Face Super-Resolution via Multilayer Locality-Constrained Iterative Neighbor Embedding and Intermediate Dictionary Learning

Junjun Jiang, Ruimin Hu, *Senior Member, IEEE*, Zhongyuan Wang, *Member, IEEE*, and Zhen Han

***Abstract*— Based on the assumption that low-resolution (LR) and high-resolution (HR) manifolds are locally isometric, the neighbor embedding super-resolution algorithms try to preserve the geometry (reconstruction weights) of the LR space for the reconstructed HR space, but neglect the geometry of the original HR space. Due to the degradation process of the LR image (e.g., noisy, blurred, and down-sampled), the neighborhood relationship of the LR space cannot reflect the truth. To this end, this paper proposes a coarse-to-fine face super-resolution approach via a multilayer locality-constrained iterative neighbor embedding technique, which intends to represent the input LR patch while preserving the geometry of original HR space. In particular, we iteratively update the LR patch representation and the estimated HR patch, and meanwhile an intermediate dictionary learning scheme is employed to bridge the LR manifold and original HR manifold. The proposed method can faithfully capture the intrinsic image degradation shift and enhance the consistency between the reconstructed HR manifold and the original HR manifold. Experiments with application to face super-resolution on the CAS-PEAL-R1 database and real-world images demonstrate the power of the proposed algorithm.**

***Index Terms*— Face super-resolution, face hallucination, manifold learning, neighbor embedding, dictionary learning.**

I. Introduction

THE past twenty years have witnessed a rapid progress in face recognition techniques. Due to the limitations of surveillance system, such as network bandwidth limitation, server storage, and long distance to the interest object, the query face images captured by surveillance camera are very Low-Resolution (LR). Based on the current technical level, the information revealed by LR face images is so limited that the accuracy of the resulting face recognition is very low. Thus the LR problem forms one of the most challenging issues in face recognition [1]. Recently, face super-resolution (or *face hallucination*) techniques have been employed to address the LR problems of imaging system. It can generate an High-Resolution (HR) face image from either a sequence of LR face images (by multi-frame reconstruction based super-resolution approaches) or a single frame LR face image (by learning-based super-resolution approaches), thus providing more facial details for the following recognition process. In this paper, we focus on the learning-based face super-resolution method for its superiority over the multi-frame approaches, especially when the magnification factor is large [2], [3].

Generally, the common idea of learning-based methods is to infer an HR image by training the relationship between LR and HR image pairs with the assistance of training samples. The pioneering work on hallucinating face image was done by Baker and Kanade [4] who proposed to infer the HR face from an input LR one under a Bayesian formulation. Liu et al. [5] proposed to integrate a global parametric model and a local nonparametric model for face super-resolution. Since then, a number of different learning-based face super-resolution methods and models have been introduced. They super-resolve the LR face either by globally using holistic face image or locally using patches or pixels of face [6]. A comprehensive overview of face super-resolution is given in [7].

Some classical face models, such as Principal Component Analysis (PCA) [8]–[10], Locality Preserving Projections (LPP) [11], Non-negative Matrix Factorization (NMF) [12] and Canonical Correlation Analysis (CCA) [13] have been used to globally super-resolve face image. They can well capture the global face appearance variations. However, these methods are highly dependent on the training set and fail to render effectively the fine individual details of an input LR face, especially when the input LR face is very different from the training faces or when the training set size is small. To alleviate the above problems, techniques of decomposing a complete image into smaller patches have been introduced recently [12]–[21].

For those local patch based methods, the common idea is first to represent the input LR patch with the LR training patches, and then transform the reconstruction weights to faithfully represent each target (unknown) HR patch with the corresponding HR training patches. Their basic assumption is that LR and HR patch manifolds are locally isometric.

Manuscript received November 5, 2013; revised March 2, 2014 and May 19, 2014; accepted July 30, 2014. Date of publication August 12, 2014; date of current version August 29, 2014. This work was supported in part by the National Key Technologies Research and Development Program under Grant 2013AA014602, in part by the National Natural Science Foundation of China under Grant 61231015, Grant 61172173, and Grant 61303114, in part by the Major Science and Technology Innovation Plan of Hubei Province under Grant 2013AAA020, in part by the Guangdong-Hongkong Key Domain Breakthrough Project of China under Grant 2012A090200007, in part by the China Post-Doctoral Science Foundation under Project 2013M530350, in part by the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20130141120024, and in part by the Key Technology Research and Development Program, Wuhan, China, under Grant 2013030409020109, in part by Guangdong-Hongkong Key Domain Breakthrough Project of China under Grant 2012A090200007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David Frakes.

The authors are with the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan 430072, China (e-mail: jiangjunjun@whu.edu.cn; hrm1964@163.com; wzy_hope@163.com; hanzhen_2003@hotmail.com).

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/TIP.2014.2347201

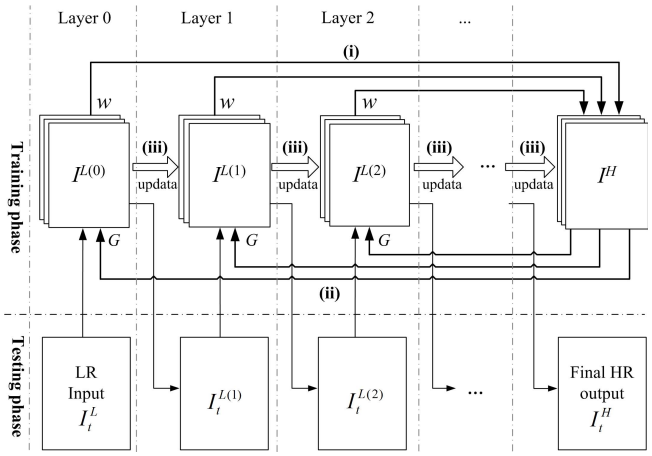


Fig. 1. Flowchart of the proposed framework. (i) Obtaining the optimal reconstruction weights w; (ii) Embedding the geometry G of the original HR space for the super-resolved HR image; (iii) Learning the intermediate dictionaries, $I^{L(1)}, I^{L(2)}, \ldots, I^{L(B)}$. Note that the symbol "$G$" represents the neighborhood relationship in the HR training set, which is characterized by the the distance between image patches (for more details, please refer to Section III-A).

However, due to the *one-to-many* mapping between the LR image and HR ones, the manifold assumption is not true in practice. The geometry of the degraded LR space cannot reflect the true neighborhood relationship in the original HR space.

A. Motivation and Contribution

To address the above issue, in this paper, after considering the combined problem of the HR space geometry preservation and iterative neighbor embedding, we propose a *coarse-to-fine* framework for face super-resolution via *multilayer Locality-constrained Iterative Neighbor Embedding* (multilayer LINE for short) and *intermediate dictionary learning*. The term "coarse-to-fine" has twofold implications: the estimated HR image refining step by step and the LR training set updating. Fig. 1 is the flowchart of the proposed framework. In the training phase, we update the LR training set to obtain the intermediate dictionaries, $I^{L(1)}, I^{L(2)}, \ldots, I^{L(B)}$, given the original LR and HR training sets, $I^{L(0)}$ and I^{H}. In the testing phase, the input LR image is super-resolved by iteratively generating its representation w in the LR space and the geometry G of the estimated HR patch in the HR space.

