SRLSP: A Face Image Super-Resolution Algorithm Using Smooth Regression With Local Structure Prior

Junjun Jiang, *Member, IEEE*, Chen Chen, Jiayi Ma, *Member, IEEE*, Zheng Wang, Zhongyuan Wang, *Member, IEEE*, and Ruimin Hu, *Senior Member, IEEE*

***Abstract*—The performance of traditional face recognition systems is sharply reduced when encountered with a low-resolution (LR) probe face image. To obtain much more detailed facial features, some face super-resolution (SR) methods have been proposed in the past decade. The basic idea of a face image SR is to generate a high-resolution (HR) face image from an LR one with the help of a set of training examples. It aims at transcending the limitations of optical imaging systems. In this paper, we regard face image SR as an image interpolation problem for domain-specific images. A missing intensity interpolation method based on smooth regression with a local structure prior (LSP), named SRLSP for short, is presented. In order to interpolate the missing intensities in a target HR image, we assume that face image patches at the same position share similar local structures, and use smooth regression to learn the relationship between LR pixels and missing HR pixels of one position patch. Performance comparison with the state-of-the-art SR algorithms on two public face databases and some real-world images shows the effectiveness of the proposed method for a face image SR in general. In addition, we conduct a face recognition experiment on the extended Yale-B face database based on the super-resolved HR faces. Experimental results clearly validate the advantages of our proposed SR method over the state-of-the-art SR methods in face recognition application.**

***Index Terms*—Face image super-resolution (SR), face recognition, local structure prior (LSP), low-resolution (LR), smooth regression.**

I. Introduction

IMAGES with high quality and high-resolution (HR), which means that objects in the images are sharp and finely detailed, have many applications in remote sensing [1]–[3], medical diagnostic [4], intelligent surveillance [5]–[7], and so on. An HR image can offer more details than its low-resolution (LR) counterpart and these details may be critical in many applications. However, due to the limitations on generation, storage, and transmission of high-quality images, face images appear in an LR form in many cases, e.g., LR face images captured by surveillance cameras [8], [9]. Therefore, in order to gain more details, it is necessary to infer an HR image from one or a series of LR images. This technique is called super-resolution (SR) [10]. It is a very active research area in computer vision and machine learning since it offers the promise of overcoming some of the inherent resolution limitations of low-cost electronic imaging systems (e.g., cell phone cameras and surveillance cameras) and better utilization of the growing capability of HR displays (e.g., HD LCDs). Currently, image SR methods can be divided into three categories: functional-interpolation methods, reconstruction-based methods, and learning-based methods.

These three categories of image SR methods all have their advantages and limitations: (i) the reconstruction fidelity of functional-interpolation methods and reconstruction-based methods are better than that of learning-based methods, while the magnification ratio of functional-interpolation methods and reconstruction-based methods is smaller than that of the learning-based methods; (ii) compared with reconstruction-based and learning-based methods, functional-interpolation methods are more computationally efficient and they are simple and easy to implement; (iii) functional-interpolation methods and reconstruction-based methods mainly focus on exploring prior information from internal example (i.e., input LR image), while learning-based methods employ external examples (i.e., a universal set of example images) as additional information to predict missing (high-frequency) information for HR images.

Motivation and contributions: Combining advantages of the above three categories of methods, in this paper we propose a novel face image SR method, namely smooth regression with local structure prior (SRLSP for short). On one hand, it adopts reconstruction constraints to ensure consistency between the reconstructed image and the input image; on the other hand, it adaptively utilizes both external and internal examples for the face image SR task. More specifically, it uses the statistical properties (by smooth regression) of the facial images in a training set as well as patch structure information (by local structure prior (LSP)) of the input LR face image to infer the missing HR pixel information. Fig. 1 presents the schematic diagram of the proposed SRLSP algorithm. In the training phase, we extract LR patches (illustrated as black circles) and missing HR pixels (illustrated as white circles) to form the training pairs. In

Manuscript received July 30, 2015; revised December 22, 2015, March 14, 2016, and July 1, 2016; accepted July 29, 2016. Date of publication August 18, 2016; date of current version December 14, 2016. This work was supported by the National Natural Science Foundation of China under Grant 61501413, Grant 61503288, and Grant 61671332, by the Fundamental Research Funds for the Central Universities at China University of Geosciences (Wuhan) under Grant CUGL160412, by the China Postdoctoral Science Foundation under Grant 2016T90725, and by the Natural Science Fund of Hubei Province under Grant 2015CFB406. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Martha Larson. *(Corresponding author: Jiayi Ma.)*

J. Jiang is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China, and also with the Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China (e-mail: junjun0595@163.com).

C. Chen is with the Center for Research in Computer Vision, University of Central Florida, Orlando, FL 32816 USA (e-mail: chenchen870713@gmail.com).

J. Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jyma2010@gmail.com).

Z. Wang, Z. Wang, and R. Hu are with the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan 430072, China (e-mail: wangzwhu@whu.edu.cn; wzy_hope@163.com; hrm1964@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.2016.2601020

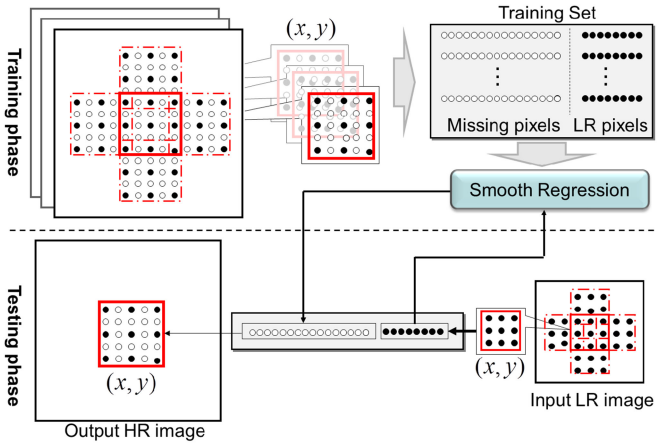


Fig. 1. Schematic diagram of the proposed SRLSP approach. The black circles are known LR pixels and the white circles are missing pixels to be interpolated.

the testing phase, we introduce a smooth regression model to construct the relationship between an LR patch and missing HR pixels with LSP. Thus, the missing HR pixel information can be predicted by the learned SRLSP model. The proposed method has the following distinct features:

1) Instead of learning a linear regression function for the entire face, we choose to learn a smooth mapping for each position-patch by introducing a weight matrix. Therefore, the learned smooth regression can be tuned towards a specific region (position-patch) of the input LR face image.
2) By exploiting the structure prior of human face, the proposed method is able to obtain more reasonable and reliable reconstruction results from external and internal examples than these methods that learn statistical properties from an external training set only.
3) Traditional local patch representation-based face SR methods use strong regularization of "same representation" for learning. In our method, we relax the "same representation" assumption to learn the regression relationship between LR and HR images, thus providing more flexibility to the learned regression function.
4) Since our proposed method is an interpolation-based approach, it meets all the reconstruction constraints needed to ensure the consistency between reconstructed HR image and input LR image. Therefore, the reconstructed results are credible.

