Noise Robust Face Hallucination via Locality-Constrained Representation

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

***Abstract*—Recently, position-patch based approaches have been proposed to replace the probabilistic graph-based or manifold learning-based models for face hallucination. In order to obtain the optimal weights of face hallucination, these approaches represent one image patch through other patches at the same position of training faces by employing least square estimation or sparse coding. However, they cannot provide unbiased approximations or satisfy rational priors, thus the obtained representation is not satisfactory. In this paper, we propose a simpler yet more effective scheme called Locality-constrained Representation (LcR). Compared with Least Square Representation (LSR) and Sparse Representation (SR), our scheme incorporates a locality constraint into the least square inversion problem to maintain locality and sparsity simultaneously. Our scheme is capable of capturing the non-linear manifold structure of image patch samples while exploiting the sparse property of the redundant data representation. Moreover, when the locality constraint is satisfied, face hallucination is robust to noise, a property that is desirable for video surveillance applications. A statistical analysis of the properties of LcR is given together with experimental results on some public face databases and surveillance images to show the superiority of our proposed scheme over state-of-the-art face hallucination approaches.**

***Index Terms*—Face hallucination, locality-constrained representation, neighbor embedding, position-patch, sparse representation, super-resolution.**

I. Introduction

With the rapid development of intelligent surveillance systems, surveillance cameras have been deployed in various areas including security and protection systems. Surveillance images, especially face images, can provide very important clues to criminal investigation. However, the resolution of a video camera is usually not High-Definition (HD) (see Fig. 1(a)), and the low resolution of the interested face in the picture resulted from the long distance between the object and the camera (see Fig. 1(c)) makes it almost impossible to provide useful information (see Fig. 1(b)). Moreover, in real surveillance scenarios, the qualities of the surveillance images are deteriorated by many environmental factors, such as underexposure, optical blurring, and defocusing. Consequently, the face images of interest are too blurred to be identifiable by humans. In order to obtain enough facial feature details for recognition, a new technique called face super-resolution or face hallucination is adopted to generate High-Resolution (HR) face image from Low-Resolution (LR) images. Existing image hallucination methods mainly fall into two categories: reconstruction-based techniques and learning-based techniques. Based on registration and alignment of multiple LR images of the same scene in sub-pixel accuracy, the former are more susceptible to ill-conditioned registration and inappropriate blurring operators [1], while the latter can generate better performance and higher magnification factor—with the help of a set of training examples. We focus on learning-based method in the sequel.

(a) (b) (c)

Fig. 1. Typical frames from surveillance videos. (a) and (c) are the surveillance images from a camera with CIF size (352×288 pixels) and a camera with 720P size (1280×720 pixels) respectively; (b) shows two interested faces extracted from (a) and (c).

Learning-based image hallucination has attracted much attention in recent years [1]–[12], [20]–[23], [27], [28], [30]–[33]. With the help of a training set of LR and HR images, one can obtain an HR image from an LR one by building a co-occurrence model. The first learning-based super-resolution method was proposed by Freeman *et al.* [20] who utilized a patch-wise Markov network to model the relationship between local regions of images and the underlying scenes. But this approach is computationally intensive and sensitive to training examples. Baker and Kanade [21] developed a learning-based super-resolution method specifically for human face. Given an LR input image, their algorithm infers the missing high-frequency components from a parent structure with LR and HR training samples. In [2], Liu *et al.* proposed to integrate a global parametric Principal Component Analysis (PCA) model with a local non-parametric Markov Random Field (MRF) model for face hallucination. The work of [21] and [2] spurred much follow-up research that fall into two categories: global face based parameter estimation methods [3], [8], [10], [11], [12], [22] and local

Manuscript received December 05, 2012; revised September 27, 2013 and February 23, 2014; accepted March 07, 2014. Date of publication March 11, 2014; date of current version July 15, 2014. This work was supported by the major national science and technology special projects (2010ZX03004-003-03), the National Key Technologies R&D Program (2013AA014602), the National Natural Science Foundation of China (61231015, 61070080, 61003184, 61172173, 61303114, and 61170023), and the China Postdoctoral Science Foundation funded project (2013M530350). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shin'ichi Satoh.

The authors are with the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan, China (e-mail: junjun0595@163.com; hrm1964@163.com; hanzhen_1980@163.com; wzy_hope@163.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/TMM.2014.2311320

patch image restoration methods [1], [4], [5], [6], [7], [9], [27], [30], [31].

Global face based parameter estimation methods: Wang and Tang [3] proposed a face hallucination approach using an eigentransformation. In this approach, an LR face image is first decomposed into a linear combination of LR face images in the training set using PCA. Then, the target HR face image is reconstructed by replacing the LR training images with the corresponding HR ones, while using the same coefficients. This approach is easy to be implemented and its performance is reasonably good. Because of its selection of eigenfaces and the maximization of facial information from the LR face image, it is robust against noise to some extent. In [22], Chakrabarti *et al.* utilized a Kernel Principal Component Analysis (KPCA) model to hallucinate face images. Park *et al.* [11] decomposed an LR input face into many prototype faces via PCA and applied a recursive error back-propagation method to reconstruct the HR face. All these PCA based global approaches can well capture the global face appearance variations. However, they fail to render effectively the fine individual details of an input face, especially when it is different from the training samples or when the size of the training samples is small. To alleviate the above problem, techniques of decomposing a complete image into smaller patches have been introduced recently.

Local patch image restoration methods: Inspired by Locally Linear Embedding (LLE) [19], Chang *et al.* [4] developed a Neighbor Embedding (NE) algorithm for super-resolution of general images. Assuming the LR image patch space and the HR one share the same local geometry, the local geometry of the LR patch space is mapped to the HR patch space to generate the HR image patches through linear combination. In [5], from the perspective of manifold alignment, Li *et al.* extended the approach of [4] to perform face hallucination on a synthesized common manifold by two explicit mappings. Different from those patch based approaches using a fixed number of neighbors for reconstruction, a single image super-resolution method based on sparse coding that adaptively selects the most relevant neighbors to minimize the reconstruction error was recently introduced by Yang *et al.* [8]. However, this method has limited subjective visual effects because it fails to make full use of the prior knowledge of human faces.

Recognizing the fact that human face is a class of highly structured object and consequently position information plays an important role in its reconstruction, several position-patch based face hallucination methods have been proposed [6], [7], [9]. For example, Ma *et al.* [6], [7] obtained the representation vector by solving a constrained least square problem, which is referred as Least Square Representation (LSR) in this paper. Compared with some manifold learning-based methods which do not incorporate the position priors, LSR is more efficient because it reasonably utilizes position information. However, when the number of the training samples is much larger than the dimension of a patch, the solution to least square estimation in LSR is not unique [8]. In order to find an unbiased solution, Jung *et al.* [9] employed a sparsity-constrained optimization to replace least square estimation, which is referred as Sparse Representation (SR) in this work. By combing the sparse coding method [8] and the LSR method [6], [7], SR outperforms both [8] and LSR. It is well known that locality is more important than sparsity in revealing the non-linear manifold structure of face image patches [13], yet SR overemphasizes sparsity and neglects locality. As a result, to reconstruct the input image patch, very distinct patches may be chosen. In addition, both LSR and SR fail to consider the effect of noise. In fact, LSR magnifies the noise rather than suppresses it, and the SR solution is usually not stable, especially when the noise is strong. Thus, the obtained weights of LSR and SR are not robust for actual surveillance images.

In [30], [31], Zhang and Cham proposed a K-nearest neighbors embedding based position-patch method for inferring local features of face image in the Discrete Cosine Transform (DCT) space. In this method, the DC coefficient is estimated by an interpolation-based method, while the AC coefficients (the high-frequency face features) are estimated by linearly combining the K nearest sample patches through a LLE based simplified MRF model. This method is very efficient and robust to low illumination. Most recently, Yang *et al.* [32] proposed to represent a face by three categories including face components, edges, and smooth regions, and exploit these local image structures for face hallucination under various poses and expressions.

Since local image patches are similar, it will be more accurate to represent one image patch using a few neighbor patches, leading to a local representation of image patches. Moreover, locality based representation is robust against noise because it replaces the noisy image patch with similar "clean" ones rather than synthesizes noisy image patch as in LSR and SR. Inspired by this, in this paper, we introduce a novel patch representation method for face image hallucination, called Locality-constrained Representation (LcR) in which a locality constraint is incorporated into the least square inversion problem. Unlike existing methods that represent one input image patch collaboratively [6], [7] or sparsely [8], [9], our proposed method projects each image patch into its neighborhoods in the training set adaptively so that both sparsity and locality are preserved. When compared with [30], which preserves the locality in a hard way (using a fixed number K neighbors for reconstruction) and may lead to over- or under-fitting problem [8], our method can adaptively select the neighbors by a locality adaptor. The proposed method has the following distinct features.

