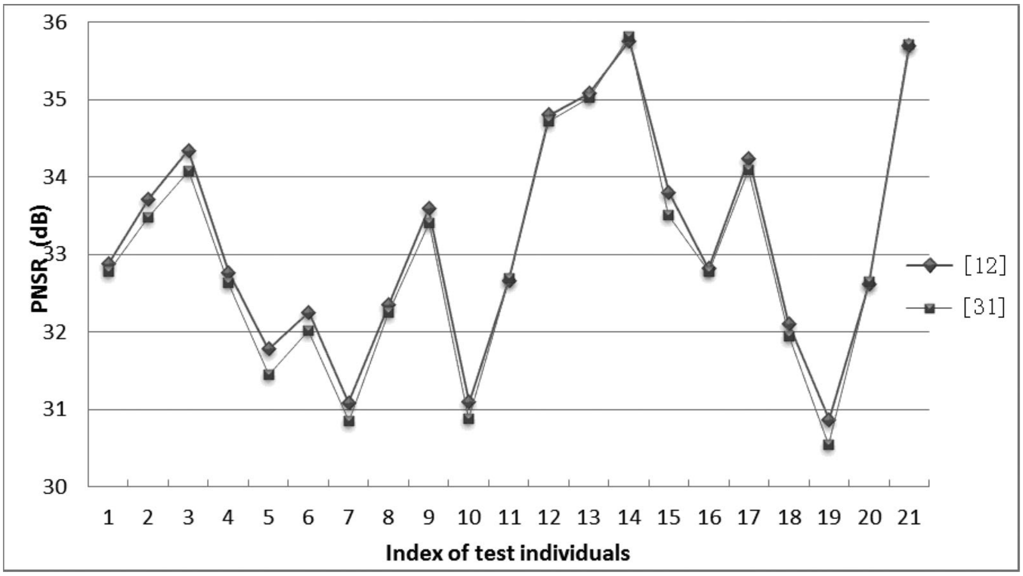


Fig. 10. PSNR values of face hallucination results.

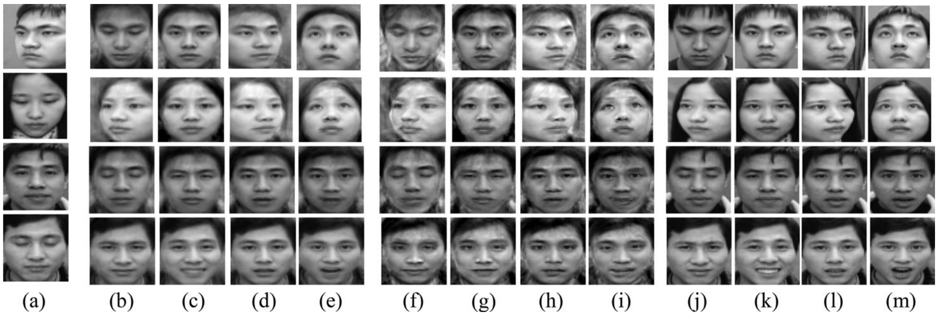


Fig. 11. Comparison of face transformation (CAS-PEAL-R1 database). (a) Input image with different poses and expressions. (b)–(e) Our method. (f)–(i) Jia's method [26]. (j)–(m) Ground truth image.