A. Organization of This Paper

The rest of this paper is organized as follows. Section II is the related work. Section III presents the proposed SRLSP method. We detail the difference between the proposed method and prior works in Section IV. Experimental results and analysis are presented in Section V. Finally, Section VI concludes the paper.

II. Related Work

In this section, we will review some related work on generic and domain specific image SR and simultaneous face SR and recognition methods.

Functional interpolation methods apply a fixed polynomial approximation model or adaptive-structure kernels on an input LR image to estimate unknown pixels in an HR grid and obtain a processed image [11]–[13]. However, in many cases, the reconstructed images are unsatisfactory due to aliasing, blocking and blurring artifacts.

On the other hand, *reconstruction-based methods* usually use a set of consecutive LR frames of the same scene to generate one or a sequence of HR images. Accurate image registration is a crucial step to the success of reconstruction-based SR methods [14], [15]. In addition, as reported in Lin *et al.*'s work [16], under practical conditions, the magnification factor of reconstruction-based methods is limited to no more than two if denoising and registration are not good enough.

Recently, *learning-based SR methods* have received substantial attention. They assume that high-frequency details lost in an LR image can be predicted from a training data set. These methods can be broadly categorized into two major classes based on their tasks: generic image SR [17]–[23] and domain-specific image SR [24]–[31]. While generic SR algorithms are developed for all kinds of images where the priors are typically based on primitive image properties such as edges and segments, domain-specific image SR algorithms focus on specific classes of images such as faces, scenes, and graphic artwork [10]. By utilizing the additional training set, learning-based SR methods exhibit strong SR capability. A comprehensive review of current advances in face image SR is given in [5]. In this paper, we focus on the SR problem of face images. In order to predict the high-frequency information, domain-specific image SR algorithms aim to learn the relationship between HR and LR images or coefficients from a training set. Specifically, they can be further classified into the two categories: global face based parameter estimation methods and local patch-based restoration methods.

Global face-based parameter estimation methods take a face image as a whole and model it by some classical face models, such as principal component analysis (PCA) [25], [26], [32], locality preserving projections (LPP) [28], non-negative matrix factorization (NMF) [33] and canonical correlation analysis (CCA) [30]. These approaches are easy to implement and their performances are reasonably good. However, they often fail to recover the fine details of a face.

Local patch-based restoration methods are able to enhance the representation ability of the training set by decomposing an image into small patches. The target HR image can be inferred implicitly coding (by representing the input LR patches locally [8], [34]–[37], collaboratively [38], and sparsely [33], [39], [74]) or explicitly regression [31], [40], [41]. The implicitly coding based methods assume that image patches from an LR image and their HR counterparts share similar local geometry (manifold assumption [34]). Thus, patches in HR space can be reconstructed as a weighted average of local neighbors using the same weights as in LR space. For example, Chang *et al.* developed a neighbor embedding based super-resolution method by K-NN searching, and it is improved by Jiang *et al.* [42] through introducing the Tikhonov regularization. Ma *et al.* [38] proposed a position-patch based face image SR method that works by performing collaboratively over all training face patches of

the same position given an LR input patch. To improve the representation ability, sparsity [33], [39], [43], [44] and locality [8], [36], [45], [46] regularization terms have been incorporated into the patch coding objective function. However, these implicitly coding based methods may perform poorly when the manifold assumption, which states that LR and HR image patches share the same representation, is not satisfied due to "*one-to-many*" mapping between LR and HR images in practice [47]. To mitigate this problem, our previously proposed locality-constraint iterative neighbor embedding method explores the local structure by both considering the LR patch and HR patch manifolds instead of only considering one manifold (i.e., LR patch manifold), giving rise to improved performance compared with traditional neighbor embedding approaches. In addition, these explicitly regression based methods directly model the mapping function between LR and HR patch pairs. For example, Huang *et al.* [31] proposed to model the relationship between LR and HR images through linear regression to achieve good results.

Simultaneous face SR and recognition: Recently, some face image SR algorithms focused on the face recognition task have been introduced [48], [49]. For example, Li *et al.* [50] proposed coupled locality preserving mappings to project LR and HR face images onto a unified feature space. Based on the multi-manifold assumption, Jiang *et al.* [51] proposed a coupled discriminant multi-manifold analysis method for matching LR face images. In order to simultaneously recognize and super-resolve LR faces, Hennings Yeomans *et al.* [52] expressed constraints between LR and HR images in a regularization formulation. Jian *et al.* [53] proposed a simultaneous SR and recognition method based on singular value decomposition (SVD). In [9], Yang *et al.* suggested a joint face SR and recognition approach based on sparse representation with a learned person-specific face super-resolution model.

III. Proposed Method

A. The Image Degradation Model

To comprehensively analyze the image SR reconstruction problem, the first step is to formulate an observation model that relates the original HR image to the observed LR image. Concretely, let I_h and I_l denote an HR and corresponding LR facial images, respectively. The relationship between the original HR image I_h and an LR observation I_l can be mathematically modeled by the following expression:

$$I_l = DBI_h + n \tag{1}$$

where B is a blurring filter for the HR image, D is a matrix representing the decimation operator, and n is the additive Gaussian white noise accounting for imaging sensor noise. In this work, we only consider a special case of the model, in which the blurring operator and the noise term are ignored. Then, the image degradation model becomes

$$I_l = DI_h. \tag{2}$$

After applying the decimation operator to an HR facial image I_h, we can obtain its corresponding LR face image I_l.

B. Local Structure Prior (LSP)

Given an LR image observation I_l, there are infinitely solutions I_h satisfy (2). In other words, many HR face images will produce the same LR face image after image degradation. This is a "*many-to-one*" mapping between the HR and LR images that cannot be inverted without additional constraints. Mathematically, it is an ill-posed inverse problem and does not have a unique solution [33]. To obtain a reasonable HR image I_h, prior constraints such as smoothness, shape semantics [54], and sparse representation [33], should be used.