- The locality constraint helps reveal the non-linear manifold structure of face image patch space because it makes the solution of the least square problem fixed on the one hand, and captures the fundamental similarities between neighbor patches on the other;
- Compared with traditional neighbor embedding methods that use a fixed number of neighbors for reconstruction, it adaptively chooses the most relevant patches to avoid over- or under-fitting while giving sharper contours and richer details;
- By adaptively choosing the neighbor patches, it is very robust against noise in real surveillance scenarios.

In addition to preliminary results published in [27], here we give a detailed description of our LcR method with: i) an LcR ensemble to improve the performance of our original LcR model; ii) analysis on the properties of sparsity and locality; iii) extensive experimental evaluations on its performance, especially on its robustness against noise.

The rest of this paper is organized as follows. Section II reviews LSR and SR. Section III presents the proposed LcR method. Section IV illustrates the sparsity and locality of LcR.

Experimental results are presented in Section V. Section VI concludes the paper.

II. Existing Position-Patch Based Representation Approaches

Let Y^m denote the training face images, $m = 1, \ldots, M$, where M is the size of the training samples. Each face image is divided into N small overlapping patch sets $\{Y^m(i,j)|1 \leq i \leq U, 1 \leq j \leq V\}$, $N = UV$, U represents the patch number in every column, V represents the patch number in every row, and the term (i,j) indicates the position information (please refer to the previous version [27] for a detailed description). For the patch located at position (i,j), it can be represented by M training samples located at the same position with a weight vector, $w(i,j) = [w_1(i,j), w_2(i,j), \ldots, w_M(i,j)]^T$.

For the input face image denoted in patches as $\{X(i,j)|1 \leq i \leq U, 1 \leq j \leq V\}$, different representation schemes convert each patch into a M-dimensional optimal weight vector to generate the final patch representation. In this section, we review two existing patch representation schemes, Ma *et al.*'s LSR [6], [7] and Jung *et al.*'s SR [9].

A. Least Square Representation

By incorporating the prior of position information, LSR [6], [7] uses patches from all training samples at the same position (i,j) to represent each patch $X(i,j)$ collaboratively with

$$X(i,j) = \sum_{m=1}^{M} Y^m(i,j) w_m(i,j) + e, \tag{1}$$

where e is the reconstruction error vector.

The reconstruction weights of the input image patch $X(i,j)$ can be computed by the following constrained least square fitting problem

$$w^*(i,j) = \arg\min_{w(i,j)} \left\| X(i,j) - \sum_{m=1}^{M} Y^m(i,j) w_m(i,j) \right\|_2^2 \quad \text{s.t.} \sum_{m=1}^{M} w_m(i,j) = 1. \tag{2}$$

It is a constrained least squares problem whose closed-form solution [4] can be solved by computing a Gram matrix.

B. Sparse Representation

In practice, the solution of equation (2) may be unstable or does not exist. The way around is to impose regularization terms onto the objective function. Jung *et al.* [9] introduced the sparse representation theory and used a small subset of patches to represent $X(i,j)$ in the place of collaborative performance over the whole training samples. It converts equation (2) to a standard sparse representation problem

$$\min_w \|w(i,j)\|_0 \text{ s.t.} \left\| X(i,j) - \sum_{m=1}^{M} Y^m(i,j) w_m(i,j) \right\|_2^2 \leq \varepsilon, \tag{3}$$

where the ℓ_0-norm counts the number of nonzero entries in a vector. Since there are both the numerical unstable and NP-hard problems in combinatorial ℓ_0-norm minimization [24], recent theories developed from sparse representation [25] suggest that if the solution is sufficiently sparse, then the sparsest solution can be recovered via ℓ_1-norm minimization

$$\min_w \|w(i,j)\|_1 \text{ s.t.} \left\| X(i,j) - \sum_{m=1}^{M} Y^m(i,j) w_m(i,j) \right\|_2^2 \leq \varepsilon, \tag{4}$$

where ℓ^1-norm sums up the absolute weights of all entries in a vector, and $\varepsilon \geq 0$ is the allowed error tolerance. Note that the sparsity constraint not only leads to the exact solution of the under-determined problem, but also allows the learned representation for each patch $X(i,j)$ to capture salient properties, yielding minimized reconstruction error.

III. Proposed Method

A. Locality-Constrained Representation

The work of [9] emphasizes that strong sparsity of the weight vector $w(i,j)$ is important in representing the input patch, however, it neglects a locality constraint, which is more important than sparsity in revealing the true geometry of a nonlinear manifold [13]. In other words, SR might represent one input patch by distinct training patches.

Following the intuition that Local Coordinate Coding (LCC) [13] explicitly encourages a local representation, we incorporate a locality constraint into the objective function. Additionally, we adopt shrinkage measures as in ridge regression on the weight vector. Thus, our objective is to

$$\min \|d(i,j) \circ w(i,j)\|_2^2 \text{s.t.} \|X(i,j) - \sum_{m=1}^{M} Y^m(i,j) w_m(i,j)\|_2^2 \leq \varepsilon \sum_{m=1}^{M} w_m(i,j) = 1, \tag{5}$$

where $\circ$ denotes point-wise vector product and $d(i,j)$ is a M-dimensional vector that penalizes the distance between $X(i,j)$ and each training patch at the same position. It is simply determined by the Euclidean distance

$$d_m(i,j) = \|X(i,j) - Y^m(i,j)\|_2, 1 \leq m \leq M. \tag{6}$$

The Lagrangian for equation (5) becomes

$$w^*(i,j) = \arg\min_{w(i,j)} \left\{ \left\| X(i,j) - \sum_{m=1}^{M} Y^m(i,j) w_m(i,j) \right\|_2^2 + \tau \sum_{m=1}^{M} [d_m(i,j) w_m(i,j)]^2 \right\}. \tag{7}$$

Equation (7) consists of two parts: the first term measures the reconstruction error while the second one preserves locality, with τ representing the regularization parameter that balances the contribution of the reconstruction error and locality of the solution. When $\tau = 0$, LcR reduces to LSR. More discusses about τ are given in Section V-C. In our proposed LcR method, the roles of the locality constraint are twofold: On one hand, it makes the solution fixed; on the other hand, as discussed in Section IV-A,

it introduces a locality-constrained sparse representation to each patch, yet this "sparsity" is much weaker than that in the sense of ℓ^0-norm.

B. Optimization

The objective function (7) can be represented in the following matrix form

$$w^*(i,j) = \arg\min_{w(i,j)} \left\{ \|X(i,j) - Y(i,j)w(i,j)\|_2^2 + \tau \|Dw(i,j)\|_2^2 \right\}, \quad (8)$$

where $Y(i,j)$ is a matrix with its columns being training patches $Y^m(i,j)$ and D is the $M \times M$ diagonal matrix with

$$D_{mm} = d_m(i,j), 1 \le m \le M. \quad (9)$$

Following [4], [7], the solution of a regularized least square in equation (8) can be derived analytically as

$$w(i,j) = (G(i,j) + \tau D^2) \backslash ones(M,1), \quad (10)$$

where $ones(M,1)$ is a $M \times 1$ column vector of ones, the operator "\" denotes the left matrix division operation, and G is the covariance matrix for $X(i,j)$ as

$$G(i,j) = C^T C, \quad (11)$$

with

$$C = X(i,j)ones(M,1)^T - Y(i,j). \quad (12)$$

The final optimal weight is obtained by rescaling to satisfy $\sum_{m=1}^{M} w_m(i,j) = 1$.

C. Face Hallucination via LcR

For face hallucination, the training set is composed of LR and HR face image pairs. HR face images are denoted as $\{Y_H^m\}_{m=1}^M$ while their LR counterparts are denoted as $\{Y_L^m\}_{m=1}^M$. The primary task is to reconstruct the HR face image X_H from the observed LR face image X_L.

At the beginning, we divide the training face images and the LR input face image into patches using the same dividing scheme as in [27]. For each LR input image patch, it is approximated by a linear combination of the LR patch at the same position using LcR, and we obtain a set of weights on the LR training image patches. Since the LR patch image manifold and the HR one share the same topology [4], a new HR patch of the same position can be synthesized by keeping the weights and replacing the LR training image patches with the corresponding HR ones. By concatenating all the HR patches to their corresponding positions and averaging pixel values in the overlapping regions, we can get an estimation of the HR target face. The entire face hallucination process is given in Algorithm 1.

Pre-processing Based on Image Priors: In order to increase the overall efficiency of our proposed method, we employ one pre-processing step and exploit an assumption about the nature of images.

Following [20], we assume the high-frequency band I_H to be conditionally independent of the low-frequency band I_L, given the middle-frequency band I_M. Mathematically, we have

$$p(I_H|I_L) = p(I_H|I_M). \quad (13)$$

Based on this assumption, to predict the high-frequency band, only the mid-frequency band rather than all lower frequency bands of the LR input image will be utilized. As in [28], in order to obtain the middle frequency components, we linearly interpolate (*e.g.*, by using the Bicubic interpolation algorithm) each blurred LR face image back up to the original HR grids to form an LR input face image.