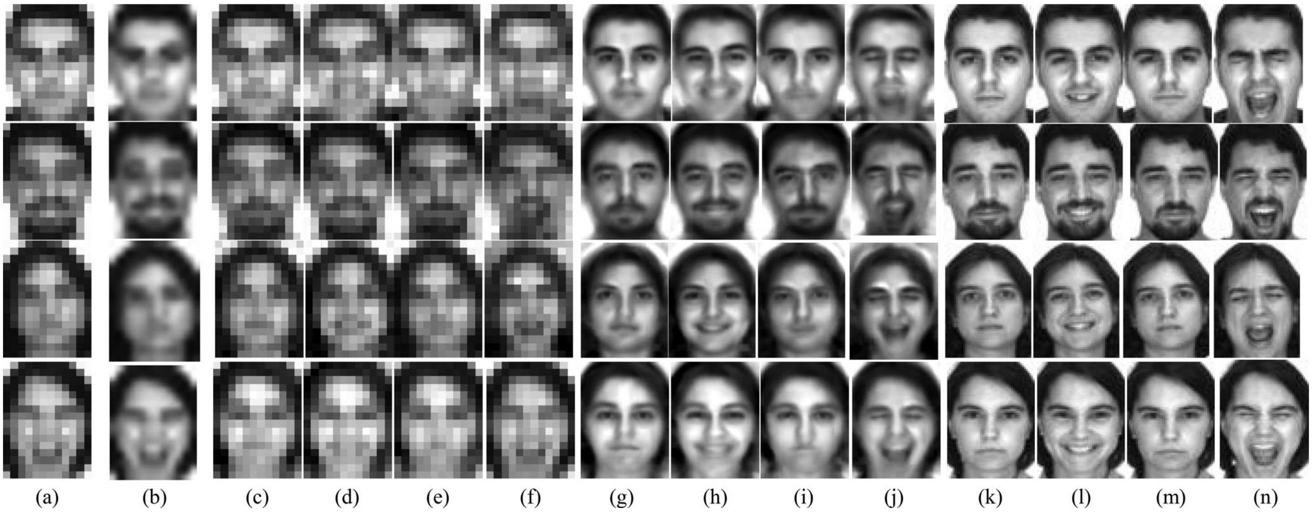


Fig. 12. Multiple facial expression hallucination (AR database). (a) LR (16 × 12) input faces at a single expression. (b) Bicubic interpolation. (c)–(f) Face transformation in LR space using our framework. (g)–(j) HR (64 × 48) results of our framework .(k)–(n).Ground truth face images.