In this paper, we assume that pixels fall into different classes such as object edges with different orientations and flat areas, and each class of pixels requires specific treatment. Particularly, for a class of highly structured objects, such as human faces, although they are different from a global point of view, there is a significant local similarity between two well-aligned faces [55]. Therefore, we introduce a novel framework that uses the local structure characteristic of facial images as a constraint to construct the facial image interpolation model.

Human faces are highly structured. Upon cropping (to the same size) and aligning (by the eye centers), patches at the same position on all facial images will have same local structure. The structure is characterized by the relationship between LR pixels and missing HR pixels in a patch (x, y)

$$p_1(x, y) = F_{(x,y)}(p_0(x, y)) \tag{3}$$

where (x, y) indicates the patch position on a facial image, $p_0(x, y)$ and $p_1(x, y)$ are pixel vectors. $F_{(x,y)}$ is the regression function for patch (x, y). If the regression function of each patch position is obtained, we can construct the HR facial image by interpolating the missing pixels from the LR input facial image.

To address the facial image interpolation problem using the LSP, we divide the procedure into two steps. First, for each patch, we learn the LSP characterized by $F_{(x,y)}$ with the help of a set of LR and HR training face image patch pairs, $\{I_L^i\}_{i=1}^N$ and $\{I_H^i\}_{i=1}^N$, where N denotes the training set size. LR image patches and the missing HR pixels are represented by two sets, $\{p_0^i(x, y)\}_{i=1}^N$ and $\{p_1^i(x, y)\}_{i=1}^N$, $1 \le x \le u$, $1 \le y \le v$. u and v are the numbers of patches in row and column, respectively. Next, we use the interpolation function $F_{(x,y)}$ to interpolate an HR facial image $I_H^t = \{p_1^t(x, y)\}$, from an LR input $I_L^t = \{p_0^t(x, y)\}$. Here, the subscript "t" is used to distinguish the test sample from training samples.

C. Facial Image Interpolation via Smooth Regression With Local Structure Prior (SRLSP)

The simplest way to define the regression function is using linear regression as follows:

$$F_{(x,y)}(P_0(x, y)) = A^{\mathrm{T}}(x, y)P_0(x, y) \tag{4}$$

where $P_0(x, y)$ denotes the LR pixel set $P_0(x, y) = [p_o^1(x, y), p_o^2(x, y), ..., p_o^N(x, y)]$ at position (x, y), $p_o^i(x, y)$ is the LR pixel vector of the $i-$th LR training sample at position (x, y), A^{T} is the linear mapping function corresponding to the position (x, y), and T denotes the matrix transpose.

Although highly structured faces are very similar to each other, slight nuances always exist in different human faces, and $F_{(x,y)}$ is not strictly linear. Each sample may have its own optimal $F_{(x,y)}$. Thus, we introduce a local linear regression model and fit a different linear regression for each test sample by weighting the training samples based on how close they are to the test sample.

A common approach to implement local linear regression is to take a window of a fixed width around the test sample and include only the samples within the window. This is essentially a simple 0/1 hard threshold weighting. It generally works better to have the weights change more smoothly with the distance, starting with large values and then gradually approaching to zero. Now we have the smooth regression model, which trains the relation between LR pixels (feature) and missing HR pixels (outcome) at every patch position (x, y) (note that for notational convenience we drop the patch position term (x, y) from now on) as follows:

$$\begin{aligned} A^* &= \min_A F(A) \\ &= \min_A \sum_{i=1}^{N} w_i(p_0^i)(p_1^i - A^{\mathrm{T}} p_0^i) \\ &= \frac{\sum_{i=1}^{N} w_i(p_0^i) p_1^i}{\sum_{j=1}^{N} w_j(p_0^j)}. \end{aligned} \tag{5}$$

Our proposed smooth regression model can be seen as the kernel version of linear regression, and the weights are proportional to the kernels, $w_i(p_0^i) \propto \mathrm{K}(p_0^i, p_0^t)$. Without loss of generality, we can take the constant of proportionality to be 1. In this paper, we define the weights as the following:

$$w(p_0^i) = \frac{1}{(dist(p_0^i, p_0^t))^{\alpha}} \tag{6}$$

where $dist(p_0^i, p_0^t)$ is the squared Euclidean distance between p_o^i and p_o^j and α is a smoothing parameter. From (6), we see that the weights change smoothly with the distances. w_i determines how much each observation in the data set influences the final parameter estimation. As can be seen from Fig. 2, the samples that are most similar to the sample of interest are given more weight than the samples that are most dissimilar (i.e., when $dist(p_0^i, p_0^t)$ is small, $w(p_0^i)$ is large). Specifically, when α is set to zero, w_i is equal to one for all training samples. Then the proposed method reduces to our previous proposed method [55].

Following some matrix algebraic properties, (5) can be rewritten in the following matrix form:

$$\begin{aligned} A^* &= \min_A F(A) \\ &= \min_A ||P_1 - A^{\mathrm{T}} P_0||_W \end{aligned} \tag{7}$$

where P_1 denotes the missing HR pixel set $P_1 = [p_1^1, p_2^1, ..., p_1^N]$, and $||X||_W$ is the weighted norm ($||X||_W = tr(XWX^{\mathrm{T}})$). Here $\mathrm{tr}(\cdot)$ is the trace (sum of the diagonal elements) of a matrix. The weight matrix W is a diagonal matrix that takes the form of

$$W = \begin{bmatrix} \frac{1}{(dist(p_0^1, p_0^t))^{\alpha}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{(dist(p_0^N, p_0^t))^{\alpha}} \end{bmatrix}. \tag{8}$$

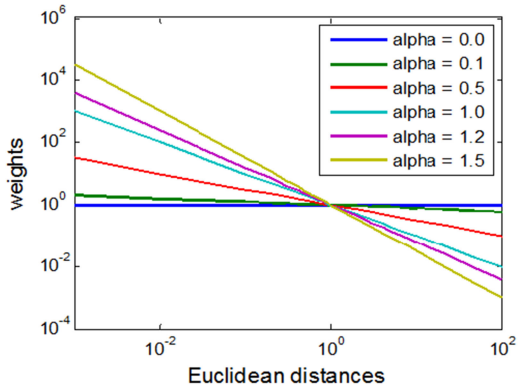


Fig. 2. (Best viewed in color and with magnification.) Plots of weights according to the squared Euclidean distances between a test sample and different training samples.