The contribution of the paper is highlighted in the following:

- With the local consistency assumption, many conventional methods try to preserve the geometry of the LR space for super-resolution, yet only the LR space was considered and the original HR space was neglected. In contrast, the proposed method iteratively uses the geometry of the original HR space unaffected by the image degradation process to regularize the reconstruction weights in the LR space.
- Instead of modeling the relationship between LR images and corresponding HR images in one single layer (the LR training set is fixed), we propose to update the LR training set by constructing intermediate dictionaries and modeling the relationship in much more consistent LR and HR spaces.

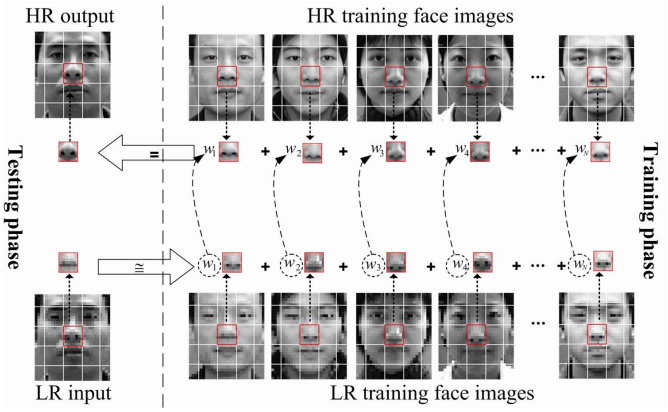


Fig. 2. Flow diagram of the position-patch based face super-resolution framework. To obtain the optimal reconstruction weights is the key issue.

Built upon our preliminary work reported in [22], this paper is an improved version with the following contributions: (i) a multilayer LINE ensemble to improve the performance of our original LINE model; (ii) analysis on the different parameter settings of our proposed approach; (iii) extensive experimental evaluations on its performance, especially on real-world images.

The remainder of the paper is organized as follows. In Section II, we first give some notations, then introduce some related local patch representation approaches. We present the proposed method in Section III and conduct the experiments on simulated LR image and real-world image in Section IV. Finally, we conclude this paper in Section V.

II. Related Work

A. Notations

Given an input LR image I_t^L (subscript "t" distinguishes the test sample from the training set), the goal of face super-resolution is to construct its HR version I_t^H by learning the relationship between the LR and HR training sets, $I^L = \{I_1^L, I_2^L, \cdots, I_N^L\}$ and $I^H = \{I_1^H, I_2^H, \cdots, I_N^H\}$, where N is the size of training set. As for a local patch based method, we divide the input LR image into M patches, $\{x_t(p,q)|1 \le p \le U, 1 \le q \le V\}$, according to the predefined patch size and overlap pixels, where U represents the patch number in vertical axis and V represents the patch number in horizontal axis, thus we have $M = UV$. Similarly, we divide the LR and HR training face images into M patches, $\{x_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$ and $\{y_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$, respectively. $x_i(p,q)$ $(y_i(p,q))$ denotes a small patch at the position (p,q) of the i-th training sample in the LR (HR) training set. For more details about the dividing strategy, please refer to our previous work [20] and [21].

B. Local Patch Representation Approaches

For the input LR patch $x_t(p,q)$ located at position (p,q), e.g., the nose patch as shown in Fig. 2, local patch based method tries to obtain the optimal reconstruction weights in

the LR space,

$$J(w(p,q)) = ||x_t(p,q) - \sum_{i=1}^{N} w_i(p,q)x_i(p,q)||_2^2 + \tau\Omega(w(p,q)). \quad (1)$$

The first term in (1) represents the data-fidelity, and the second term represents the desired properties of the reconstruction weights $w(p,q)$. Here, the regularization parameter τ represents the trade-off between the closeness to the data and the prior of the reconstruction weights. Acquiring the optimal reconstruction weights

$$w^*(p,q) = \arg\min_{w(p,q)} J(w(p,q)),$$

the target HR patch $y_t(p,q)$ can be estimated by the HR training patches of the same position:

$$y_t(p,q) = \sum_{i=1}^{N} w_i^*(p,q)y_i(p,q). \quad (2)$$

Once we have obtained all the super-resolved HR patches $\{y_t(p,q)|1 \le p \le U, 1 \le q \le V\}$, the final HR image I_t^H can be generated by averaging pixel values in the overlapping regions according to the original position. When it does not lead to a misunderstanding, we drop the term (p,q) for convenience from now on. From the process presented above, we learn that the key issue of these local patch based methods is to obtain the optimal reconstruction weights w^* as shown in Fig. 2.

The representative work for local patch based super-resolution was proposed by Chang et al. [14]. Assuming that the HR patches and LR patches share the same geometry structure, they utilized Locally Linear Embedding (LLE) [23] to estimate an HR patch by linearly combining K candidate HR patches in the HR training set. It does not require a large number of samples and achieves very competitive performance. In [19], Zhang and Cham further proposed a K-Nearest Neighbors (K-NN) embedding based method for inferring local features of face image in the Discrete Cosine Transform (DCT) space. Hu et al. [18] proposed a local structure learning based face super-resolution method. This method assumes that two similar face images share similar local pixel structures so that each pixel could be generated by a linear combination of its neighbor pixels weighted by coefficients. Li et al. [15] claimed that there is a common hidden manifold between the LR and HR manifold spaces and proposed a manifold alignment based face super-resolution method.

Inspired by the face analysis research, which states that the position information is very important for face analysis and synthesis, Ma et al. [16] proposed a position-patch based face super-resolution method by performing collaboratively over the whole training face patches of the same position given an LR input patch. However, when the number of training samples is larger than the dimensions of the observed patch, the solution is not unique. To overcome this, sparse regularization [12], [17], [24] was introduced for the objective function of patch reconstruction to obtain the optimal reconstruction weights for face super-resolution, i.e., $\Omega(w) = ||w||_1$.

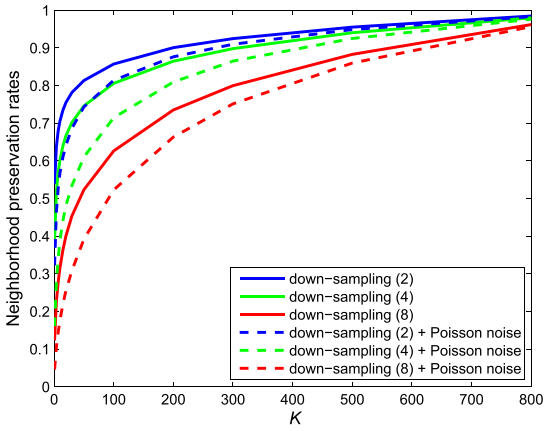


Fig. 3. (Best viewed in colors and magnification). The neighborhood preservation rates between the LR degradation image manifold and the original HR image manifold according to different neighbor number K. Note that the number in the parentheses of the legend denotes the down-sampling factor of the image degradation process.