Algorithm 1 (Face hallucination via LcR).

1. Input: Training set $\{Y_H^m\}_{m=1}^M$ and $\{Y_L^m\}_{m=1}^M$, an LR image X_L, *patch_size*, *overlap* and regularization parameter τ.

2. Divide each of the training images and the LR input image into N small patches according to the same location of face respectively.

3. For each patch of X_L **do**

- Calculate the Euclidean distance between the LR input image patch $X_L(i,j)$ and all the M training image patches $\{Y_L^m(i,j)\}_{m=1}^M$ at position (i,j):

$$d_m(i,j) = \|X_L(i,j) - Y_L^m(i,j)\|_2, 1 \le m \le M.$$

- Compute the optimal weight vector $w^*(i,j)$ for the LR input image patch $X_L(i,j)$ with the LR training image patches $\{Y_L^m(i,j)\}$:

$$w^*(i,j) = \arg\min_{w(i,j)} \left\{ \left\| X(i,j) - \sum_{m=1}^{M} Y^m(i,j)w_m(i,j) \right\|_2^2 + \tau \sum_{m=1}^{M} [d_m(i,j)w_m(i,j)]^2 \right\}.$$

- Construct the HR patch by

$$X_H(i,j) = \sum\nolimits_{m=1}^{M} Y_H^m(i,j)w_m^*(i,j).$$

4. End for

5. Integrate all the reconstructed HR patches above according to the original position. The final HR image X_H can be generated by averaging pixel values in the overlapping regions.

6. Output: HR hallucinated face image X_H.

IV. Sparsity and Locality of LcR

Recently the importance of sparsity and locality for data representation and classification has attracted a lot attention. Inspired by biological visual systems, many researchers have been arguing that sparse features of signals are useful for settling computer vision problems [24], [26]. On the other hand, in general pattern recognition problems such as data representation and dimension reduction, data locality has been proved to be

critical [13], [19]. In this section, we show that the sparsity and locality of LSR, SR and LcR through qualitative and quantitative analysis, respectively. All the experiments in this section are conducted on FEI Face Database [14]—more details about the database are given in Section V-A.

A. Sparsity of LcR

For a better understanding of our approach, we plot the concatenated optimal weight vector $\vec{w}$ for different representation methods, LSR, SR and LcR. Note that the optimal weight vector $\vec{w}$ is formed by concatenating the optimal weight vector $w^*(i,j)$ of each patch of all 40 test face images, thus the length (L) of $\vec{w}$ can be calculated by L = (the number of test face images) $\times$ (the length of the optimal weight of one patch) $\times$ (the patch number in every column) $\times$ (the patch number in every row) $= 40 \times 360 \times 15 \times 12 = 2592000 = 2.592 \times 10^6$. Note that the length of the optimal weight of one patch is equal to the sample number of the training set, and the patch number in every column and the patch number in every row can be obtain from $\text{ceil}\{\frac{imrow-overlap}{patch\_size-overlap}\}$ and $\text{ceil}\{\frac{imcol-overlap}{patch\_size-overlap}\}$ by substituting $imrow = 120$, $imcol = 100$, $patch\_size = 12$ and $overlap = 4$ to them respectively. ceil(x) is the function that rounds the elements of x to the nearest integers towards infinity.

As in the first row on the right of Fig. 2, the sorted plot of $\vec{w}$ shows that the reconstruction weight vector of LSR is not sparse since it treats all the samples as equals. Our proposed LcR approach (third row) obtains a sparse result to some extent compared with SR (second row). It shows that LcR can truly reveal the image space, which is embedded in a nonlinear manifold [4], [19]. At the same time, we also observe an encouraging phenomenon: the weights obtained by LSR and SR are equably positive and negative, whereas the reconstruction weights generated by LcR are nearly all non-negative (most values are larger than zero). This indicates that when several neighbor patches are used to represent the target image patch, all the image patches used contributes positively.

We also quantitatively evaluate and compare the sparsity of these three methods. The Gini Index[1] (GI) [18]

$$GI(\vec{w}) = 1 - 2\sum_{i=1}^{L} \frac{|w_{[i]}|}{\|\vec{w}\|_1}\left(\frac{L-i+1/2}{L}\right), \qquad (14)$$

is another metric of measurement introduced to measure the sparsity of the optimal weight vector $\vec{w} = [w_1, w_2, \cdots, w_L]$, where $w_{[i]}$ is the i-th element of re-ordered optimal weight vector $\vec{w}_{GI}$, $\vec{w}_{GI} = [|w_{[1]}|, |w_{[2]}|, \cdots, |w_{[L]}|]$ and $|w_{[1]}| \leq |w_{[2]}|, \cdots \leq |w_{[L]}|$, where $[1], [2], \cdots, [L]$ are the new indices after the sorting operation. Table I gives the GI results of three different representation methods. It indicates that although the sparsest representation method is SR, the optimal weight vector obtained by LcR is indeed sparse (24% less than that obtained by the SR method and 85% more than that obtained by the LSR method).

[1]GI is normalized, and assumes values between 0 and 1 for any vector. Further, it is 0 for the least sparse signal with all the coefficients having an equal amount of energy; and 1 for the sparsest one which has all the energy concentrated in just one coefficient.

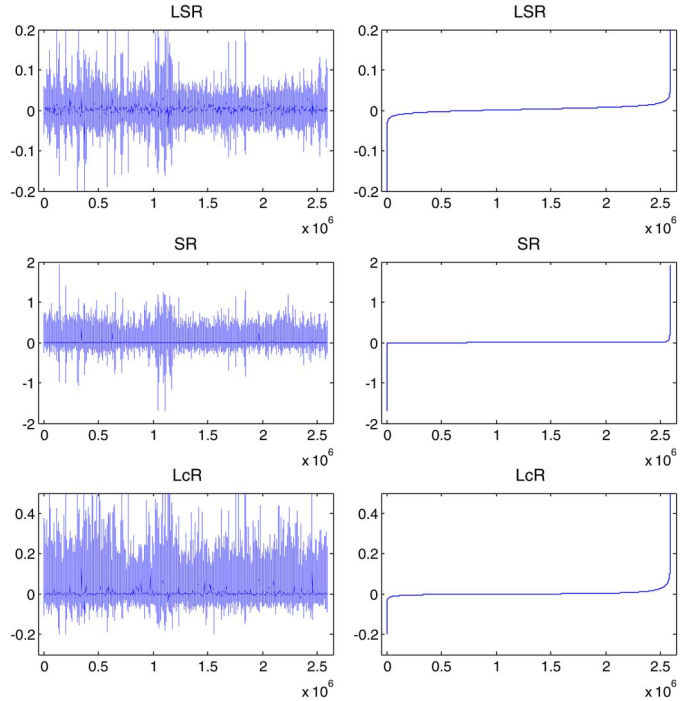


Fig. 2. The sparsity of the concatenated optimal weight vector for different representation methods: LSR (top row), SR (middle row), LcR (bottom row). The left figures are the plots of the concatenated optimal weight vector and the right figures are the corresponding plots of sorted ones. The more smooth (at the same time close to zero) the sorted ones, the sparser the concatenated optimal weight vector.

TABLE I
THE GINI INDEX OF THREE DIFFERENT REPRESENTATION METHODS

Methods	LSR	SR	LcR
GI	0.2181	0.5325	0.4040

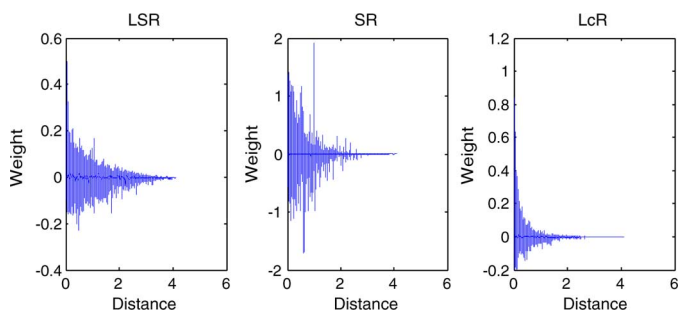


Fig. 3. The weight vector of different representation methods according to the sorted distance.