In order to make the mapping smooth, we add a regularization term to (7). Thus, we have

$$\begin{aligned} A^* &= \min_A F(A) \\ &= \min_A ||P_1 - A^{\mathrm{T}} P_0||_W + \lambda ||A||_F^2 \end{aligned} \tag{9}$$

where $||A||_F^2$ is the Frobenius norm ($||A||_F^2 = \mathrm{tr}(AA^{\mathrm{T}})$), and λ is the regularization parameter that balances the contribution of reconstruction error and the simplicity of the mapping function A. In this paper, we set the regularization parameter λ to 10^{-6} for all the experiments. From the definitions of $||X||_W$ and $||A||_F^2$, we can rewrite (9) as

$$\begin{aligned} A^* &= \min_A F(A) \\ &= \min_A (P_1 - A^{\mathrm{T}} P_0) W (P_1 - A^{\mathrm{T}} P_0)^{\mathrm{T}} + \lambda A A^{\mathrm{T}}. \end{aligned} \tag{10}$$

Obviously, the objective function in (10) is convex with respect to A. By taking the derivative of $F(A)$ with respect to A, we have (11)

$$\begin{aligned} \frac{\partial F(A)}{\partial A} &= \frac{\partial F(A)}{\partial A} \Big\{ (P_1 - A^{\mathrm{T}} P_0) W (P_1 - A^{\mathrm{T}} P_0)^{\mathrm{T}} \\ &\qquad + \lambda A A^{\mathrm{T}} \Big\} \\ &= \frac{\partial F(A)}{\partial A} \Big\{ P_1 W P_1{}^{\mathrm{T}} - P_1 W P_0{}^{\mathrm{T}} A - A^{\mathrm{T}} P_0 W P_1{}^{\mathrm{T}} \\ &\qquad + A^{\mathrm{T}} P_0 W P_0{}^{\mathrm{T}} A + 2\lambda A A^{\mathrm{T}} \Big\} \\ &= -2 P_1 W P_0{}^{\mathrm{T}} + 2 P_0 W P_0{}^{\mathrm{T}} A + 2\lambda A. \end{aligned} \tag{11}$$

By setting $\frac{\partial F(A)}{\partial A} = 0$, we then have the following equation:

$$(P_0 W P_0{}^{\mathrm{T}} + \lambda I) A = P_1 W P_0{}^{\mathrm{T}}. \tag{12}$$

Here, the term $P_0 W {P_0}^{\mathrm{T}} + \lambda I$ is invertible (non-singular) by choosing a proper λ. So the solution of (9) is

$$A = (P_0 W {P_0}^{\mathrm{T}} + \lambda I)^{-1} P_1 W {P_0}^{\mathrm{T}}. \tag{13}$$

With the mapping matrix A, we can project the observed LR pixel vector p_0^t onto A via $A^{\mathrm{T}} p_0^t$ to obtain the missing HR pixel vector p_1^t. A face image is first decomposed into smaller patches according to their positions. The patches are processed in the raster-scan order, from left to right and top to bottom. Then, the proposed smooth regression model is applied to each LR patch to predict its missing HR pixels, and the target HR patch can be connected according to the pixel positions. Finally, following [33], [34], the compatibility between adjacent patches is enforced by averaging pixel values in the overlapping regions.

IV. Relation to Prior Work

Note that our proposed SRLSP method is similar to the local linear transformation (LLT) based method proposed by Huang *et al.* [31] and our previously proposed locality-constrained representation (LcR) based method [36]. However, there are essential differences among LLT [31], LcR [36] and the proposed SRLSP method.

The key insights of our work lie in the LSP and the smooth weighting. LLT [31] and LcR [36] both learn the relationship between the LR and HR training patches, whereas our proposed SRLSP method takes into consideration the LSP of human face and learns the relationship between the LR patch and the missing HR pixels (instead of the whole HR patch as in [31]). In other words, we utilize the external and internal prior for the face image SR task simultaneously. To learn the relationship, LcR uses implicit coding based technology with the manifold assumption that the LR and HR image patches share the same representation, while LLT [31] and our proposed SRLSP method directly construct the regression model and avoid utilizing the strong regularization of "same representation" for learning. Instead of learning a linear regression relationship for each position patch in LLT [31], our proposed SRLSP method fits a different linear regression for each test sample by weighting the training samples based on how close they are to the test sample, i.e., by replacing the 1–0 hard thresholding weighting with a smooth weighting based on the distance.

V. Experimental Results

In this section, we describe the details of the extensive experiments performed to evaluate the effectiveness of the proposed method for face image SR. We compare our method with several state-of-the-art algorithms and use peak signal-to-noise ratio (PSNR) and structure similarity (SSIM) index [56] to evaluate the performance of different methods on the FEI face database [57], which will be introduced in the following subsection. In addition, we also demonstrate some objective results of different methods. In order to further verify the superiority of SRLSP over other methods, we repeat the experiments on another public face database, namely the CAS-PEAL-R1 face database [58], and analyze the influence of parameter settings, training set size, and the magnification factor. To testify the effectiveness of our proposed method, we also conduct an experiment on some real-world images from CMU+MIT face database and test the face recognition performance of super-resolved HR faces.

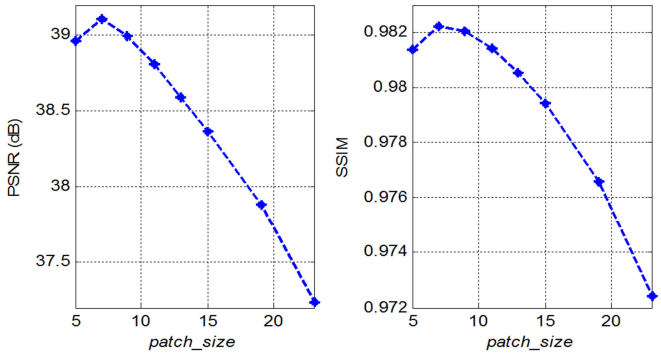


Fig. 3. Performance (PSNR and SSIM) of the proposed method using various values of $patch\_size$ on the FEI face database.