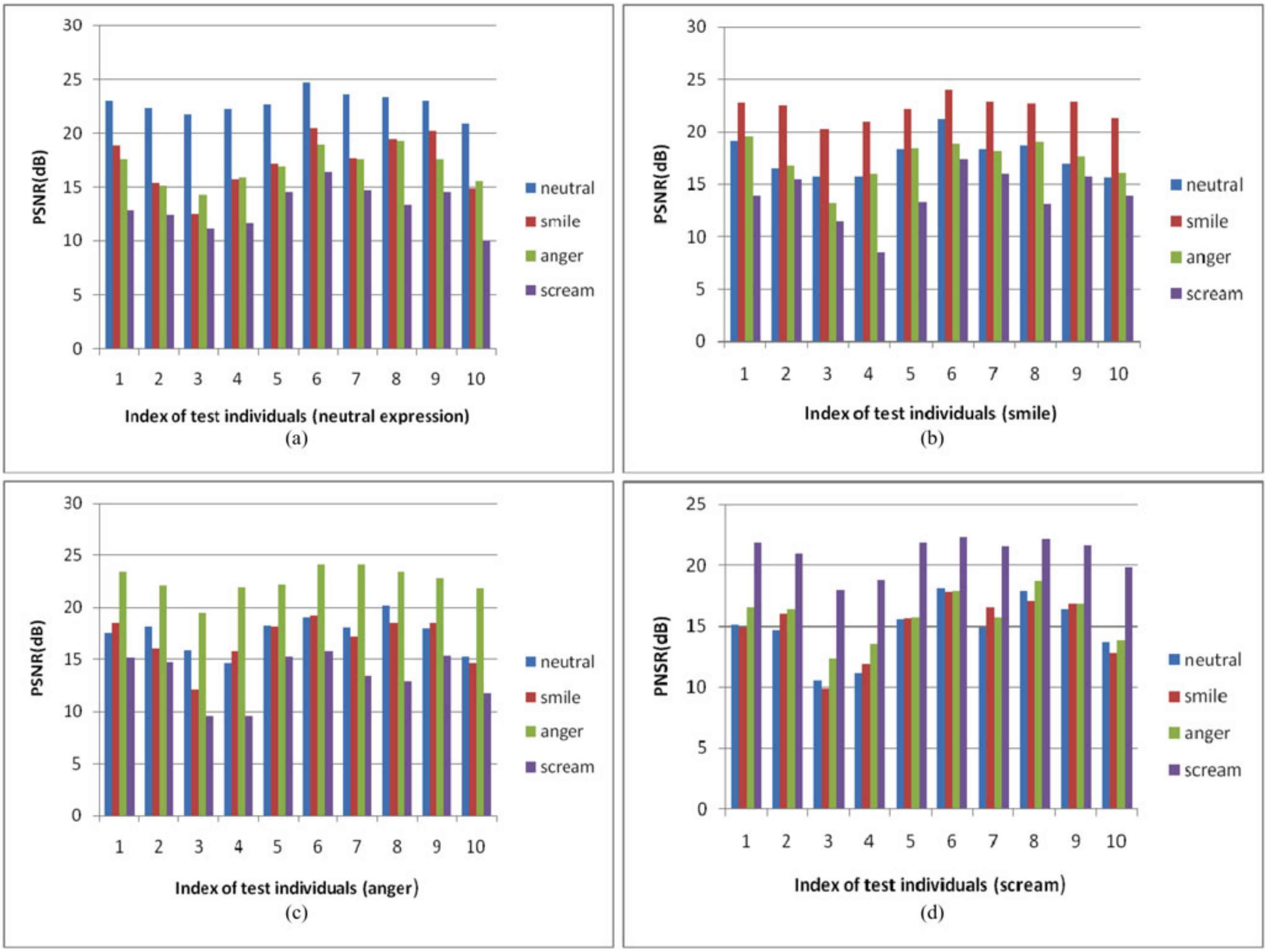


Fig. 13. PSNR values of the hallucinated results of four different expressions from each test expression of ten individuals. (a) LR test inputs with neutral expression. (b) LR test inputs are with smile expression. (c) LR test inputs with anger expression. (d) LR test inputs with scream expression.

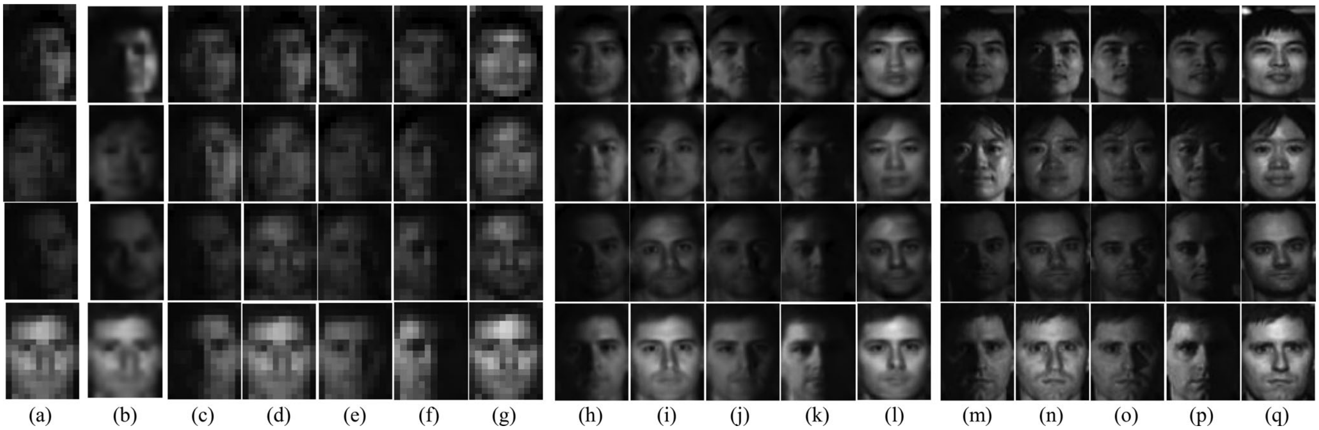


Fig. 14. Multiple facial illumination condition hallucination (CMU PIE database). (a) LR (16 × 12) input faces at a single illumination. (b) Bicubic interpolation. (c)–(g) Face transformation of our method in LR space. (h)–(l) HR (64 × 48) results of our framework. (m)–(q) Ground truth face images.