The sparsity constraint can not only ensure that the under-determined equation have an exact solution but also reduce reconstruction error. However, it overemphasizes the sparsity and neglects the locality constraint, which is more important than sparsity in revealing the non-linear manifold structure of data [25], [26]. Thus, the obtained patch representation based sparsity constraint is far from the best. In order to solve this problem, in our previous work [20] and [21] we further improved the position-patch method by incorporating the local geometry constraint of manifold instead of the sparsity constraint into the patch representation objective function:

$$J(w) = ||x_t - \sum_{i=1}^{N} w_i x_i||_2^2 + \tau||dist \odot w||_2^2, \quad (3)$$

where "$\odot$" denotes the element-wise multiplication, and $dist$ is an N-dimensional locality adaptor that gives different freedom for each LR training patch x_i to its similarity to the input LR patch x_t. It is simply determined by the squared Euclidean distance between the input LR patch and the LR training patches:

$$dist_i = ||x_t - x_i||_2, \quad 1 \le i \le N. \quad (4)$$

By assigning different freedom to the training samples, e.g., those patches close to the input LR patch will be preferably selected while those patches distant from the input LR patch should be penalized severely, it will achieve sparsity and locality simultaneously [20], [21].

Remark 1: Whether the constraint is collaborative [16], sparse [12], [17], [24] or local [20], [21], all the methods mentioned above are based on the basic assumption that the LR manifold and the HR one share the same local geometry structure. Although the local geometry preserving assumption is taken for granted, it may not be held in practice. This is mainly due to the *one-to-many* mapping between the LR image and HR ones. As shown in Fig. 3, the neighborhood preservation rates [27] decrease with the increase of down-sampling factor or by adding Poisson noise. For example, when $K = 20$, the neighborhood preservation rate falls sharply

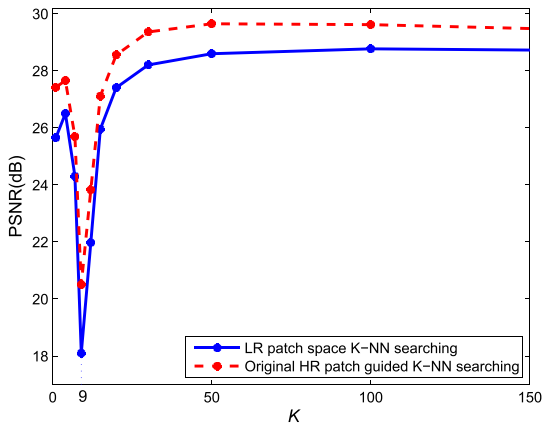


Fig. 4. The performance of two different K-NN searching and embedding based super-resolution methods according to different neighbor number K. Blue plot: the traditional K-NN searching in the LR space; red plot: the original HR patch guided K-NN searching in the HR space, which can be seen as the ideal case.

from 76% of "down-sampling (2)" to 25% of "down-sampling (8) + Poisson noise".

Remark 2: All methods mentioned above obtain the reconstruction weights and the target HR patch separately. They only consider the LR manifold and do not fully utilize and exploit the geometry structure information of the HR manifold, which is unaffected by image degradation process [28]. In order to prove the usefulness of the geometry of the original HR space, we assess the effect of different manifold geometry preservation approaches, i.e., the LR manifold geometry preservation versus the original HR manifold geometry preservation (can be seen as the ideal case where the target HR image is known in advance). From the performance comparisons shown in Fig. 4, it can be seen that the preservation scheme is vitally important to those local patch based neighbor embedding algorithms and the traditional LR patch guided preservation method has a comparatively huge rise space. Note that the fall around 9 is because that the least squares solution of neighbor embedding is *too fitted* on the LR data [29].

III. The Proposed Coarse-to-Fine Method

In light of the above discussion, we attempt to address the following two problems: (i) how to preserve the HR manifold geometry while representing the input LR patch; (ii) how to mitigate the inconsistence between the LR and HR manifolds.

For the first problem, we advance an iterative neighbor embedding strategy to iteratively update the LR patch representation and the estimated HR patch (Section III-A). Specifically, we calculate the locality adaptor in the HR space, i.e., the distance between the estimated HR patch and each HR patch in the HR training set. And then, the reconstruction weights and the newly estimated HR patch are calculated iteratively. In this coarse-to-fine manner, we can expect a good HR patch estimation.

To address the second problem, we propose to update the LR training set (Section III-B) to gradually reduce the the inconsistence between the LR and HR manifolds. In this paper we hypothesize the existence of a virtual path which smoothly connects the LR image and corresponding HR one. Intuitively, face images which gradually transform from the LR image to the HR one will form a smooth transition path in the image space. As shown in Fig. 5, constructing some intermediate dictionaries along the transition path allows us to more likely capture the underlying image degradation process. Therefore, we can expect to gradually refine the estimated HR face with the intermediate dictionaries.

A. Locality-Constrained Iterative Neighbor Embedding (LINE)

Rather than calculating the locality adaptor in the LR space and seek a package solution to construct the target HR patch once and for all as in [20] and [21], we obtain the locality adaptor in the HR space and iteratively update the LR patch representation and the estimated HR patch via the following bivariate objective function:

$$J(w, y_t) = ||x_t - \sum_{k \in C_K(y_t)} w_k x_k||_2^2 + \tau ||dist|_K \odot w||_2^2, \quad (5)$$

where $C_K(y_t)$ denotes the indexes of the K-NN of y_t in the HR training set, thus

$$C_K(y_t) = \text{support}(dist|_K), \quad (6)$$

where $dist|_K$ refers to the smallest K entries of $dist$, and $dist \in R^N$ is the measurement of distances between y_t and patches in the HR training set $\{y_i\}_{i=1}^N$:

$$dist_i = ||y_t - y_i||_2, \quad 1 \le i \le N. \quad (7)$$

To solve Eq. (5), for the s-th block, we first set $y_{t(s)}$ to the Bicubic interpolation version of the input LR patch and then update the reconstruction weights $w_{(s)}$ with $y_{t(s)}$ by minimizing Eq. (5) as a constrained least squares problem, which can be solved by an analytical solution [26].

Upon acquiring the optimal LR patch reconstruction weights w^*, we can construct the corresponding HR patch with the same weights through:

$$y_t = \sum_{k \in C_K(y_t)} w_k^* y_k. \quad (8)$$

According to the position, the patches are processed in raster-scan order in the image, from left to right and top to bottom. Following [12], we enforce compatibility between adjacent patches (the values in the overlapped regions are simply averaged). The entire face super-resolution process is summarized as Algorithm 1. Note that the *patch_size* and *overlap* denote the pixel number of one square patch and the overlap pixel number between patches respectively.