A. Database Description and Parameter Settings

The first database used in our paper is the FEI face database, which consists of 400 facial images. All the images are cropped to 120 × 100 pixels to form the HR training faces. The people in the database is mainly 19 to 40 years old with distinct appearances, e.g., hairstyles and adornments. The LR images are formed by down-sampling (by a factor of 2) the corresponding HR images, resulting in LR face images of the size 60 × 50 pixels. In our experiments, we randomly select 360 images to train the proposed smooth regression model, leaving the remaining 40 images for testing. So all the test images are absent in the training set. As for the proposed method, there are only two parameters, $patch\_size$ and the smoothing parameter α, that need to be set (note that the overlap between neighbor pathes is set according to the $patch\_size$ in our experiments, i.e., *overlap* $= patch\_size$-2). Fig. 3 shows the average PSNRs and SSIMs of all the 40 test faces using different values of $patch\_size$. Based on the performance, we can conclude that small size image patches cannot capture the structure information of a face image, while large size image patches are difficult to model. When the HR image patch size is set to 7 × 7 pixels and the corresponding LR image patch is 4 × 4 pixels, the optimal performance can be achieved. As for the smoothing parameter α, we experimentally set it to 1.2. For details about the setting of α, please refer to the following subsection.

B. Effectiveness of Smooth Weighting

To validate the effectiveness of the proposed smooth regression model for face image SR, we compare two different weighting methods including simple 0/1 hard threshold weighting and smooth weighting. The former selects K-nearest-neighbors (K-NN) to construct the linear regression model with, while the latter weights the training samples smoothly according to the distances.

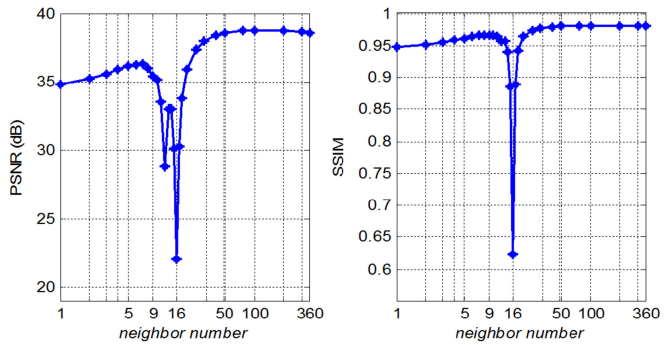


Fig. 4. Average PSNR and SSIM results versus the number of nearest neighbors K for the simple 0/1 hard threshold weighting-based regression method on the FEI face database. The best performance is achieved at $K = 100$ (PSNR $= 38.80$ dB and SSIM $= 0.9814$). It is worth noting that the input LR patch is 4 $\times$ 4 pixels (16-dimentional feature vector), and the large performance drop and instability around neighbor number $= 16$ can be explained by overfitting of the linear least squares solution to the input LR patch image.

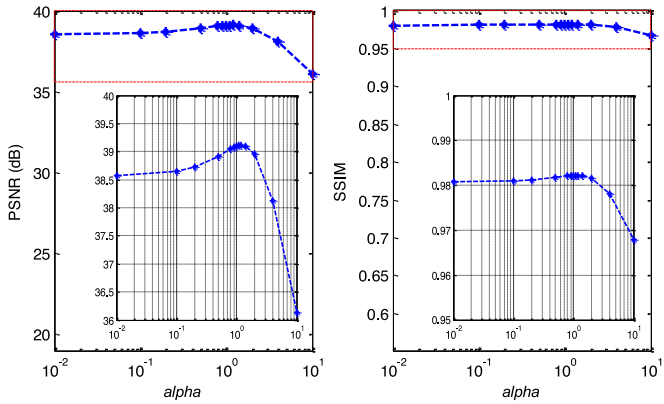


Fig. 5. Objective performance (PSNR and SSIM) with various values of α on the FEI face database. The best performance is achieved at $\alpha = 1.2$ (PSNR $=$ 39.11 dB and SSIM $= 0.9822$). In order to see more clearly, we amplify the lines in the red dashed boxes and show them on the lower right of each subfigure.

Figs. 4 and 5 show the PSNR and SSIM results varying with different neighbor number K for the 0/1 hard threshold weighting based regression method and different α for our proposed smooth regression method, respectively. As shown, the parameters K and α have influence on the performance of the respective method. When $K = 360$ or $\alpha = 0$, the weight matrix is an identity matrix, i.e., $w_i = 1$ for all training samples. Under this condition, the simple 0/1 hard threshold weighting based regression method and the proposed smooth regression method both reduce to the traditional linear regression algorithm, whose performance (PSNR $= 38.57$ dB and SSIM $= 0.9808$) is marginally worse than that of the 0/1 hard threshold weighting based regression method and our proposed smooth regression method. From Figs. 4 and 5, when compared with the 0/1 hard threshold weighting based method, the smooth regression based method is better, e.g., 0.31 dB and 0.0008 improvement in terms of PSNR and SSIM, respectively. This demonstrates the effectiveness of our proposed smooth weighting strategy for face image SR.

TABLE I
PSNR (dB) AND SSIM COMPARISON OF DIFFERENT METHODS ON THE FEI FACE DATABASE

Methods	PSNR	SSIM
Bicubic	32.44	0.9444
NEDI	28.58	0.8655
SAI	27.70	0.8607
NE	38.02	0.9758
GPR	31.34	0.9161
ANR	38.06	0.9786
EigTran	29.30	0.8127
LLT	38.02	0.9784
LSR	38.02	0.9784
SC	37.95	0.9780
LcR	38.23	0.9783
LINE	38.33	0.9783
SRLSP ($\alpha = 0$)	38.57	0.98 08
SRLSP	**39.11**	**0.9822**
Improvement	0.78	0.0038

C. Comparison Results on the FEI Face Database

In order to evaluate the superiority of the proposed method, we compare our method with several state-of-the-art algorithms including three functional interpolation based methods (i.e., Bicubic interpolation, Li *et al.*'s new edge-directed interpolation (NEDI) method [11] and Zhang *et al.*'s soft-decision adaptive interpolation (SAI) method [12]), and eight learning-based methods of which three are general image SR methods (i.e., Chang *et al.*'s neighbor embedding (NE) based method [38], He *et al.*'s Gaussian process regression (GPR) based method [59], and Timofte *et al.*'s anchored neighborhood regression (ANR) based method [20]) and the remaining five specially focus on face image SR (i.e., Wang *et al.*'s Eigen-transformation (EigTran) based method [25], Huang *et al.*'s local linear transformation (LLT) based method [31], Ma *et al.*'s least square representation (LSR) based method [38], Yang *et al.*'s sparse coding (SC) based method [33], and our previously proposed locality-constrained representation (LcR) based method [36] and locality-constrained iterative neighbor embedding (LINE) based method [37]). In this subsection, we show the subjective and objective results of the comparison methods and the proposed method. Table I tabulates the average results of different methods. In this table, SRLSP ($\alpha = 0$) denotes the proposed SRLSP method with $\alpha = 0$, which is our preliminary work reported in [55].