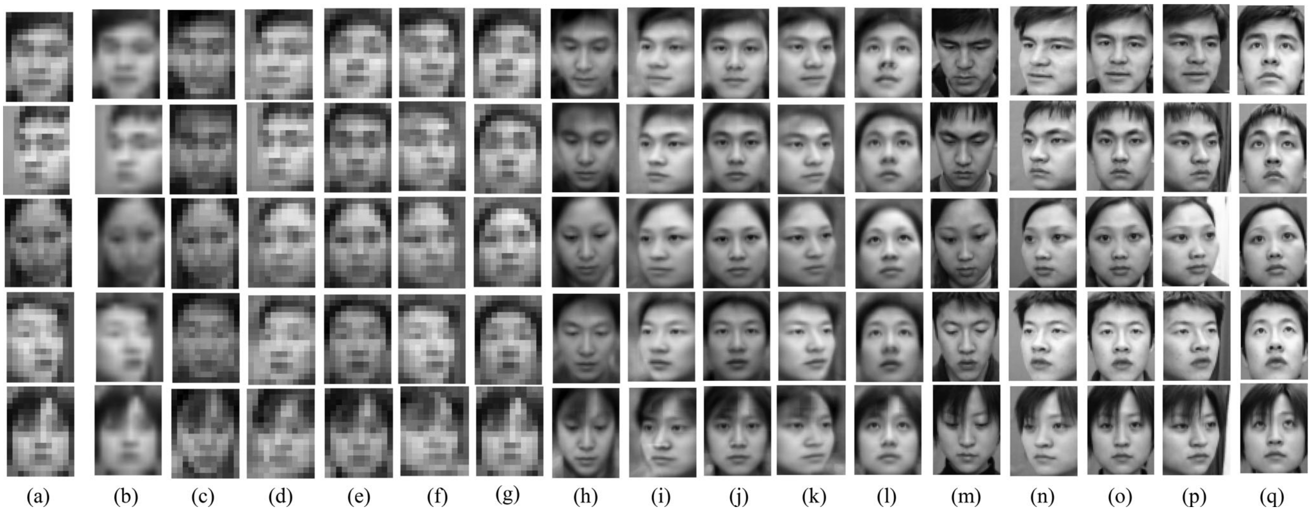


Fig. 15. Multiview face hallucination (CAS-PEAL-R1 database). (a) LR (16 × 12) input faces at single view. (b) Bicubic interpolation. (c)–(g) Face transformation of our method. (h)–(l) HR (64 × 48) results of our method. (m)–(q) Ground truth face images.

A. Comparison of Global Transformation With Traditional Hierarchical Tensor Framework

We compared face transformation of different poses and expressions with the method in [26] on the CAS-PEAL-R1 Face Database [38], which contains 1350 face images of 270 different individuals. Two hundred and fifty individuals were chosen at random. Each individual has five different views and five different expressions. We cut out the interesting region of the faces and unified the images to the size of 64 × 48 pixels. The 1250 images of 250 individuals were used as HR training images, and the rest of the images of the 20 individuals were used as image inputs. Some input faces were transformed from one pose to multiple poses, e.g., from nonfront to front, and some were transformed from one expression to multiple expressions. Representative results are shown in Fig. 11. The average PSNR value for our method is 20.6 dB and for the method in [26] is 18.1 dB. The proposed framework achieves better image qualities and higher PSNR values than in [26]. Our method is superior than that in [26] with respect to face transformation.

B. Multiple Facial Expression Hallucination

Our approach was applied to the AR face database [32] for multiple facial expression hallucination. The original AR dataset consists of 126 people, and for each individual, it includes images of different facial expressions, illumination conditions, and occlusions. The setup was adopted from [26]. We obtained four HR results with different expressions from every LR input. Representative results are in Fig. 12. The PSNR values of the hallucinated results are in Fig. 13. If the hallucinated results have the same expression with LR inputs, they will obtain the highest PSNR values compared with other expressions. Our method can produce multiple HR faces with four expressions only given a single LR face.