B. The Locality of LcR

As argued in [13], locality is more important than sparsity. Thus, [19] states that each point on the well-sampled manifold can be linearly represented by a few neighbors of the given data. We test the locality of LSR, SR and LcR, respectively. As proposed in Section IV-A (the formation of the optimal weight vector $\vec{w}$), we combine all the distances of each patch of all 40 test images to form a L dimensional distance vector $\vec{d}$. Then, we sort the distance vector $\vec{d}$ in

$$sort(\vec{d}) = \left[\vec{d}_{[1]}, \vec{d}_{[2]}, \cdots, \vec{d}_{[L]}\right], \vec{d}_{[1]} \leq \vec{d}_{[2]} \leq \cdots \leq \vec{d}_{[L]}. \qquad (15)$$

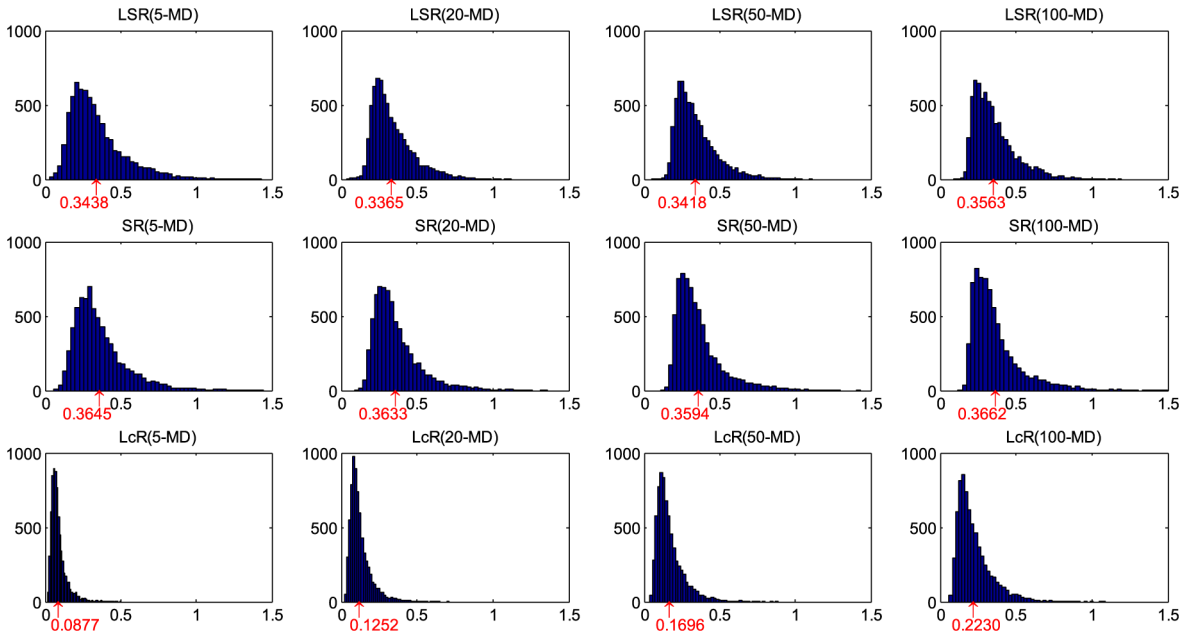


Fig. 4. Histograms of the K-mean distance for different methods: LSR (top row), SR (middle row), LcR (bottom row). The red arrows indicate the average values.

According to the sorted distance vector $sort(\vec{d})$, the new optimal weight vector $\vec{w}_d$ is

$$\vec{w}_d = \left[w_{[1]}, w_{[2]}, \cdots, w_{[L]}\right]. \tag{16}$$

We plot in Fig. 3 the sorted distance $sort(\vec{d})$ as abscissa and the new optimal weight vector $\vec{w}_d$ as coordinate for LSR, SR and LcR, respectively. It is seen that the weights decrease with the increase of distance. This trend is more pronounced for LcR than for LSR and SR. As for LcR, large weights mostly concentrate on small distances. We also notice that the weights of LSR and SR have many large spikes as the sorted distance increases, while our proposed LcR approach is capable of removing these spike coefficients, thereby leading to improved locality.

In addition, we also quantitatively check the locality of LSR, SR and LcR. At the beginning, we define an evaluation metric of locality called $K - MeanDistance$ (K-MD). Let $X(i,j)$ and $\{Y^m(i,j)|m \in C_K(X(i,j))\}$ denote the inference patch and the K most significant patches in training set with the same position (i,j) respectively, where $C_K(X(i,j))$ is the patch index set. The larger the entries in the weight vector $w(i,j)$, the more important the corresponding patch. The K-MD is defined as

$$K\text{-MD} = \left\{ \frac{\sum_{C_K(X(i,j))} \|X(i,j) - Y^m(i,j)\|_2^2}{K} \right\}^{\frac{1}{2}}. \tag{17}$$

According to the K-MD definition, the locality of a representation approach is measured by computing the distance between the LR input patch and K largest weight patches in the representation of the inference patch. Therefore, the smaller K-MD, the better locality.

The locality comparison is given in Fig. 4, from which we see that the average values of K-MD of LcR are much smaller than those of LSR and SR (regardless of the values of K, such as $K = 5$, $K = 20$, $K = 50$ or $K = 100$). Thus LcR can better capture the locality property than LSR and SR. Meanwhile, we also notice that the K-MDs with different K either remain constant or change very little in the LSR and SR cases. But for LcR, the value of K-MD increases with K. This implies that those image patches given by LcR are closer to the input one with larger weights, while in LSR and SR cases, all the training patches are treated in the same way. This indicates that our proposed LcR method pays more attention to those patches with small distance to the input (to preserve locality).

V. Experiment Results

To evaluate the proposed LcR algorithm, we compare it with some other state-of-the-art methods for face hallucination: Wang *et al.*'s global face method [3], NE method [4], LSR [7], SR [9] and DCT [30]. Experiments are performed on FEI Face Database[2] [14] and brief description of the dataset is provided along with the details of the experiments in Section V-A. The objective results and the objective metrics, *i.e.*, PSNR and SSIM index [15], will be reported in Section V-B. We also test the effect of parameter settings in Section V-C and the robustness against noise in Section V-D. In order to further verify the superiority of LcR over other methods, we repeat the experiments on some real-world images in Section V-E. Matlab code available upon e-mail request (junjun0595@163.com).

A. Database and Parameter Settings

Experiments described in this paper are conducted on the frontal and pre-aligned images of FEI Face Database. It contains 400 images from 200 subjects (100 men and 100 women) and each subject has two frontal images, one with a neutral expression and the other with a smiling facial expression. Human faces in the database are mainly from 19 to 40 years old with distinct appearances, hairstyles and adornments (some samples are shown in Fig. 5). All the images are cropped to 120×100 pixels and we randomly choose 360 images (180 subjects) as the training set, leaving the rest 40 images (20 subjects) for

[2]It is publicly available on http://fei.edu.br/~cet/facedatabase.html.

Fig. 5. Some training faces in FEI Face Database.

Fig. 6. Comparison of results based on different methods. From left to right: LR input faces, hallucinated faces by Wang *et al*. [3], NE [4], LSR [7], SR [9], DCT [30], LcR, and PreLcR, and the last column is the original HR faces. (Note that the effect is more pronounced if the figure of the electronic version is zoomed.)

testing. Therefore, all the test images are absent completely in the training set. The LR images are formed by smoothing (an averaging filter of size 4×4) and down-sampling (Unless otherwise specified, the down-sampling factor is 4, thus the size of LR face images are 30×25 pixels) corresponding HR images.

To pursue the best performance, we tune the parameters for all comparative methods their best possible results. In particular, for Wang *et al.*'s global face method, we let the variance accumulation contribution rate of PCA be 99% (around 100 bases). The number of neighbors in NE is set to around 50. For the SR method, we set error tolerance to 1.0. For Wang *et al.*'s DCT method, the number of neighbors is set to 10 (we change the value and find 10 is the best choice). For these patch based methods, such as NE, LSR, SR and our proposed LcR, we recommend to use the size of 12×12 pixels for HR image patch and the overlap between neighbor patches is 4 pixels, while corresponding LR image patch size is set to 3×3 pixels with an overlap of 1 pixel. For more details about the performance under different path size and overlap, please refer to Section V-C. Our algorithm has only one free parameter τ. We carefully tune it to gain the best performance and we will demonstrate the performance versus the value of τ in Section V-C. The following reported hallucination results are obtained under the best parameter, $\tau = 0.04$ for LcR model and $\tau = 0.02$ for the enhanced LcR approach with pre-processing presented in Section III-C. In the following, we use PreLcR to denote the enhanced LcR.