B. Multilayer LINE Model

In LINE, by incorporating the geometry of HR manifold and introducing the iterative scheme, it obtains the optimal reconstruction weights from the LR manifold in nature. As given in Eq. (5), the optimal reconstruction weights are constructed by minimizing the reconstruction error of x_t with the LR training set, whose geometry is inconsistent with that of the HR one. Therefore, the performance of LINE will be inevitably affected by the degraded LR training samples.

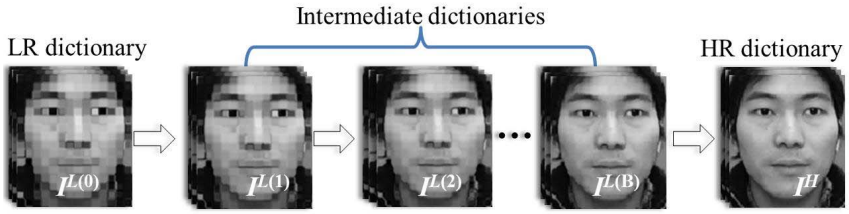


Fig. 5. Given LR and HR training sets, $I^{L(0)}$ and I^H, our procedure learns a set of intermediate dictionaries, $I^{L(1)}, I^{L(2)}, \ldots, I^{L(B)}$, to capture the intrinsic image degradation process between LR and HR images and to enhance the consistency between the updated LR manifold and the original HR manifold.

Algorithm 1 Face Super-Resolution via Locality-Constrained Iterative Neighbor Embedding (LINE)

Input: LR and HR training sets, $I^L = \{I_1^L, I_2^L, \cdots, I_N^L\}$ and $I^H = \{I_1^H, I_2^H, \cdots, I_N^H\}$, and an input LR face image I_t^L. The parameters: $patch\_size$, $overlap$, τ, K, and $maxIter$.
Compute U and V:
$U = \text{ceil}((imrow - overlap)/(patch\_size - overlap))$
$V = \text{ceil}((imcol - overlap)/(patch\_size - overlap))$
Divide each of the LR and HR training images and the input LR image into M small patches according to the same location of face, $\{x_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$, $\{y_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$ and $\{x_t(p,q)|1 \le p \le U, 1 \le q \le V\}$, respectively.
Initialize: $s = 0$,$I_{t(0)}^H = Bicubic(I_t^L)$.
for $p = 1 : U$ **do**
  for $q = 1 : V$ **do**
    repeat
      $dist_{i(s)}(p,q) = ||y_{t(s)}(p,q) - y_i(p,q)||_2, \quad 1 \le i \le N$.
      $C_K(y_{t(s)}(p,q)) = \text{support}(dist_{i(s)}(p,q)|_K)$.
      Calculate $w_{(s)}^*(p,q)$ according to Eq. (5) at a specific iteration number s.
      $y_{t(s+1)}(p,q) = \sum_{k \in C_K(y_{t(s)}(p,q))} w_{k(s)}^*(p,q) y_k(p,q)$.
      $s = s + 1$.
    until $s > maxIter$
  end for
end for
Integrating all the above obtained HR patches $\{y_t(p,q)|1 \le p \le U, 1 \le q \le V\}$ according to the position (p,q). The final HR image can be generated by averaging pixel values in the overlapping regions.
Output: HR super-resolved face image I_t^H.

To this end, we extend the LINE model to multilayer LINE one, in which the LR training set is updated constantly. Therefore, the objective function of our proposed multilayer LINE approach is expressed as:

$$J(w^b, y_t^b, X^b) = ||x_t^b - \sum_{k \in C_K(y_t^b)} w_k^b x_k^b||_2^2 + \tau||dist^b|_K \odot w^b||_2^2, \ b = 0, 1, \cdots, B. \quad (9)$$

Here, b denotes the layer index and B is the number of layer, as shown in Fig. 5. When $b = 0$, x_t^b is the input LR patch $x_t^0 = x_t$, $X^0 = [x_1, x_2, \cdots, x_N]$ is the original LR training patch set. With the updated LR training set and the original HR training set, the super-resolution reconstruction can be performed in much more consistent LR and HR spaces, giving rise to improved performance compared with traditional neighbor embedding approaches.

Solving the joint optimization problem of Eq. (9) seems to be very challenging. And thus, we look for a way around it. Once we fix the LR training set X^b, the optimization of w^b and y_t^b will reduce to Eq. (5). Therefore, the only remaining problem of solving Eq. (9) is how to learn the intermediate dictionaries.

To obtain the intermediate dictionaries, at each layer, we utilize the LINE method presented in Section III-A to update all the LR training faces by a so-called *leave-one-out* strategy. For each LR training image $I_i^{L(b)}$ at the b-th layer, a new LR training set of size $N-1$ is composed by all other LR training samples, $\hat{I}^{L(b)} = \{I_j^{L(b)}|j = 1, 2, \cdots, i-1, i+1, \cdots, N\}$. Similarly, the corresponding newly obtained HR training set is composed by all the HR training samples except for I_i^H, $\hat{I}^H = \{I_j^H|j = 1, 2, \cdots, i-1, i+1, \cdots, N\}$. The target HR image $I_i^{*H(b)}$ can be constructed by Algorithm 1 with the input LR image, $I_i^{L(b)}$, and the new LR and HR training sets, $\hat{I}^{L(b)}$ and $\hat{I}^H$. Algorithm 2 summarizes the entire intermediate dictionary learning process.

C. Computational Complexity

In this subsection, we discuss the computational complexity of the proposed multilayer LINE algorithm. Because the intermediate dictionary learning process can be performed off-line and completed once for all, therefore, we only discuss the computational complexity of the on-line super-resolution reconstruction process involved in the proposed method. As illustrated in Algorithm 1, the proposed super-resolution approach takes major cost on two parts: (i) K-NN searching; and (ii) reconstruction weights calculating. We learn that there are four major factors determining the computational cost in stages (i) and (ii): the number of nearest neighbors K, the training set size N, the patch number in one image M, and the dimension of one patch p^2.

According to [30], it takes $O(Kp^2N)$ for the K-NN searching stage and $O(K^3p^2)$ for the reconstruction weights calculating stage. Therefore, it costs $O(Kp^2N+K^3p^2)$ at each iteration in Eq. (5). In addition, the maximum iteration times $maxIter$ in each layer and the total layer number B affect the practical running time. Therefore, the computational cost of

Algorithm 2 Learning Intermediate Dictionaries

Input: LR and HR training sets: $I^L = \{I_1^L, I_2^L, \cdots, I_N^L\}$ and $I^H = \{I_1^H, I_2^H, \cdots, I_N^H\}$. The parameters: τ, K, and B.
Initialize: $b = 0$, $I^{L(0)} = I^L$.
repeat
  for the i-th LR training image $I_i^{L(b)}$ **do**
    Obtain a new LR training set composed by all other LR training samples except for $I_i^{L(b)}$: $\hat{I}^{L(b)} = \{I_j^{L(b)} | j = 1, 2, \cdots, i-1, i+1, \cdots, N\}$.
    Obtain a new HR training set composed by all other HR training samples except for I_i^H: $\hat{I}^H = \{I_j^H | j = 1, 2, \cdots, i-1, i+1, \cdots, N\}$.
    Infer the target HR image $I_i^{*H(b)}$ according to Algorithm 1 with $\hat{I}^{L(b)}$, $\hat{I}^H$ and $I_i^{L(b)}$.
    $I^{L(b+1)}(:, i) = I_i^{*H(b)}$.
  end for
  $b = b + 1$.
until $b > B$
Output: The learned intermediate dictionaries, $I^{L(1)}, I^{L(2)}, \cdots, I^{L(B)}$.

the proposed method is about $O(maxIter(Kp^2N + K^3p^2)B)$. We will further compare the proposed method with the state-of-the-art methods in term of the CPU time in Section IV-C.