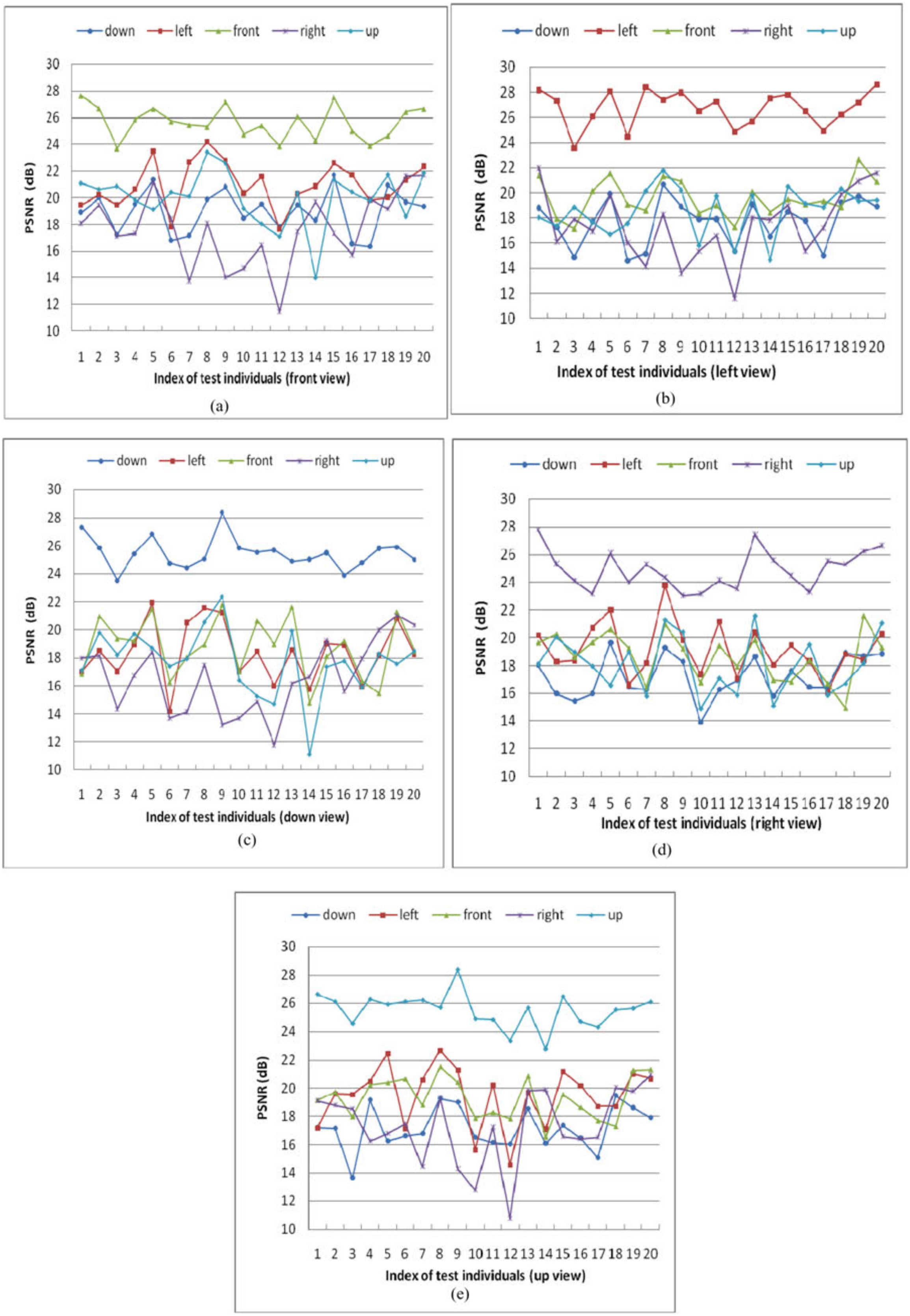


Fig. 16. PSNR values of the hallucinated results of five different views from each test view of 20 individuals (better shown in electronic version). (a) LR test inputs with front view. (b) LR test inputs with left view. (c) LR test inputs with down view. (d) LR test inputs with right view. (e) LR test inputs with up view.

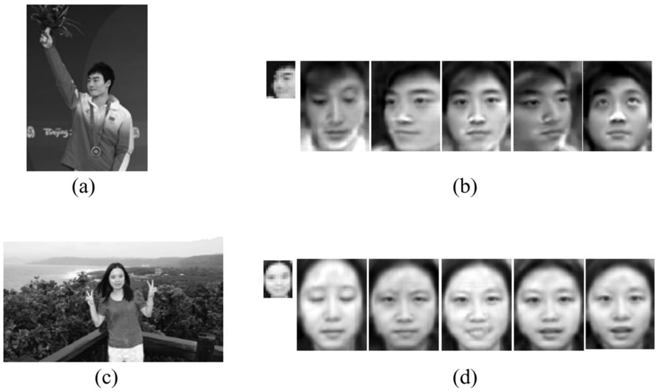


Fig. 17. Experimental results on real-world pictures. (a) Picture from the Internet. (b) Multiview experiment. (c) Picture from a camera. (d) Multiexpression experiment.