B. Results on FEI Face Database

Fig. 6 shows some samples of the hallucinated results generated by different methods. The first column are the LR input faces, the last column are the ground true HR faces, while column 2 to column 8 are the hallucinated HR faces based on seven different methods. From the visual results of hallucinated faces, the following points can be observed:

- Patch based methods outperform global faces reconstruction methods. Wang *et al.*'s global face method [3] did not maintain the global smoothness of images, but leads to "ghost" artifacts around contours on the contrary. In comparison, patch based methods show their superiority in further enhancing the edges and textures. This is mainly because that [3] is based on a statistical mode, which could not reveal the distribution of data when the training samples are not sufficient (*e.g.,* 360 training samples in our experiments while the feature dimension is $30 \times 25 = 750$). When the LR input face image is very different from those in the training database, [3] does not capture the novel structures, but cause serious distortions (see the second row);
- Position-patch based methods are better than NE. The third column indicates that the face images hallucinated by NE are blurring and have obvious artifacts. This is primarily due to over- or under-fitting. Since the incorporating of position information of human faces, position-patch based face hallucination methods can capture the latent semantic information, while NE considers only the aspect of reconstruction and ignores the semantic features of faces;
- Further examination shows that the proposed LcR method successfully captures more high frequency components and fewer artifacts than the other three position-patch based methods, LSR, SR and DCT (see the mouth in the second row). The excellent visual quality of our method owes to the LcR strategy. In addition, we also show the results of the PreLcR method. By pre-processing, the hallucinated faces of PreLcR are more visually pleasant (see the mouth in the first two rows and the face contour in the last two rows, which are pointed by red arrows);
- The PSNR and SSIM index[3] of different methods are shown in Table II (see the second column whose down-sampling factor is set to 4). The Bicubic Interpolation (BI) is taken to be the baseline for comparison. Again, we can see that patch based methods perform much better than Wang *et al.*'s global face method and BI. We found that DCT based method is no better than some other patch based methods, and this is due to the fail of manifold assumption in the DCT space. In particular, the consistence between LR and HR patches in pixel space is better than that in the DCT space. LcR achieves the highest PSNR and SSIM values. The average PSNR and SSIM improvements of LcR method over the second best

[3]The higher the SSIM value, the better is the face hallucination quality. The maximum value of SSIM is 1, which means a perfect reconstruction. Compared with the measure of PSNR, SSIM can better reflect the structure similarity between the target image and the reference image.

TABLE II
PSNR (dB) AND SSIM COMPARISON OF DIFFERENT METHODS WITH DIFFERENT DOWN-SAMPLING FACTORS

Factors	4	4	8	8	16	16
Methods	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicibic	27.49	0.8417	22.94	0.6617	19.01	0.5322
Wang [3]	27.75	0.7582	26.05	0.7297	24.16	0.6954
NE [4]	31.23	0.8975	**27.75**	**0.8088**	**24.60**	**0.7283**
LSR [7]	31.90	0.9032	26.88	0.7813	23.94	0.7063
SR [9]	**32.11**	**0.9048**	26.88	0.7814	23.90	0.6993
DCT [30]	30.55	0.9011	26.55	0.7881	21.79	0.6327
LcR	32.76	0.9145	27.86	0.8102	24.60	0.7320
PreLcR	**33.13**	**0.9216**	**28.64**	**0.8345**	**25.07**	**0.7429**
Improvemen	1.02	0.0168	0.89	0.0257	0.47	0.0146

method, *i.e.*, SR, are 0.65 dB and 0.0097, respectively. The improvement of PreLcR approach is much more effective, which gains 1.02 dB and 0.0168 more than SR method. This serves to show the importance of exploring prior information on the image representation model.

To compare the ability of hallucinating very LR face images, the performances of different methods with the down-sampling factor of 8 and 16 are evaluated. The third and fourth columns of Table II tabulate the results. We note that the gain of our proposed method over the comparison methods is still evident although with a high down-sampling factor. In the meantime we see that when using very LR input, the superiority of employing the holistic structures of facial images will be more apparent (see the forth column, Wang *et al.*'s global face method is better than some local patch based methods). In Fig. 7, we have given results for different methods with different down-sampling factors. The overall quality of results for high down-sampling factor is lower because of increased inconsistency between LR image HR images [17]. In spite of this, the proposed method yields reasonable results, while the comparison methods produce several artifacts, particularly around the eyes, the mouth and the contours. From the table and the figure, it seems that the best alternative is a position-patch based method that incorporates suitable prior information, *e.g.*, the locality constraint.

Fig. 7. Hallucinated face images with different down-sampling factors. From left to right: LR input faces, hallucinated faces by Wang *et al*. [3], NE [4], LSR [7], SR [9], DCT [30], and PreLcR, and the last column is the original HR faces. Note that the first two rows and last two rows are the hallucinated results with down-sampling factor of 8 and 16 respectively.

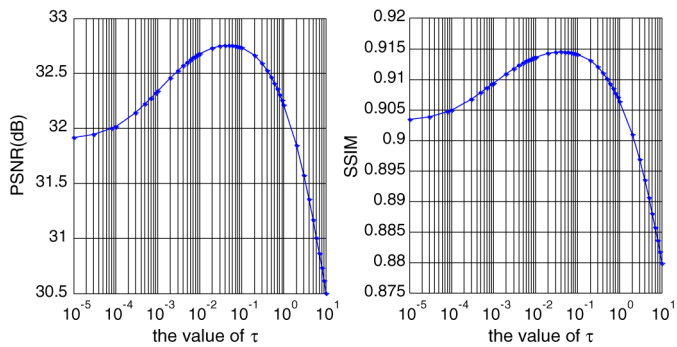


Fig. 8. The average PSNR and SSIM values of the proposed LcR method with different τ.

C. Parameter Analysis

In this subsection, we investigate the effect of the different parameter settings.

1) The Performance of Different τ: To demonstrate the effectiveness of locality, we evaluate the influence of the proposed LcR method by choosing different regularization parameters (τ), which controls the weights of locality constraint in the objective function. As shown in Fig. 8, when $\tau = 0$, which can be regarded as the case of LSR, the performance of LcR is restricted. With the increase of τ, more benefits on performance can be gained. This implies that the locality constraint is essential for patch reconstruction. However, we should also see that the value of τ could not be set too high. Therefore, with a proper regularization parameter τ, LcR will gain good results.

2) The Performance of Different Patch Size: For the local patch based method, the size of a patch is important for getting reliable results. If a patch is too small, it would give little information and cannot capture the geometric structure of a human face and the hallucinated face becomes noisy. On the other hand, as a patch becomes larger, the hallucinated face may be smooth and lost some visual details, in addition, the dimension of a patch becomes larger, it needs much more training images to extract reliable generalized basis, especially when the test image is far different from the training face images. Table III tabulates the PSNR and SSIM indexes of different methods under different patch size and overlap pixels. Given the same patch size, the larger the overlap degree is, the better the performance is. However, a larger overlap degree means higher computational complexity. From Table III, we can learn that the superiority of the proposed method over the comparison methods, such as NE, LSR and SR, is unaffected by the patch size and overlap, *e.g.*, the patch size takes the value between 4 pixels (the smallest patch) and 24 pixels while the overlap takes the value between 0 pixels and 20 pixels.

3) The Influence of Registration Error: Accurate registration of the input face image is vital for the position-patch based methods. The registration error will relieve the effects of face position prior. In all above experiments, we test the proposed face hallucination method based on the assumption that LR face

TABLE III
PSNR (dB) AND SSIM COMPARISON OF DIFFERENT METHODS UNDER DIFFERENT PATCH SIZE AND OVERLAP PIXELS

Patch Size	Overlap	NE [4]		LSR [7]		SR [9]		LcR		PreLcR	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
4	0	28.80	0.8140	28.28	0.8042	27.91	0.7991	28.81	0.8149	32.63	0.9105
8	0	31.36	0.8838	30.58	0.8715	30.33	0.8661	31.51	0.8875	32.92	0.9163
	4	32.17	0.9044	31.18	0.8918	31.23	0.8912	32.19	0.9049	33.17	0.9216
12	0	30.03	0.8556	31.40	0.8897	31.45	0.8883	32.22	0.9030	32.87	0.9164
	4	31.23	0.8975	31.90	0.9032	32.11	0.9048	32.76	0.9145	33.11	0.9210
	8	31.47	0.9015	32.04	0.9083	32.50	0.9145	32.90	0.9186	33.16	0.9224
16	0	30.66	0.8906	31.79	0.8973	31.76	0.8934	32.40	0.9074	32.73	0.9142
	4	32.17	0.9037	32.15	0.9067	32.28	0.9061	32.81	0.9163	32.92	0.9185
	8	32.43	0.9099	32.27	0.9100	32.58	0.9138	32.90	0.9185	32.97	0.9196
	12	32.60	0.9146	32.36	0.9135	32.77	0.9191	32.99	0.9213	33.01	0.9208
24	0	31.18	0.8835	31.88	0.8970	31.50	0.8865	32.29	0.9056	32.41	0.9077
	8	31.58	0.8951	32.31	0.9086	32.05	0.9030	32.66	0.9146	32.64	0.9138
	16	31.73	0.8999	32.41	0.9122	32.46	0.9132	32.71	0.9167	32.70	0.9156
	20	31.78	0.9020	32.41	0.9134	32.52	0.9157	32.72	0.9177	32.71	0.9160