IV. Experiments

In this section, we describe the details of extensive experiments performed to evaluate the usefulness of the proposed method to face super-resolution. The experiments are designed to answer the following questions:

- Is it necessary to iterate the process of geometry generation in the HR space and weights reconstruction in the LR space? Can the iterative process converge?
- Is it necessary to update the LR dictionary to obtain the intermediate dictionaries? Can the iterative process converge?
- How does the proposed approach compare against other state-of-the-art methods?
- Does the proposed approach succeed in enhancing real-world images?

Matlab code available upon e-mail request (junjun0595@163.com or jiangjunjun@whu.edu.cn).

A. Database Description

Experiments described in this paper are conducted on the public face database, the CAS-PEAL-R1 database [31], with 30871 images of 1040 subjects. We only use the neutral expression and normal illumination face of each subject from the frontal subset for experiments. In all the 1040 frontal face images, we randomly select 1000 images for training and leave the other 40 images for testing. Therefore, all the test subjects are not present in the training set. All the images are aligned by the two points of eye center and cropped to 128×112 pixels (as shown in Fig. 6). The LR images are formed by smoothing (an averaging filter of size 4×4) and down-sampling (by a factor of 4) the corresponding HR images, thus the size of LR face images are 32×28 pixels.

Fig. 6. Some training faces in the CAS-PEAL-R1 database.

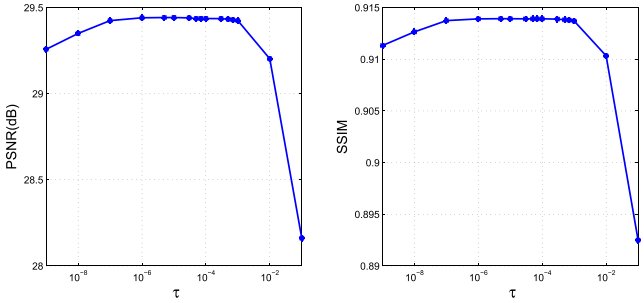


Fig. 7. The performance of different values of τ in terms of PSNR (left) and SSIM (right).

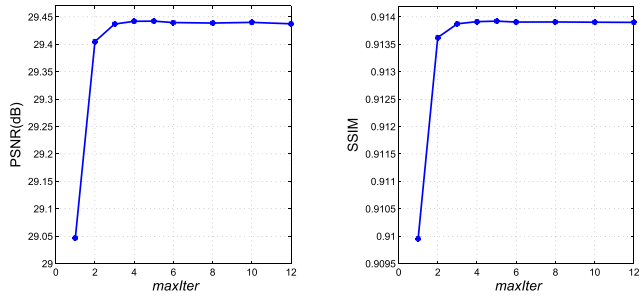


Fig. 8. The performance of different values of $maxIter$ in terms of PSNR (left) and SSIM (right).

B. Parameter Analysis

In this subsection, we investigate the effect of the locality regularization parameter τ, the maximum iteration number $maxIter$ and the layer number B of the proposed method.

1) The Influence of the Locality Regularization Parameter: In order to test the effect of the locality regularization parameter τ, we evaluate the performance of LINE with different τ. Fig. 7 is the plot of the average PSNR and Structural SIMilarity (SSIM)[1] [32] values of all 40 test face images according to different values of τ. We can see that the locality regularization parameter has big implications for the performance of the proposed method: as τ increases, the gain of the proposed method becomes larger, which

[1]The SSIM index is a method for measuring the similarity between two images, which have proven to be inconsistent with human eye perception. The higher the SSIM value, the better is the face super-resolution quality. The maximum value of SSIM is 1, which means a perfect reconstruction.

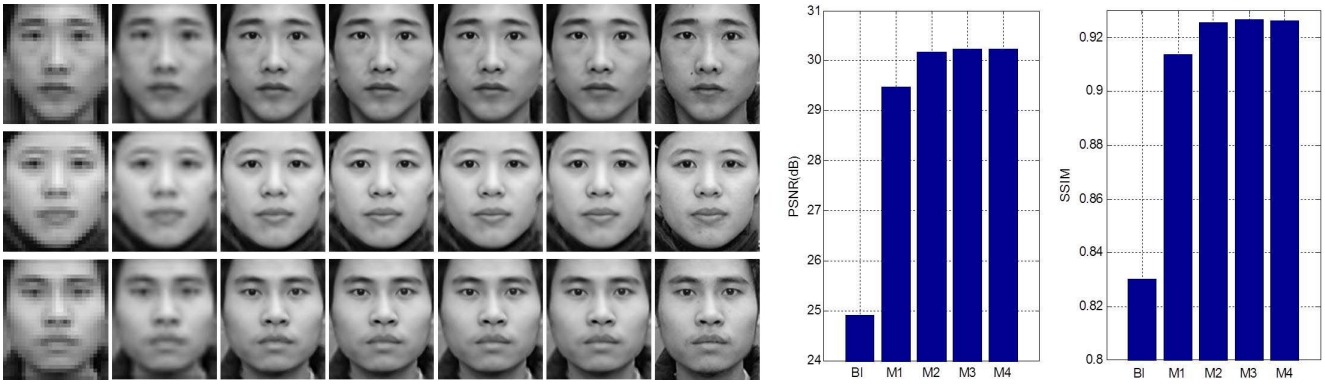


Fig. 9. Face super-resolution results with different layer numbers. The left subfigures are three groups of super-resolved face images, from left to right: the input LR face images, BI's results, M1's results, M2's results, M3's results, M4's results and the original HR face images; the right plots are the objective results in terms of PSNR and SSIM respectively, BI (PSNR = 24.92 dB, SSIM = 0.8301), M1 (PSNR = 29.48 dB, SSIM = 0.9138), M2 (PSNR = 30.16 dB, SSIM = 0.9256), M3 (PSNR = 30.22 dB, SSIM = 0.9265), M4 (PSNR = 30.23 dB, SSIM = 0.9265).

TABLE I
NSR EVALUATIONS AT DIFFERENT ITERATION

iteration	1	2	3	4	5
NSR	83.33%	94.82%	95.05%	95.07%	95.07%

implies the importance of the locality in representing the image patch. When τ is between 1e-7 and 1e-3, LINE can continually obtain a stable performance. For balancing the reconstruction error, the parameter τ cannot be set too high. Therefore, we set τ to 1e-5 for the LINE.