C. Multiple Facial Illumination Condition Hallucination

We also evaluated our approach on the CMU PIE database [33] for multiple illumination hallucination. The database contains 41 368 images obtained from 68 subjects. We chose the frontal face images of all 68 individuals, in which each individual has five different illumination conditions. We manually aligned these face images and established a standard training dataset and used the "leave-one-out" methodology to perform the multiple facial illumination hallucination experiments. Representative results are shown in Fig. 14. Our framework can be applied to LR input with different illumination conditions.

D. Multiview Face Hallucination

Our multiview hallucination approach was evaluated on the CAS-PEAL-R1 Face Database [38]. The 1250 images of 250 different individuals were used as HR training images, which were smoothed and downsampled to 16 $\times$ 12 as LR training images. The images of the remaining 20 individuals were also smoothed and downsampled to 16 $\times$ 12 as LR image inputs. The view of face input was called input view. For hallucination of one view, we chose the HR face of target view and LR face of input view from training set to build our training pairs. Finally, we obtained the five facial view hallucinations for each test view input. Representative results are shown in Fig. 15.

We can see that any hallucination of LR input with the same view is always better than those with other views, which is due to generating nonlinear variations across different facial views.

The performance was also quantified by evaluating PSNR between the ground truth face images and the multiview hallucinated images. The PSNR values from the hallucinated results of multiple factors is given in Fig. 16.

The results with the same views of the inputs have relatively higher values of PSNR than the others, which appear on the tops of the wave in Fig. 16(a)–(d). The results with the distinct views of the inputs have relatively lower values of PSNR than the others, which appear on the bottoms of the wave in Fig. 16(a)–(d). The computational time of our face transform method is around 60 s when working on a 16 $\times$ 12 LR face using a PC with four cores, 1.7-G CPU, while the first step of [26] already about 300 seconds. Thus, our method has an advantage of efficiency.

E. Real-World Images

Two real-world images are used in our evaluation. We chose the CAS-PEAL-R1 Face Database [38] as the training set. The LR faces in Fig. 17(a) and (c) were aligned, extracted manually, and standardized to the size of 16 $\times$ 12 as LR inputs. Multiple facial expression hallucination and multiview face hallucination experiments are conducted using the two pictures as inputs. We show the results in Fig.17 (b) and (d). It shows that our framework is applicable to the real-world pictures.

The image qualities are not as good as the results from the standard face database. The reason for this is that the LR image is from the real-world environment and contains noise. Because sometimes we do not know anything about the LR input beforehand in real-world applications, superresolution should ensure all original information of the input is not lost. If we know beforehand that the LR input has noise which is useless information, we will remove the noise using a method before face hallucination. Otherwise, we may remove useful information as noise, e.g., a scar or other personal characteristics, which are useful for recognition.

IV. Conclusion

Most face hallucination methods are limited to the frontal face without consideration of illumination, head pose, and expression variations. A few methods consider face variations, but they are limited in that the HR output and LR input have the same view. In this paper, we proposed a robust framework of face superresolution across multiple factors. Specifically, we propose a redundant transformation with diagonal loading for modeling the mappings among different new face factors, and a local reconstruction with geometry and position constraints for incorporating image details in the new factor spaces. While a complex tensor model is traditionally used, the experiment results illustrate that our model is superior to the traditional hierarchical tensor framework. Compared with the existing method, our framework is effective with regard to the computational cost. The comparison of our two proposed redundant and sparse strategies are also discussed. It is not necessary to adopt sparse representation in the proposed framework. The experimental results demonstrate that the proposed framework offers robustness when dealing with the inputs that have different expressions, head poses, and illuminations compared to the state-of-the-art methods, can generate HR face images with better image qualities than the hierarchical tensor based method, and improves the state of the art from single one output to multiple outputs with new factors.