For the Bicubic interpolation method, we use the Bicubic function in Matlab. We take the source code of NEDI [11] and SAI [12] from their webpages,[1] and use the default parameter settings. These three methods all explore the prior information of the internal example (i.e., the input LR image), such as the smooth constraint, edge prior and the non-local prior. However, the prior information learned from the general images without considering the facial structure may be not suitable for human faces. Therefore, the performance of these interpolation methods is considerably lower than the other methods.

[1][Online]. Available: http://www.csee.wvu.edu/xinl/code/nedi.zip; http://www.ece.mcmaster.ca/xwu/executables/ARInterpolation.rar

We also report the results of three state-of-the-art learning-based SR methods. For NE [34] and ANR [20], we modify the original source code[2] to make them appropriate for face images. Specifically, we use the idea of neighbor embedding and anchored neighborhood regression to learn the relationship between the LR and HR patches for each position. The neighborhood number for NE [34] and ANR [20] is set to 75 and 200 respectively. For all these local patch based methods, the HR patch size is set to 12×12 pixels with an overlap of 8 pixels and the corresponding LR patch is 6×6 pixels with an overlap of 4 pixels unless otherwise stated. For GPR [59], we directly use the source code[3] from the author's personal homepage to carry out SR. NE [34] and ANR [20] have better performance than GPR [59]. We can conclude from the above results that learning from external examples (the LR and HR training pairs) is much more effective than learning from the internal example (the input LR image).

In addition, six recently proposed SR methods that specifically designed for human face images are also employed as comparison baselines. For Wang *et al.*'s EigTran method [25], we let the variance accumulation contribution rate of PCA be 99.9%. In Yang *et al.*'s SC based method [36], we set error tolerance to 1.0. As for our previous proposed LcR [36], we set the locality-constraint parameter to 0.04. Wang *et al.*'s method [25] is a global face method, and its representation ability is very limited, especially when the observed face image is very different from the training samples or when the size of the training samples is small. The coding-based methods, such as LSR [38], SC [33] and LcR [36], can achieve relatively good results by assuming that the LR and HR patch manifolds share similar local structure. However, due to the "one-to-many" mapping between the LR and HR images, the assumption may not hold in practice. LINE method [37] simultaneously explores the structure of LR and HR patch manifolds and obtains better results. Instead of using the strong regularization of "same representation" for learning, LLT [31] and SRLSP both aim at constructing the regression relationship between the LR and HR patch pairs to achieve better SR performance. The performance gain of SRLSP over LLT [31] is substantial due to: (i) LLT [31] assumes that each data point provides equally precise information and treats all the training samples equally, while SRLSP assigns less weight to the less precise measurements and more weight to more precise measurements, which can yield the most accurate parameter estimates possible; (ii) LLT [31] learns the relationship between the LR and HR patch pairs, while SRLSP incorporates the LPS as an additional constraint. In other words, SRLSP deeply exploits the facial structure prior whereas LLT [31] only considers the statistical properties. To prove this point, we list the results of SRLSP when α is set to 0. The only difference between the SRLSP ($\alpha = 0$) and LLT [31] is that the former incorporates the LPS and LLT [31] does not.

Fig. 6 shows one group of examples of the reconstructed results and the reconstruction error maps using different methods

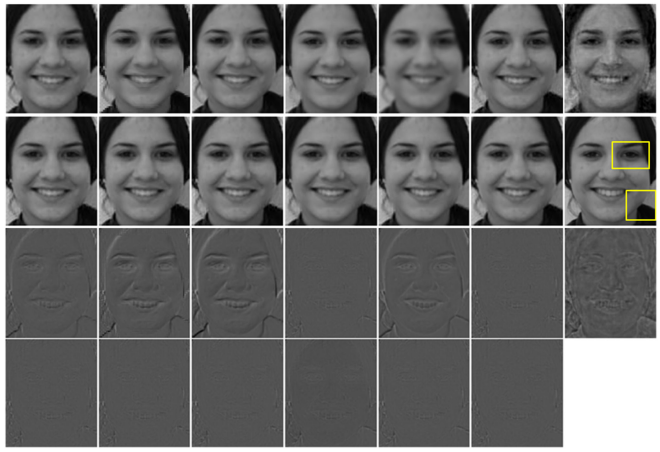


Fig. 6. One group of face images that were reconstructed from the FEI face database by different methods. From left-to-right and top-to-bottom are the super-resolved results of Bicubic interpolation, super-resolved results of NEDI [11], SAI [12], NE [34], GPR [59], ANR [20], EigTran [25], LLT [31], LSR [38], SC [33], LcR [36], LINE [37], and the proposed SRLSP, and finally the ground truth HR face image. The first two rows are the super-resolved results and the last two rows are the corresponding reconstruction error maps. (Note that the effect is more pronounced if the figure of the electronic version is zoomed, same as Fig. 7.)

(for more results, please refer to the supplementary material). In each group of images, the top-left is the input LR face, the bottom-right is the ground truth HR face, and the rest are the reconstructed HR faces based on twelve different methods. From these results, we see that the reconstructed HR faces of Bicubic interpolation are very smooth and miss many facial details, e.g., edges and corners. Wang *et al.*'s method [25] is not able to recover a clear face and the reconstructed HR faces have obvious ghosting effects. Yang *et al.*'s method [33] cannot fully recover the detailed features leading to the reconstructed HR face images having some artifacts. Compared with other methods, our method can generate better results both in global faces and fine details, especially in face contour and eyes (please refer to the yellow boxes in Fig. 6).

Simulation experiments demonstrate that our approach is able to generate HR face images with visually satisfactory global face appearance and local detailed features. The reconstructed faces are much more similar to the ground truth HR faces. We attribute this superiority of our method over other methods to the introduction of LSP and the smooth regression model.