2) The Influence of Iteration Number: In order to test the effect of the iteration number, we give the performance of LINE with different $maxIter$. As shown in Fig. 8, we plot the average PSNR and SSIM values of all 40 test face images according to the iteration number, and find that: (i) as the iteration number increases, the gain of the proposed method becomes larger, which implies that the HR space geometry preservation strategy is crucial to the neighbor embedding based super-resolution approaches; (ii) the proposed method can be fast-converging with only several steps, i.e., the iteration number is set to 4 for all experiments, which implies a potential application in practice.

We additionally examine the component of $C_K(y_t)$ at each iteration. Specifically, we define the Neighborhood Similar Rate (NSR for short) as follows:

$$NSR = \frac{|C_K(y_t) \cap C_K(y^*)|}{|C_K(y^*)|}, \tag{10}$$

where y_t and y^* denote the estimated HR patch and the ideal HR patch of the input LR patch x_t respectively. Here, $|\cdot|$ is the cardinality operator. Table I tabulates the variability of NSR at each iteration. After few iteration, e.g., 4, the support $C_K(y_t)$ has no major changes, approximating the ideal case (only 4.93% disparity).

3) The Influence of Layer Number: Fig. 9 shows the subjective results (due to space limitation, we only give three groups of super-resolved face images out of the 40 test face images) and objective results with different layer numbers. The Bicubic Interpolation (BI) is taken to be the baseline for comparison. Note that "M1" denotes the single-layer LINE method, "M2" denotes the two layer LINE method, and so on. From the subjective results, we learn that BI gives the smooth results and blurs the fine details while our proposed method can recover most of the facial details. With the increase of layer number, the face contours and texture become much clearer. In the caption of Fig. 9, we list the average PSNR and SSIM values of all the 40 test face images in parentheses following each method. From the right plots, we can see that the improvement of our proposed method over BI is remarkable, e.g., 4.56 dB in term of PSNR and 0.0837 in term of SSIM respectively. With the increase of layer number, the gain of the proposed method is becoming more marked. The PSNR and SSIM improvements of multilayer LINE model over single-layer LINE model are 0.74 dB and 0.0127 respectively. It must be realized that the proposed method will reach a stable performance when the layer number is larger than three, and this also shows its quick convergence. All the above results verify the effectiveness of the LR training set updating scheme.

C. Comparisons With the State-of-the-Art Methods

We have compared our methods with one global face super-resolution method, Wang et al.'s global face method [8], and three other local patch based face super-resolution methods, e.g., Chang et al.'s Neighbor Embedding (NE) method [14], Yang et al.'s Sparse Representation (SR) method [12], and Zhang et al.'s DCT space based method [19]. To pursue the best performance, we tune the parameters for all comparative methods to achieve their best possible results. Specifically, for Wang et al.'s global face method [8], we let the variance accumulation contribution rate of PCA be 99% (around 250 bases). The number of neighbors in NE [14] is set to 50. For the SR method [12], we use the parameter settings as in the source code provided by the authors. For DCT method [19], the number of neighbors is set to 200 (we change the value and find 200 is the best choice). As for our proposed method, the number of nearest neighbors K for neighbor embedding is experimentally set to 150 in the first layer. Note that as the layer b increases (and so does the quality of the target patches), we use higher value for K, since more atoms are required to

Fig. 10. Comparisons of super-resolved results based on different methods on the CAS-PEAL-R1 face database. From left to right are the Input LR faces, the results of Wang et al. [8], NE [14], SR [12], DCT [19], and our method, and the original HR faces. (Note that the effect is more pronounced if the figure of the electronic version is zoomed.)

represent the fine details at the output of the higher level. As for these local patch based methods (NE [14], SR [12], DCT [19], and our method), we recommend to use the size of 12×12 pixels for the HR patch and the overlap between neighbor patches is 4 pixels as in [20] and [21]. Note that more overlap pixels will lead to better performance, for fair comparison, we choose the same patch size and overlap pixels for all local patch based methods.

1) Subjective Results Comparisons: Fig. 10 shows some samples of the super-resolved results generated by different methods. The first column are the input LR faces, the last column are the ground true HR faces, and second to sixth columns are the super-resolved HR faces based on five different methods. From these visual results of super-resolved faces, we learn that the proposed method generates the most competitive results with more facial details in the nose, mouth and face contour than the others and meanwhile are much more similar to the original HR faces. Note that the results of Wang et al.'s global face method [8] have severe "ghosting" effects, especially on locations around face contours and margins of the mouth. NE method [14] may construct some "facial details" that don't exist in the original HR face due to

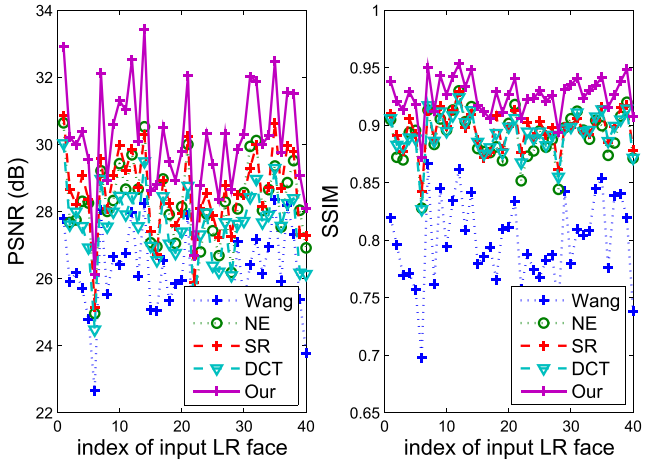


Fig. 11. (Best viewed in colors and magnification). Plots of PSNR and SSIM values for all the 40 test face images of different methods. Wang et al. [8] (PSNR = 26.04 dB, SSIM = 0.7989), NE [14] (PSNR = 27.76 dB, SSIM = 0.8841), SR [12] (PSNR = 28.56 dB, SSIM = 0.8962), DCT [19] (PSNR = 27.59 dB, SSIM = 0.8929), Our method (PSNR = 30.22 dB, SSIM = 0.9265).

under or over-fitting problem of fixed K neighbor embedding. SR method [12] and DCT method [19] are very competitive, but the results are either too smooth or with some noticeable artifacts inevitably in the eyes and around face contours.

2) Objective Results Comparisons: To further assess the objective quality of different methods, we give the quantitative comparisons in terms of PSNR and SSIM as in Fig. 11. It can be seen from the plots that the proposed method gives the best performance for all the test face images. In the caption of Fig. 11, we give the average PSNR and SSIM values of different methods on the 40 test face image. We can learn that: (i) local patch based methods substantially surpass Wang et al's global face method [8], and this can be attributed to their strong expressive power because of decomposing a complete image into smaller patches; (ii) by incorporating the position information of face, position-patch based methods (SR [12] and our method) can perform better than NE method [14], which takes no consideration of the position semantics priors; (iii) the gain of our method over the second best method, e.g., SR method [12], is remarkable (1.66 dB in term of PSNR and 0.0303 in term of SSIM). We attribute this superiority to the iterative neighbor embedding of LINE and the intermediate dictionary learning.