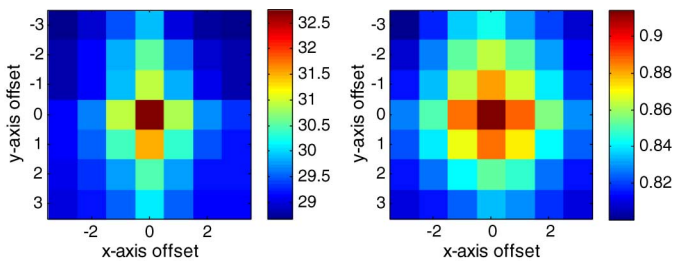


Fig. 9. The changes of PSNR (left) and SSIM (right) values with the offset of a misaligned face from original face regions at (0, 0).

image is previously extracted by manual or algorithmic operations and aligned to the training set. In order to simulate a real-world environment, we evaluate the impact of registration error to the performance of the proposed method. We firstly generated new test LR face images by translating HR face images on the horizontal and vertical directions respectively. Translation size (offset) varied from -3 to 3 pixels for the test LR face images. In this experiment, we also used the all the 40 test face images from FEI face database. As shown in Fig. 9, PSNR and SSIM indexes were inferior, as the size of the offset is large. Fig. 10 gives one of the 40 hallucinated faces when the translation size is set to different values, and we can see that the misalignment can dramatically degrade the hallucination results.

D. Robustness Against Noise

Most methods in previous work do not consider the influence of noise, and they simply assume that the LR input face image is noiseless. However, in the practical environment, the observed LR image is inevitably affected by noise. Denoising and super-resolution are usually adopted by previous approaches to offset the noise impact. However, artifacts introduced by denoising will be kept or even magnified in the latter super-resolution process. In this subsection, we show that by formulating the problem into our LcR model, our proposed method can well handle both noise impact and super-resolution simultaneously.

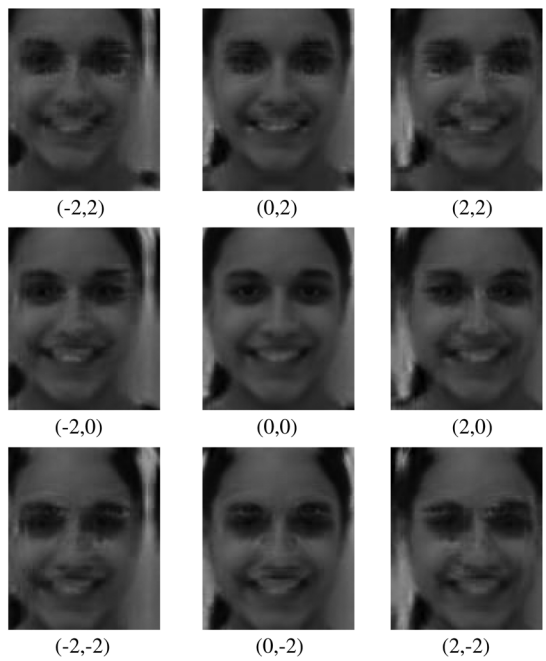


Fig. 10. Examples of face hallucination using various translation sizes. (0,0) indicates that the LR input face is well aligned.

In our experiments, we add zero mean Gaussian noises with seven different standard deviations (σ) on the LR input face images and adjust the parameters for NE (neighbor number K), SR (error tolerance ε) and LcR method (regularization parameter τ) to achieve the best performance. Note that larger patch size and more overlap pixels will lead to better noise robustness performance. For fair comparison, we choose the same patch size and overlap pixels as presented in Section V-A. With the increase of noise level, the performances of NE, LSR and DCT reduce rapidly as shown in Fig. 11. When $\sigma > 10$, they cannot remove the noise anymore. However, SR and the proposed method can remove most of the noise; when the noise is very strong, *i.e.*, $\sigma > 20$, the SR method will cause severe distortion, while the

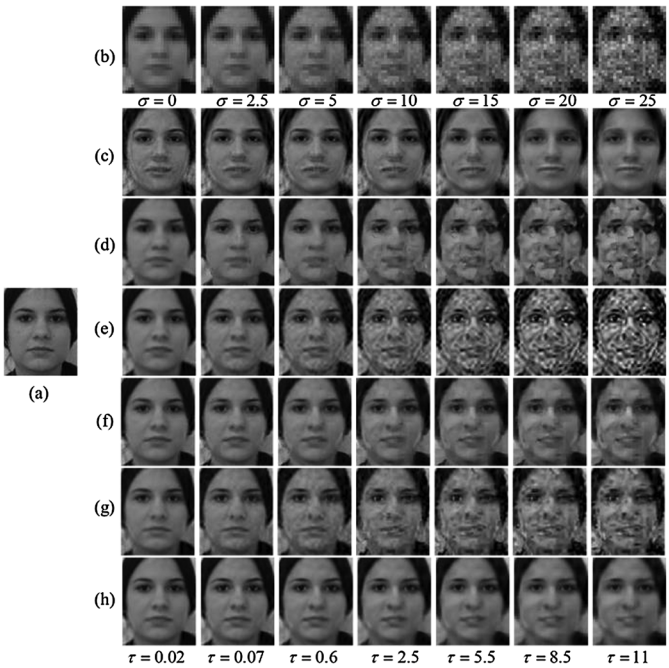


Fig. 11. Hallucinated faces with noise using different methods: (a) Original HR face image; (b) LR face images with noise; (c) Wang *et al.* [3]; (d) NE [4]; (e) LSR [7]; (f) SR [9]; (g) DCT [30]; (h) our method. Note that the values of τ are the best parameters under different noise levels of our proposed method.

LcR can maintain the primitive facial feature information as shown in Fig. 11(h), though the hallucinated face images tend to become smooth with minor blocking effects. This is consistent with the sparse representation theory which states that sparse recovery is robust against small magnitude noise in the observation. In other words, it is not feasible to achieve the noise robustness upon relying only on sparsity constraint of the weights. From Fig. 11, we also find that Wang *et al.*'s global face reconstruction method can well preserve the characteristics of human face and remove the noise well, but the hallucinated face images are dirty and different from the ground truth especially when the noise increases.

From Fig. 11(h), we learn that the regularization parameter τ plays an important role in removing noise. The value of τ depends on the noise level of the LR input faces. To be specific, we can give a larger value to τ to gain a good performance as the LR input face images become noisier (For more details about the relationship between τ and σ, please refer to the Appendix A). We can also explain this potential of denoising of LcR from another perspective: an extremely noisy LR input image patch. In that situation, it is quite sensible to replace the noisy patch with similar "clean" ones rather than synthesize and "explain" the noisy patch as in LSR and SR. What must be emphasized here is that the number of selected patch must be small (*sparsity*) and the selected patch similar to the input one (*locality*), which are the typical characteristics of LcR discussed in Section IV.

In this paper, we attribute the robustness against noise of LcR to the capability of adaptively selecting the neighbor patches. Specifically, by adaptively choosing the neighbor patches, the proposed LcR approach can choose the most relevant patches to avoid over- or under-fitting while giving sharper contours and richer details. When compared with the approach that uses a fixed number of neighbors for reconstruction (we denote this approach by K-NN, and it differs from Chang's NE method, which doesn't consider the position prior of human faces), our method is much more robust against noise. In the following, we will give some comparative experiments (adaptive K versus fixed K) and analysis to validate the merit of choosing adaptive K neighbor patches. As in Section IV-C, we conduct the experiments on the 360 training images and 40 test images. Fig. 12 shows the performance in term of PSNR under different noise levels. Note that we try different K and τ to achieve the best results for LcR and K-NN respectively. We can learn that LcR is better than K-NN in all case. In particular, the gain of LcR over K-NN is getting larger with the increase of the noise level, *e.g.*, the gains are 0.30 dB, 0.73 dB, 0.94 dB, 1.09 dB, 1.36 dB and 1.57 dB when σ varied from 0 to 25. Fig. 12 shows that, when $\sigma \geq 10$, the optimal neighbor number K for K-NN is very small, *e.g.*, 1 or 3. It once again proves that locality constraint is very important for noise robust face hallucination, since too many training samples can only lead to the reconstruction of added noise, not noise removal.

E. Experiments with Real World Image

The LR input face images of all the above experiments are formed by smoothing and down-sampling HR images, which cannot represent the true spatial feature relationship between the HR image and the degraded LR one [23]. 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 an image's structure and statistics. Therefore, in order to further testify the efficacy of our method, we perform two more experiments: i) experiment on CMU+MIT face database [29]; ii) experiment on real surveillance imaging condition image.