D. Experiments on the CAS-PEAL-R1 Face Database

In addition to the FEI face database, we also conduct experiments on the CAS-PEAL-R1 face database [36] which contains 30871 images of 1040 subjects. We only use the neutral expression and normal illumination faces of each subject from the frontal subset in the experiments. In all the 1040 frontal face images, we randomly select 1000 images for training and leave the other 40 images for testing. All the images are aligned by five manually selected feature points and are cropped to 128×112 pixels through automatic alignment method [60] and robust feature matching technology [61], [73]. Similarly, none of the

[2][Online]. Available: http://www.jdl.ac.cn/user/hchang/doc/code.rar; http://www.vision.ee.ethz.ch/~timofter/software/SR_NE_ANR.zip

[3][Online]. Available: http://www.eie.polyu.edu.hk/ wcsiu/softmodule/4/GPR_v1.1.zip

TABLE II
PSNR (DB) AND SSIM COMPARISON OF DIFFERENT METHODS ON THE CAS-PEAL-R1 FACE DATABASE

Methods	PSNR	SSIM
Bicubic	28.56	0.9332
NEDI	24.87	0.8460
SAI	24.22	0.8429
NE	33.93	0.9720
GPR	27.58	0.9022
ANR	33.89	0.9724
EigTran	28.46	0.8395
LLT	33.66	0.9716
LSR	33.67	0.9715
SC	33.84	0.9713
LcR	34.03	0.9730
LINE	34.15	0.9733
SRLSP ($\alpha = 0$)	33.93	0.9737
SRLSP	**34.61**	**0.9761**
Improvement	0.46	0.0024

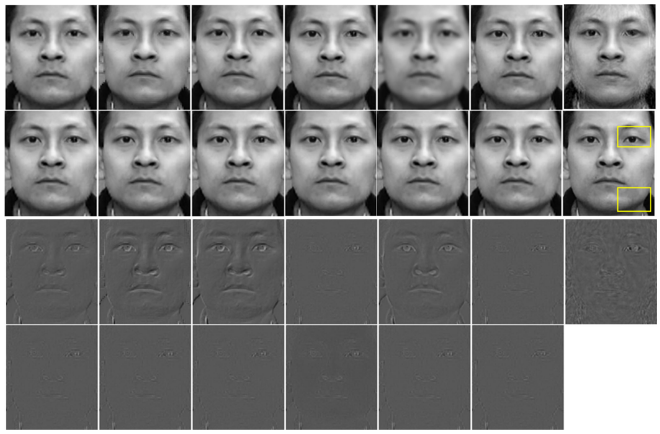


Fig. 7. One group of face images that were reconstructed from the CAS-PEAL-R1 face database by different methods. Visual comparison results on the CAS-PEAL-R1 face database.

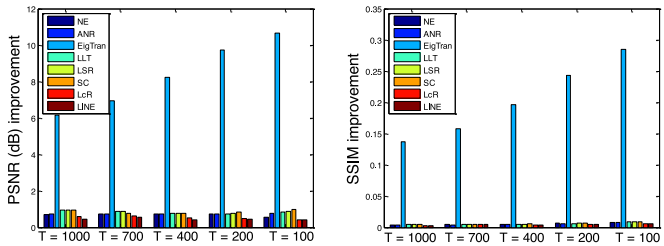


Fig. 8. (Best viewed in color and with magnification.) PSNR and SSIM improvement of SRLSP over seven comparison methods with different training sizes on the CAS-PEAL-R1 face database.

test subjects is present in the training images. The LR images are formed as described in Section V-A, thus the size of LR face images are 64 $\times$ 56 pixels. We set the values of all the parameters of SRLSP equal to those mentioned in Section V-B except for the smoothness parameter α, which is determined by carefully tuning. We choose $\alpha = 1.5$ which results in the best performance of our method (more details about setting the parameter α on the CAS-PEAL-R1 face database can be found in Section V-E). For the comparison algorithms, we experimentally set the parameters to obtain the best performance. Specifically, as for NEDI [11], SAI [12], and GPR [59], we directly use the source codes and the parameter settings therein. The neighborhood number for NE [34] and ANR [20] is set to 75 and 200, respectively. For Wang *et al.*'s global face method [25], we let the variance accumulation contribution rate of PCA be 99.9%. In Yang *et al.*'s SC method [33], we set error tolerance to 1.0. As for our previous proposed LcR [36], we set the locality-constraint parameter to 0.1. The iteration number is set to 5 and the locality parameter is set to 1e-5 in [37]. In the following, we show comparison results in terms of visual quality and objective metrics (PSNR and SSIM indexes).

1) Subjective and Objective Results Comparison: Table II tabulates the average PSNR and SSIM results of different methods. Fig. 7 presents the visual comparison of different methods (for more results, please refer to the supplementary material). SRLSP generates the best visual results (see the eyes and face contours). We can also draw the same conclusion that: (i) domain-specific image SR methods are better than generic image SR methods; (ii) smooth weighting and LSP are critical for the face image SR problem.

2) Effects of the Training Set Size on the CAS-PEAL-R1 Face Database: The above experimental results on the CAS-PEAL-R1 face database show that the smoothness regularization and the LSP are very effective in regularizing the ill-posed face image SR problem. In the experiments, we fix the training set size to be 1000. Intuitively, the larger training set should possess more representation power and thus may yield more accurate approximation at the expense of increasing the running time.

To further verify the effectiveness the proposed smooth regularization and the LSP, here we evaluate the effect of dictionary size on face image SR. We randomly select four training subsets of size 100, 200, 400 and 700, and use them to perform SR on the same 40 input LR face images described in Section V-D. In Fig. 8, we show the PSNR and SSIM improvements of SRLSP over eight comparison methods according to different training set sizes on the CAS-PEAL-R1 face database (note that the Bicubic interpolation, NEDI [11], SAI [12], and GPR [59] are independent of the training set, thus they are not considered here.). We can see from Fig. 8 that SRLSP consistently performs better than the comparison methods. The improvements of SRLSP over the other methods are more obvious as the training set size gets smaller. This effect is particularly noticeable in term of SSIM index as shown in the right figure of Fig. 8. This makes smooth regression an effective method under the condition that the training sample size is small.