3) Computational Complexity Comparisons: We further compare the proposed method with the other state-of-the-art methods in term of the CPU time. Fig. 12 shows the average CPU time spent on all the 40 test images performed using Matlab 7.14 (R2012a) on an Intel Xeon 8 CPU with 2.13 GHz and 16G memory PC at Windows platform. Due to the coarse-to-fine estimation scheme, the proposed super-resolution method requires much more time to gradually refine the result on reconstruction process in comparing with the methods of Wang et al. [8], NE [14], SR [12], and DCT [19]. Wang et al.'s global face super-resolution approach is the fastest for its simpleness, which requires only a few steps of matrix multiplications and additions. While these local patch

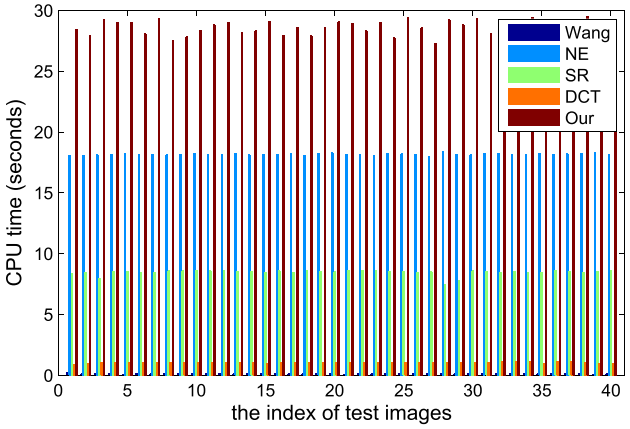


Fig. 12. (Best viewed in colors and magnification). Comparisons of CPU time between the Wang et al. [8], NE [14], SR [12], DCT [19] and the proposed method. Their average CPU times are 0.1249 seconds, 18.18 seconds, 8.480 seconds, 1.051 seconds and 28.62 seconds respectively.

representation based approaches have demonstrated successful results, they come at the price of additional complexity, often accompanied by higher computational cost. Thanks to the independence of the reconstruction of each target HR patch, we can straightforward to accelerate the algorithm via parallel computation.

D. Experiments on Real-World Images

The input LR face images of the above experiments are formed by smoothing and down-sampling the original HR face images, which cannot represent the true feature relationship between the HR image and the degraded LR one in real-world [33]. In an actual condition, it is too difficult for us to simulate the image degradation process or know how different types of image degradation processes affect the structure and statistics of one image.

In order to further testify the effectiveness of the proposed method, we conduct experiments on the real-world images as shown in the first row of Fig. 13. We first use the automatic face detection algorithm [34] to detect the faces in the captured picture, and then utilize the detected two points of eye centers to align the interested face images. Since humans are more sensitive to illuminance changes, we only perform super-resolution reconstruction in the luminance component. We therefore interpolate the color layers (Cb, Cr) using plain Bicubic interpolator. The super-resolved results for a collection of test images are shown in the second row to the fifth row. For each example, the extracted, aligned LR face is at the left, and the HR super-resolved face at the right. From the second row to the fourth row (corresponding to the face marked by green box in the picture), we can see that our approach is able to produce very reasonable results even though the test images are drastically different from the training samples (see Fig. 6). However, we also find that the eyes of the super-resolved faces have some artifacts. We attribute this to the inconsistence between test images and training samples, e.g., the eyes of persons in the test picture are squint while they are open in the training sample (see Fig. 6). The last row (corresponding to the face marked by red box in the picture) is

Fig. 13. (Best viewed in colors and magnification) Super-resolved results from real-world LR faces using automatic detection and alignment of LR face images. The first row is the input picture, the second row to the fifth row are the super-resolved results. For each example, the extracted, aligned LR face in the left, and the HR super-resolved face at the right.

the failure case, and the super-resolved results do not improve the quality as before. This is mainly because the input faces are in a variety of poses and with glasses (note that the training samples are frontal face and without glasses). The above points show that the characteristics of a learning-based vision system require certain amount of similarity between test and training samples [5].

E. Discussion

From the above real-world experiments, we can see that the super-resolved results are still noisy or superfluous and relatively poor compared with the results from our simulated experiments, in which training images and test ones are from the CAS-PEAL-R1 face database. This shows a weakness of the learning-based super-resolution techniques which depend highly upon the consistency (similarity) between the supporting training set and the input LR image to be handled [5], [35]. Specifically in actual application, such as super-resolution for surveillance face images, the captured faces are not necessarily visible in a portrait fashion but at an angle or arbitrary pose and under different illumination conditions, occlusions etc., all of which are truly representative of surveillance imagery.

Therefore, the enhancement of an interested surveillance face is an extreme complex and difficult problem. In order to get a satisfactory result, it requires a combination of a variety of techniques, such as recently proposed dense image correspondences [36], and facial component decomposition [37]. Besides, in actual surveillance environment, we often can get a sequence of face images [38]. Therefore, in addition to the prior information learned from the LR and HR co-occurrence model, the temporal redundancy information can also be used to guide the reconstruction process. Combining the advantages of multi-frame reconstruction-based methods (exploit the prior information from the given sequence) and learning-based methods (exploit the prior information from the given training set) to recover a high-quality and HR face image will be our further work. This also makes it possible to solve the occlusion problem.

Recent research in simultaneous face super-resolution and recognition indicates that most of the current face super-resolution methods can obtain visually appealing results, but they are not designed for the following face recognition task [39]–[41]. Nevertheless, we cannot arrive at the conclusion that the super-resolution process does not make any sense to the following recognition task, because the performance of face recognition algorithms relies heavily on the facial descriptor (representation of one face image). As Liang et al. [41] and Jiang et al. [42] pointed out, it is difficult for these holistic feature descriptors (e.g., Eigenface, Fisherface, or KPCA) to capture the high-frequency components of face images, which are crucial to the following recognition task. Therefore, if one face recognition algorithm utilizes the holistic-feature descriptors to represent a face image, there is no need to perform a super-resolution method, especially when the size of the input face is more than 30×30 pixels. In practice, Kurita et al. has observed that faces characterized by global features could often be more easily recognized in lower resolutions [43]. Since faces are very similar in overall configuration and only different on subtle details, which can be captured by these hand-crafted loal feature descriptors, such as Local Binary Pattern(LBP), Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG). Therefore, an efficient face super-resolution method that can effectually recover the high-frequency information will help to improve the recognition performance when using local-feature descriptors. In the further, designing an effective super-resolution approach that can capture much more discriminant facial details will be our second concern.