References

[1] X. Ma, H. Q. Luong, W. Philips, H. Song, and H. Cui, "Sparse representation and position prior based face hallucination upon classified over-complete dictionaries," *Signal Process.*, vol. 92, no. 9, pp. 2066–2074, 2012.

[2] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning, "A multi-frame image super-resolution method," *Signal Process.*, vol. 90, pp. 405–414, 2010.

[3] K. I. Kim and Y. Kown, “Single-image super-resolution using sparse regression and natural image prior,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 127–1133, Jun. 2010.

[4] S. Baker and T. Kanade, “Hallucinating faces,” *Proc. Int. Conf. Autom. Face Gesture Recognit.*, Grenoble, France, 2000, pp. 83–88.

[5] C. Liu, H. Shum, and C. Zhang, “A two-step approach to hallucinating faces: Global parametric model and local nonparametric model,” *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2001, pp. 192–198.

[6] H. Chang, D. Y. Yeung, and Y. Xiong. “Super-resolution through neighbor embedding,” *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, Washington, DC, USA, 2004, pp. 1275–1282.

[7] X. Wang and X. Tang, “Hallucinating face by eigentransformation,” *IEEE Trans. Syst. Man Cybernetics C, Appl. Rev.*, vol. 35, no. 3, pp. 425–434, Aug. 2005.

[8] Y. Zhuang, J. Zhang, and F. Wu, “Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation,” *Pattern Recognit.*, vol. 40, pp. 3178–3194, 2007.

[9] X. Ma, J. P. Zhang, and C. Qi, “Hallucinating faces: Global linear modal based super-resolution and position based residue compensation,” *Lecture Notes in Comput. Sci.*, vol. 5716, pp. 835–843, 2009.

[10] J. S. Park and S. W. Lee, “An example-based face hallucination method for single-frame, low-resolution facial images,” *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1806–1816, Oct. 2008.

[11] H. Huang, H. T He, X. Fan, and J. P. Zhang, “Super-resolution of human face image using canonical correlation analysis,” *Pattern Recognit.*, vol. 43, no. 7, pp. 2532–2543, 2010.

[12] X. Ma, J. P. Zhang, and C. Qi, “Hallucinating face by position-patch,” *Pattern Recognit.*, vol. 43, no. 6, pp. 2224–2236, 2010.

[13] T. Shan, B. Lovell, and S. Chen, “Face recognition robust to head pose from one sample image,” in *Proc. Int. Conf. Pattern Recognit.*, 2006, pp. 515–518.

[14] V. Blanz and T. Vetter, “Face recognition based on fitting 3d morphable model,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.

[15] V. Blanz and T. Vetter, “Face recognition based on frontal views generated from non-frontal images,” in *Proc. IEEE Int. Conf. Computer Vision Pattern Recognit.*, 2005, pp. 454–461.

[16] C. Tian and G. L. Fan, “Multiview face recognition: From TensorFace to V-TensorFace and K-TensorFace,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 320–331, Apr. 2012.

[17] M. A. O. Vasilesescu and D. Terzopoulos. “Multilinear image analysis for facial recognition,” in *Proc. Int. Conf. Pattern Recognit.*, 2002, pp. 511–514.

[18] M. A. O. Vasilescu and D. Terzopoulos. “Multilinear analysis of image ensembles: TensorFaces,” in *Proc. Eur. Conf. Comput. Vision*, 2002, pp. 447–460.

[19] Y. Li and X. Y. Lin, “Face Hallucination with pose variation,” in *Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2004, pp. 723–728.

[20] X. Chai, S. Shan, X. Chen, and W. Gao, “Locally linear regression for pose-invariant face recognition,” *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1716–1725, Jul. 2007.

[21] T. H. Fang, X. Zhao, O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, “3D/4D facial expression analysis: An advanced annotated face model approach,” *Image Vision Comput*., vol. 30, no. 10, pp. 738–749, 2012.

[22] W. Gu, X. Cheng, Y. V. Venkatesh, D. Huang, and H. Lin, “Facial expression recognition using radial encoding of local Gabor features and classifier synthesis,” *Pattern Recognit.*, vol. 45, pp. 80–91, 2012.