For CMU+MIT face database, firstly, we manually extract and align the input faces to the training samples according to the two center points of eyes, which have 25 to 45 pixels in each dimension. Secondly, we use Bicubic interpolation to enlarge those raw images to the size of 120×100 pixels (To avoid information lost because of down-sampling, we use the pre-processing method as in Section III-C to interpolate the input raw image back up to 120×100 pixels to form an LR input face image for all comparison methods). Finally, we construct the target HR face image through the respective approaches. Note that we set the values of all the parameters equal to those mentioned in Section V-C except for the regularization parameter τ, which is set with respect to the noise levels of the observation images. The results of the test images collection are shown in Fig. 13. Due to space limitations, we only give five groups of comparison results in Fig. 14. Obviously, the LcR method can produce reasonable results even though the test images are drastically different from the training examples and degraded by different noise levels. In addition, compared with five other methods, our proposed method is much more robust against noise with different levels (see the fifth row of Fig. 14, our result is very well).

Fig. 15 are pictures with a CIF-size (352×288 pixels) taken by a surveillance camera. The images in the first row are obtained in the condition of underexposure. The images in the second row are obtained when the light condition is normal and

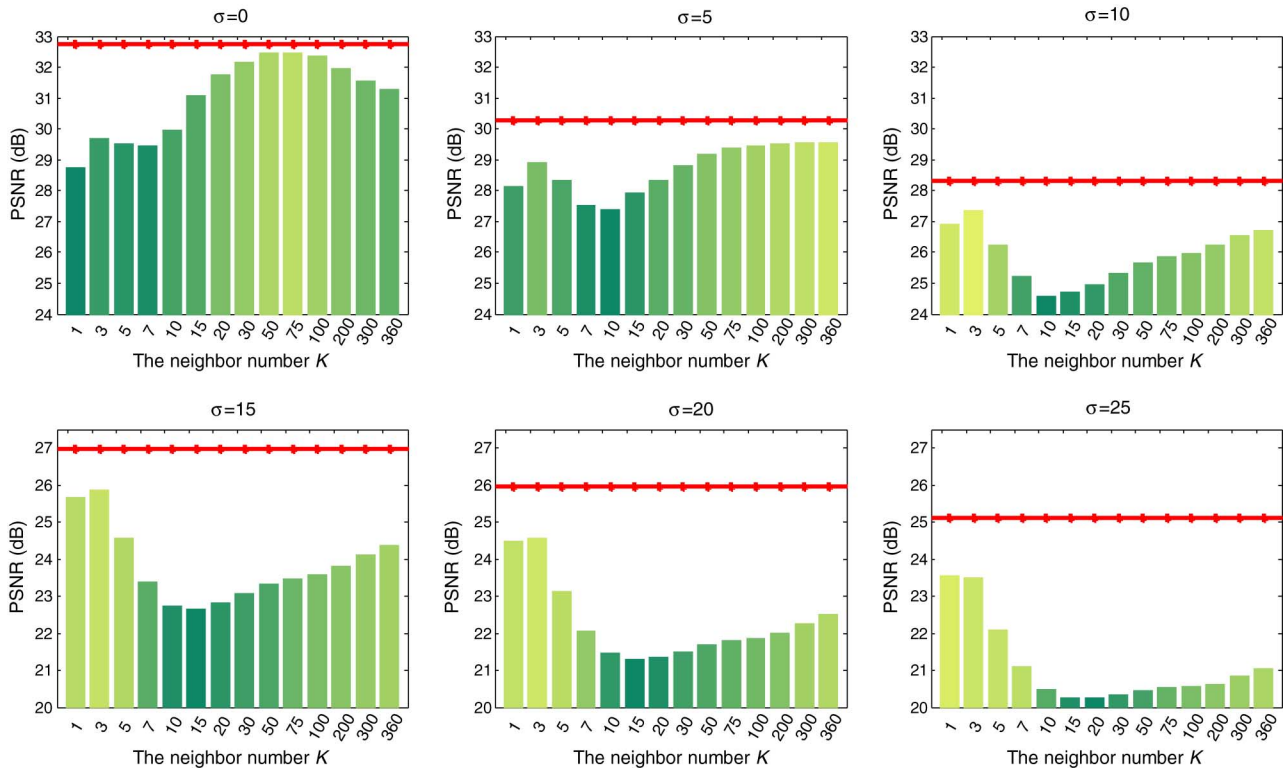


Fig. 12. PSNR comparisons between adaptive K approach and fixed K approach under different noise level σ. The red horizontal line denotes the result of the LcR method. The fall around 10 of K-NN method can be explained by that the least squares solution of neighbor embedding is "too fitted" on the LR data [33].

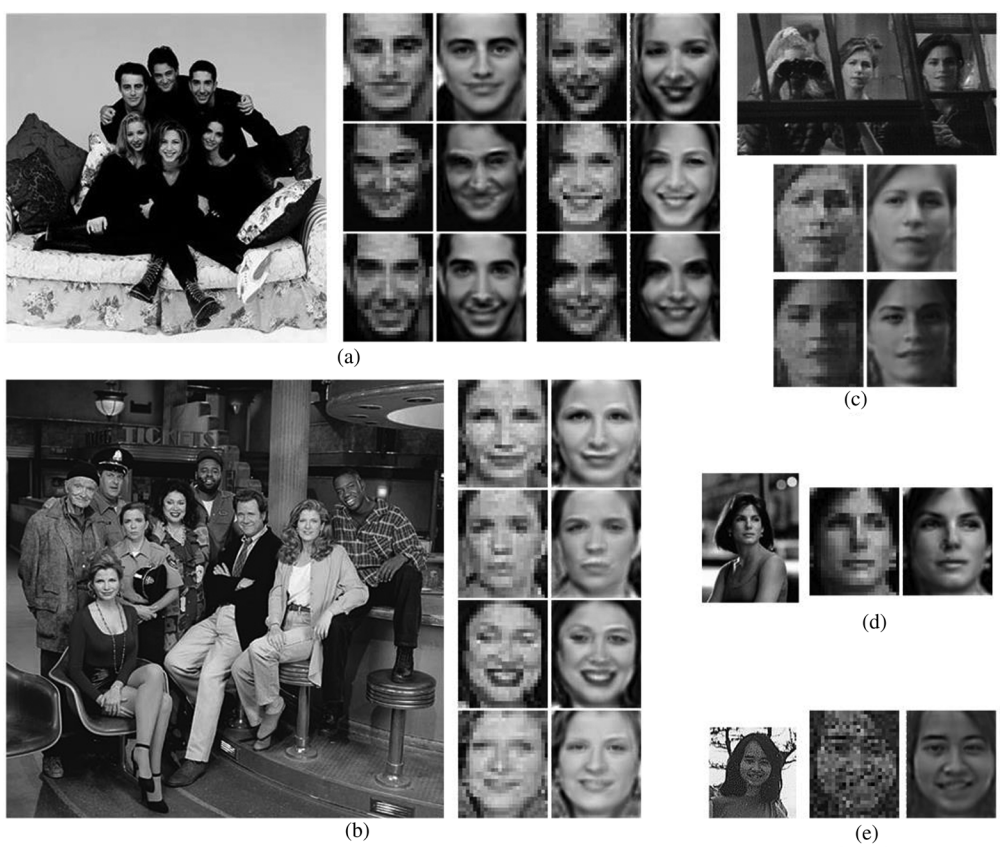


Fig. 13. Some hallucinated results of our method on CMU+MIT face database. For each example, the input image is at left, the extracted, aligned LR face in the middle, and the HR hallucinated face at the right.

Fig. 14. The comparison of visual results on five face images from CMU+MIT face database. From left to right: LR input faces, hallucinated faces by Wang *et al*. [3], NE [4], LSR [7], SR [9], DCT [30], and the proposed method.

Fig. 15. Pictures captured by surveillance camera. We extract the interest face images from the first row, which are captured in a low light and at a distance location; meanwhile, we display the "ground faces" (second row) captured in a normal light and near the camera.

the involved person is near to the camera, thus these images are used only for visual comparisons and seen as the "ground truth" of the first row. Firstly, we manually extract and align the faces of interest to the training samples according to the two center points of eyes as above, which have 50 to 63 pixels in each dimension. And then, the LR input faces are generated by converting the cropped faces to grayscale, adjusting their levels, and upsampling to the size of 120×100 pixels. The first column of Fig. 16 shows the four LR input faces from the surveillance camera. As for our proposed method, we set the values of all the parameters equal to those mentioned in Section V-C except for the regularization parameter τ, which is set to 0.6 in our experiments.

Fig. 16 compares visual results of different methods on four real surveillance LR images. When the surveillance face images are of low-quality (noisy and blurring), the proposed LcR method performs quite well for the hallucinating task. In addition, we can also see that, for non-Gaussian noise, Wang *et al.*'s global face method doesn't work. LSR adds noise to the hallucinated results rather than removes it and SR's results are very dirty. NE, DCT and our method can suppress the noise, and

Fig. 16. The comparison of visual results on surveillance images of different methods. From left to right: LR input faces, hallucinated faces by Wang *et al*. [3], DCT [30], NE [4], LSR [7], SR [9], the proposed method and the "ground truth".

we attribute this to the locality of these two representation approaches. However, due to the over- or under- fitting solution of NE and DCT, their hallucinated faces have serious blocking-artifact. In conclusion, our proposed LcR method achieves the best performance. It removes most of the noise and is to a great extent similar to the "ground truth".