V. Conclusion

In this paper, we propose a novel face super-resolution framework, namely multilayer locality-constrained iterative neighbor embedding. In this method, we firstly use Bicubic interpolator to obtain the initial HR image, then calculate the distances between the estimated HR patch and these patches in the HR patch dictionary and search the K-NN in the HR space. And then, we utilize the calculated distances to penalize the reconstruction weights of the input LR patch in the LR space, thus obtaining the HR patch with the same reconstruction weights and corresponding K-NN in the LR space. After several iteration steps, we can get the target HR patch. Concatenating and integrating all the super-resolved HR patches, we generate the target HR face image. What's more, we extend the single LINE model to multilayer LINE one through intermediate dictionary learning, thus capturing the intrinsic image degradation shift and further refining the super-resolved results. Experimental results demonstrate that the proposed method can produce the best super-resolution recovery both quantitatively and perceptually in comparison with other state-of-the-art baselines.

Acknowledgment

The authors would like to thank Wei Zhang who is the author of [19] for providing the source code of his method.

References

[1] W. W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 327–340, Jan. 2012.

[2] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[3] Z. Lin and H.-Y. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 83–97, Jan. 2004.

[4] S. Baker and T. Kanade, "Hallucinating faces," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Mar. 2000, pp. 83–88.

[5] C. Liu, H.-Y. Shum, and C.-S. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local nonparametric model," in *Proc. IEEE Comput. Soc. Conf. CVPR*, vol. 1. Dec. 2001, pp. I-192–I-198.

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

[7] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *Int. J. Comput. Vis.*, vol. 106, no. 1, pp. 1–22, 2013.

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

[9] A. Chakrabarti, A. N. Rajagopalan, and R. Chellappa, "Super-resolution of face images using kernel PCA-based prior," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 888–892, Jun. 2007.

[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] X. Zhang, S. Peng, and J. Jiang, "An adaptive learning method for face hallucination using locality preserving projections," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–8.

[12] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

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

[14] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. CVPR*, vol. 1. Jun./Jul. 2004, pp. 275–282.

[15] B. Li, H. Chang, S. Shan, and X. Chen, "Aligning coupled manifolds for face hallucination," *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 957–960, Nov. 2009.

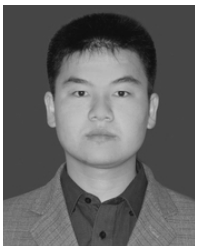
[16] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recognit.*, vol. 43, no. 6, pp. 2224–2236, 2010.

[17] 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.

[18] 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.

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

[20] J. Jiang, R. Hu, Z. Han, T. Lu, and K. Huang, "Position-patch based face hallucination via locality-constrained representation," in *Proc. IEEE ICME*, Jul. 2012, pp. 212–217.
[21] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1268–1281, Aug. 2014.
[22] J. Jiang, R. Hu, Z. Han, Z. Wang, T. Lu, and J. Chen, "Locality-constraint iterative neighbor embedding for face hallucination," in *Proc. IEEE ICME*, Jul. 2013, pp. 1–6.
[23] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[24] 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, Sep. 2012.
[25] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc., 2009, pp. 2223–2231.
[26] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conf. CVPR*, Jun. 2010, pp. 3360–3367.
[27] K. Su, Q. Tian, Q. Xue, N. Sebe, and J. Ma, "Neighborhood issue in single-frame image super-resolution," in *Proc. IEEE ICME*, Jul. 2005, pp. 1122–1125.
[28] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, Oct. 2000.
[29] M. Bevilacqua, A. Roumy, C. Guillemot, and M. Alberi, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. BMVC*, 2012, pp. 1–10.
[30] X. Gao, K. Zhang, D. Tao, and X. Li, "Image super-resolution with sparse neighbor embedding," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3194–3205, Jul. 2012.
[31] W. Gao *et al.*, "The CAS-PEAL large-scale chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 1, pp. 149–161, Jan. 2008.
[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[33] P. P. Gajjar and M. V. Joshi, "New learning based super-resolution: Use of DWT and IGMRF prior," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1201–1213, May 2010.
[34] M. Everingham, J. Sivic, and A. Zisserman, "Hello! my name is... Buffy—Automatic naming of characters in TV video," in *Proc. BMVC*, 2006, pp. 899–908.
[35] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with multiscale similarity learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 10, pp. 1648–1659, Oct. 2013.
[36] M. F. Tappen and C. Liu, "A Bayesian approach to alignment-based image hallucination," in *Proc. 12th ECCV*, vol. 7578. Oct. 2012, pp. 236–249.
[37] C.-Y. Yang, S. Liu, and M.-H. Yang, "Structured face hallucination," in *Proc. IEEE Conf. CVPR*, Jun. 2013, pp. 1099–1106.
[38] K. Nasrollahi and T. B. Moeslund, "Extracting a good quality frontal face image from a low-resolution video sequence," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 10, pp. 1353–1362, Oct. 2011.
[39] P. H. Hennings-Yeomans, S. Baker, and B. V. K. V. Kumar, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces," in *Proc. IEEE Conf. CVPR*, Jun. 2008, pp. 1–8.
[40] W. W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 327–340, Jan. 2012.
[41] Y. Liang, X. Xie, and J.-H. Lai, "Face hallucination based on morphological component analysis," *Signal Process.*, vol. 93, no. 2, pp. 445–458, Feb. 2013.
[42] J. Jiang, R. Hu, Z. Han, Z. Wang, and J. Chen, "Two-step superresolution approach for surveillance face image through radial basis function-partial least squares regression and locality-induced sparse representation," *J. Electron. Imag.*, vol. 22, no. 4, p. 041120, Oct. 2013.
[43] T. Kurita, N. Otsu, and T. Sato, "A face recognition method using higher order local autocorrelation and multivariate analysis," in *Proc. Int. Conf. Pattern Recognit.*, Aug./Sep. 1992, pp. 213–216.



Junjun Jiang received the B.S. degree in information and computing science from the School of Mathematical Sciences, Huaqiao University, Quanzhou, China, in 2009. He is currently pursuing the Ph.D. degree with the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan, China. His research interests include applications of image processing and pattern recognition in video surveillance, image superresolution, image interpolation, and face recognition.

Ruimin Hu received the B.S. and M.S. degrees from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1984 and 1990, respectively, and the Ph.D. degree in communication and electronic system from the Huazhong University of Science and Technology, Wuhan, China, in 1994. He is currently the Dean of the School of Computer with Wuhan University, Wuhan, and the Director of the National Engineering Research Center for Multimedia Software and the Key Laboratory of Multimedia Network Communication Engineering with Wuhan University. He is also the Executive Chairman of the Audio Video Coding Standard Workgroup of China in the Audio Section. He has authored two books, and over 100 scientific papers. His research interests include audio/video coding and decoding, video surveillance, and multimedia data processing.

Zhongyuan Wang received the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2008. He is currently an Associate Professor with the School of Computer, Wuhan University. He is also currently directing two projects funded by the National Natural Science Foundation Program of China. His research interests include video compression, image processing, and multimedia communications.

Zhen Han received the B.S. and Ph.D. degrees in computer science and technology and computer application technology from Wuhan University, Wuhan, China, in 2002 and 2009, respectively. He is currently a Lecturer with the School of Computer, Wuhan University. His research interests include image/video compressing and processing, computer vision, and artificial intelligence.