[23] P. Vageeswaran, K. Mitra, and R. Chellappa, “Blur and illumination robust face recognition via set-theoretic characterization,” *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1362–1372, Apr. 2013.

[24] K. Jia and S. G. Gong. “Multi-modal tensor face for simultaneous super-resolution and recognition,” in *Proc. IEEE Int. Conf. Comput. Vision*, 2005, pp. 1683–1690.

[25] K. Jia and S. G. Gong. “Multi-modal Face image super-resolution in Tensor space,” in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2005, pp. 264–269.

[26] K. Jia and S. G. Gong. “Generalized face super-resolution,” *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 873–886, Jun. 2008.

[27] J. C. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-Resolution via sparse representation,” *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[28] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 22, no. 12, pp. 2323–2326, 2000.

[29] H. S. Lee and D. J. Kim, “Generating frontal view face image for pose invariant face recognition,” *Pattern Recognit. Lett.*, vol. 27, pp. 747–754, 2006.

[30] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, “Locality-constrained linear coding for image classification,” in *Proc. IEEE Conf. Computer Vision Pattern Recognit.*, 2010, pp. 3360–3367.

[31] C. Jung, L. Jiao, B. Liu, and M. Gong, “position-patch based face hallucination using convex optimization,” *IEEE Signal Process. Lett.*, vol. 18, no. 6, pp. 367–370, Jun. 2011.

[32] A. M. Martinez and R. Benavente, “The AR face database,” CVC Tech. Report #24, 1998.

[33] H. A. Rowley, S. Baluja, and T. Kanade, “Neural network-based face detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 23–38, Jan. 1998.

[34] W. Zhang and W.-K. Cham, “Hallucinating face in the DCT domain,” *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2769–2779, Nov. 2011.

[35] Y. Hu, K.-M. Lam, G. Qiu, and T. Shen, “From local pixel structure to global image super-resolution: A new face hallucination framework,” *IEEE Trans. Image Process.*, vol. 20, no. 2, pp. 433–445, Feb. 2011.

[36] H. Li, L. Xu, G. Liu, and M. Gong, “Face hallucination via similarity constraints,” *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 19–22, Jan. 2013.

[37] X. M. Qian, X. S. Hua, P. Chen, and L. J. Ke, “PLBP: An effective local binary patterns texture descriptor with pyramid representation,” *Pattern Recognit.*, vol. 44, pp. 2502–2515, 2011.

[38] W. Gao, B. Cao, S. G. Shan, X. L. Chen, D. L. Zhou, X. H. Zhang and D. B. Zhao, “The CAS-PEAL large-scale Chinese face database and baseline evaluations,” *IEEE Trans. System Man, Cybernetics A, Syst. Humans*, vol. 38, no. 1, pp. 149–161, Jan. 2008.

[39] T. Vetter, “Synthesis of novel views from a single face image,” *Int. J. Comput. Vision*, vol. 28, no. 2, pp. 103–116, 1998.

Xiang Ma received the B.S. degree from North China Electric Power University, Beijing, China, in 1999, the Ph.D. degree from Xi’an Jiaotong University, Xi’an, China, in 2011.

He is currently an Associate Professor with the School of Information Engineering, Chang’an University, Xi’an. His research interests include face image processing and image processing in machine and intelligent transportation.

Huansheng Song received the B.S. and M.S. degrees in communication and electronic systems and the Ph.D. degree in information and communication engineering from Xi’an JiaoTong University, Xi’an, China, in 1985, 1988, and 1996, respectively.

Since 2004, he has been with the Information Engineering Institute, Chang’an University, Xi’an, where he became a Professor in 2006. In 2012, he was nominated as the Dean of the Information Engineering Institute. His current research interests include image processing and recognition and intelligent transportation systems.

Xueming Qian (M’10) received the B.S. and M.S. degrees from the Xi’an University of Technology, Xi’an, China, in 1999 and 2004, respectively, and the Ph.D. degree from the School of Electronics and Information Engineering, Xi’an Jiaotong University, Xi’an, in 2008.

He is currently a full Professor with the School of Electronics and Information Engineering, Xi’an Jiaotong University. He is the Director of the SMILES LAB. His research interests include social media big data mining and search. His research has been supported by the National Natural Science Foundation of China, Microsoft Research, and MOST.