VI. Conclusion and Future Work

In this paper, we propose a novel patch representation algorithm, called Locality-constrained Representation (LcR), for face hallucination. It imposes a locality constraint together with sparsity constraint onto least square inversion problem, aiming at obtaining the optimal representation of one image patch. In this method, each patch is represented by a small number of bases, which are adaptively selected from its neighborhoods, thus achieving sparsity and locality simultaneously. Experimental results on some public databases and surveillance images have demonstrated the superiority of the proposed method over some state-of-the-art methods.

However, there are several problems that need to be investigated in the future:

- Experiments show that the locality prior is vital for patch representation. However, in this paper, we measure the similarity (locality) between the LR input patch and the LR training patches by pair wise Euclidean distances. This fails to discover the intrinsic geometrical structure of the data set, which is essential to the real applications [34]. Therefore, designing a universal distance measurement algorithm, which can exploit the intrinsic manifold structure (accurate locality prior) of training sample patches, should be our future work.
- Note that the overlap patch representation and reconstruction is very time consuming, and this leads to the difficulty of our method in practical applications, *e.g.*, face recognition and real time 3D face synthesis. Thanks to the independence of the reconstruction of each target HR patch, we can straightforward to accelerate the algorithm via parallel computation.

APPENDIX A

In the following, we explain the relationship between the parameter τ and the noise level σ of the input data. Let $w_D(i,j) = Dw(i,j)$, the objective function (8) can be rewritten as

$$w_D^*(i,j) = \arg\min_{w_D^*(i,j)} \left\{ \|X(i,j) - Y_D(i,j) w_D(i,j)\|_2^2 + \tau \|w_D(i,j)\|_2^2 \right\}, \quad (18)$$

where $Y_D(i,j) = Y(i,j)D^{-1}$ and the optimal weight can be obtained by $w(i,j) = D^{-1}w_D(i,j)$. When it does not lead to a misunderstanding, we drop the term (i,j) for convenient. Eq. (18) can be rewritten as

$$w_D^* = \arg\min_{w_D^*} \left\{ \|X - Y_D w_D\|_2^2 + \tau \|w_D\|_2^2 \right\}. \quad (19)$$

We reformulate the objective function (19) from the Bayesian framework, and the optimal weight is estimated by

$$\begin{aligned} w_D^* &= \arg\max_{w_D} \{\log P(w_D|X)\} \\ &= \arg\min_{w_D} -\log P(w_D) - \log P(X|w_D). \end{aligned} \quad (20)$$

Here, $P(X|w_D)$ is the conditional probability and $P(w_D)$ is prior distribution of w_D. Note that X is an observation, $P(X)$ is a constant and it can be ignored in Eq. (20). In order to compute the MAP estimation, we assume the observation is contaminated with additive Gaussian noise of standard deviation, we have

$$P(X|w_D) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} \|X - Y_D w_D\|_2^2 \right). \quad (21)$$

The prior distribution of $P(w_D)$ is characterized by an i.i.d. zero-mean Gaussian probability model

$$P(w_D) = \frac{1}{\upsilon\sqrt{2\pi}} \exp\left(-\frac{1}{2\upsilon^2} \|w_D\|_2^2 \right), \quad (22)$$

where υ is the standard deviation of w_D. By plugging $P(X|w_D)$ and $P(w_D)$ into Eq. (20), we could readily derive

$$\tau = \frac{\sigma^2}{\upsilon^2}. \quad (23)$$

Suppose the standard deviation υ of w_D is fixed, the more noisy of the LR input face image (σ is bigger), the larger of the value τ should be.

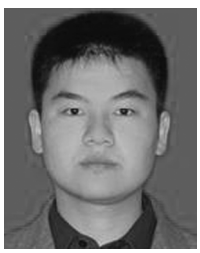
ACKNOWLEDGMENT

The authors are very grateful to Wei Zhang who is the author of [30] and [31] for providing the source code of his method. The authors would like to express the sincere gratitude for the invaluable comments and constructive suggestions by anonymous reviewers.

REFERENCES

[1] 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, 2002.
[2] C. Liu, H. Shum, and W. Freeman, "Face hallucination: Theory and practice," *Int. J. Comput. Vision*, vol. 7, no. 1, pp. 115–134, 2007.
[3] X. Wang and X. Tang, "Hallucinating face by eigen-transformation," *IEEE Trans. Syst., Man, Cybern. C*, vol. 35, no. 3, pp. 425–434, 2005.
[4] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2004, pp. 275–282.
[5] 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, 2009.
[6] X. Ma, J. Zhang, and C. Qi, "Position-based face hallucination method," in *Proc. IEEE Int. Conf. Multimedia and Expo. (ICME)*, 2009, pp. 290–293.
[7] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recognit.*, vol. 43, no. 6, pp. 3178–3194, 2010.
[8] J. Yang, H. Tang, Y. Ma, and T. Huang, "Face hallucination via sparse coding," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2008, pp. 1264–1267.
[9] 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, 2011.
[10] X. Ma, H. Huang, S. Wang, and C. Qi, "A simple approach to multi-view face hallucination," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 579–582, 2010.
[11] J. Park and S. 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, 2008.
[12] K. Jia and S. Gong, "Generalized face super-resolution," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 873–886, 2008.
[13] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2009, pp. 2223–2231.
[14] C. Thomaz and G. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vision Comput.*, vol. 28, no. 6, pp. 902–913, 2010.
[15] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
[16] E. Candes and J. Rombergt, "ℓ^1-Magic: Recovery of sparse signals via convex programming," 2005 [Online]. Available: http://www.acm.caltech.edu/l1magic/
[17] K. Su, Q. Tian, N. Sebe, and J. Ma, "Neighborhood in single-frame image super-resolution," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, 2005, pp. 1122–1125.
[18] D. Zonoobi, A. A. Kassim, and Y. V. Venkatesh, "Gini index as sparsity measure for signal reconstruction from compressive samples," *IEEE J. Select. Topics Signal Process.*, vol. 5, no. 5, pp. 927–932, 2011.
[19] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
[20] W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision," *Int. J. Comput. Vision*, vol. 40, no. 1, pp. 25–47, 2000.
[21] S. Baker and T. Kanade, "Hallucinating faces," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition (FG)*, 2000, pp. 83–88.
[22] A. N. Ayan Chakrabarti, Rajagopalan, and Rama Chellappa, "Super-resolution of face images using kernel PCA-based prior," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 888–892, 2007.
[23] 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, 2010.
[24] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.
[25] D. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006.
[26] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.
[27] J. Jiang, R. Hu, Z. Han, T. Lu, and K. Huang, "Position-patch based face hallucination via locality-constrained representation," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, 2012, pp. 212–217.

[28] X. Gao, K. Zhang, D. Tao, and X. Li, "Joint learning for single image super-resolution via coupled constraint," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 469–480, 2012.
[29] 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, 1998.
[30] W. Zhang and W.-K. Cham, "Hallucinating face in the DCT domain," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2769–2779, 2011.
[31] W. Zhang and W.-K. Cham, "Learning-based face hallucination in DCT domain," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1–8.
[32] C.-Y. Yang, S. Liu, and Y. Hsuan, "Structured face hallucination," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 1099–1106.
[33] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi, "Low-complexity single-image super-resolution based on non-negative neighbor embedding," in *Proc. British Machine Vision Conf. (BMVC)*, 2012, pp. 1–10.
[34] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schlkopf, "Ranking on data manifolds," *Adv. Neural Inf. Process. Syst.*, pp. 169–176, 2004.



Junjun Jiang received his B.S. degree in Information and Computing Science from School of Mathematical Sciences, Huaqiao University, Quanzhou, China, in 2009. He is currently pursuing his Ph.D. degree in 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 super-resolution, image interpolation, and face recognition.

Ruimin Hu received the B.S. and M.S. degrees from Nanjing University of Posts and Telecommunications, Nanjing, China, in 1984 and in 1990, and Ph.D. degree in Communication and Electronic System from Huazhong University of Science and Technology, Wuhan, China, in 1994. Dr. Hu is the director of National Engineering Research Center for Multimedia Software, Wuhan University and Key Laboratory of Multimedia Network Communication Engineering in Hubei province. He is Executive Chairman of the Audio Video coding Standard (AVS) workgroup of China in Audio Section. He has published 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. Dr. Wang is now an associate professor with School of Computer, Wuhan University, Wuhan, China. He is 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 etc.

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