3) Effects of the Smoothing Parameter α on the CAS-PEAL-R1 Face Database: Smooth regression is an important component in SRLSP. In the following, we provide a deep analysis about the influence of the smoothing parameter α which controls the smooth regression process. In Fig. 9, we show the PSNR and SSIM according to different values of α on the CAS-PEAL-R1face database with various training sizes, i.e., training size = 1000, 700, 400 and 100. The maximum PSNR or SSIM

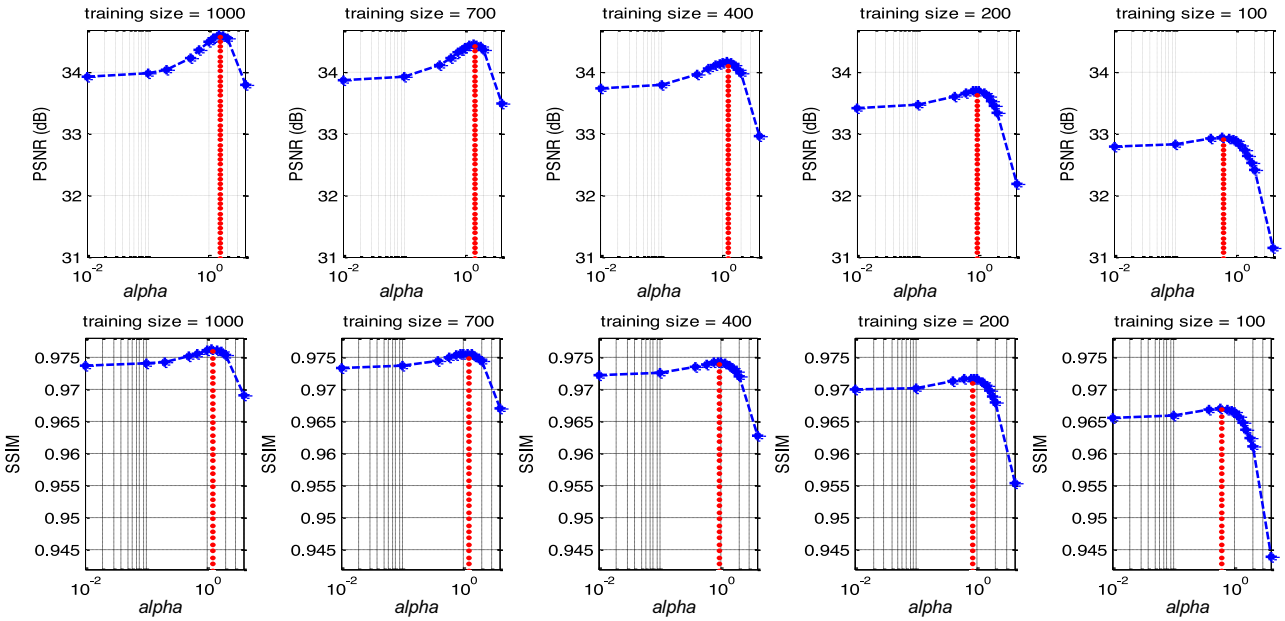


Fig. 9. Objective performance in terms of PSNR (first row) and SSIM (second row) according to different values of α on the CAS-PEAL-R1 face database with various training sizes, i.e., training size = 1000, 700, 400, and 100. The maximum PSNR or SSIM values are achieved at 1.5, 1.4, 1.2, 0.9, and 0.6 (indicated by the red lines), respectively.

values are achieved at $\alpha > 0$, which implies that the smooth constraint is essential for regression. In addition, we also find that the optimal value of α becomes smaller with the decrease of the training set size. This phenomenon could be explained as follows: when the training set size is large, the training samples exhibit rich diversity. In order to fit the observation, it is easy to find several similar samples to model the observation. By setting a large α, several neighbor samples can be chosen by giving large weights to these neighbor samples. In contrast, when the training set size is small, we may have to use as many as possible training samples to fit the observation rather than selecting only a few neighbor samples. In this case, we can set a relatively small α to gather as many samples as possible to fit the observation. This shows the characteristics of a learning-based system which requires a certain amount of similarity between test and training samples [24].

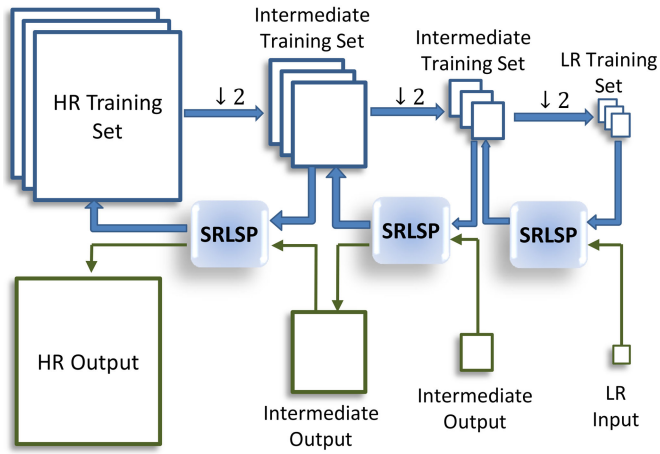


Fig. 10. 8X magnification framework of the proposed method.

E. SR Results of Very Low-Resolution Faces

In the above experiments, the down-sampling factor is set to 2 for our proposed method is essentially an interpolation-based method. Like many interpolation based methods [11], [62], our method can only amplify the input LR face image by a factor of 2 each time. To test the effectiveness of our method when the input faces are very low-resolution, e.g., 8×7 pixels (corresponding to 16X magnification) and 16×14 pixels (corresponding to 8X magnification), we conduct some experiments to super-resolve the input LR face by stepwise interpolation. Fig. 10 is an example of 8X magnification framework of the proposed method, and we super-resolve the input LR face by three SRLSP based interpolations. We compare our method with Bicubic interpolation and one global face SR method (EirTran [25]) and one position-patch based representation method (LcR [36]). Note that we randomly choose 400 images from the CAS-PEAL-R1 face database as the training set and the 40 test images are the same with Section V-D. For our method, the $patch\_size$, $overlap$ and parameter α are set to the same with experiments above. Fig. 11(a) and (b) visually compare the different reconstructions on three test images by a magnification of 8 and 16 respectively. Bicubic interpolation cannot work anymore. Results of EirTran [25] are similar to the mean face and are not be trusted. LcR [36] and the proposed method can produce reasonable HR faces but may also lose some detailed features [see Fig. 11(b)]. The average PSNRs and SSIMs of all 40 test faces obtained using different methods are shown in Table III. In terms of objective quality, we can see that the proposed method results in larger PSNR and SSIM values. This is mainly due to the proposed method simultaneously integrating external and internal

Fig. 11. Visual SR results of very low-resolution faces. (a) 8X magnification results. (b) 16X magnification results. For each subfigure, from left to right, there are the LR input faces, super-resolved faces by bicubic interpolation, EirTran [25] and our method, and the last column is the original HR faces.

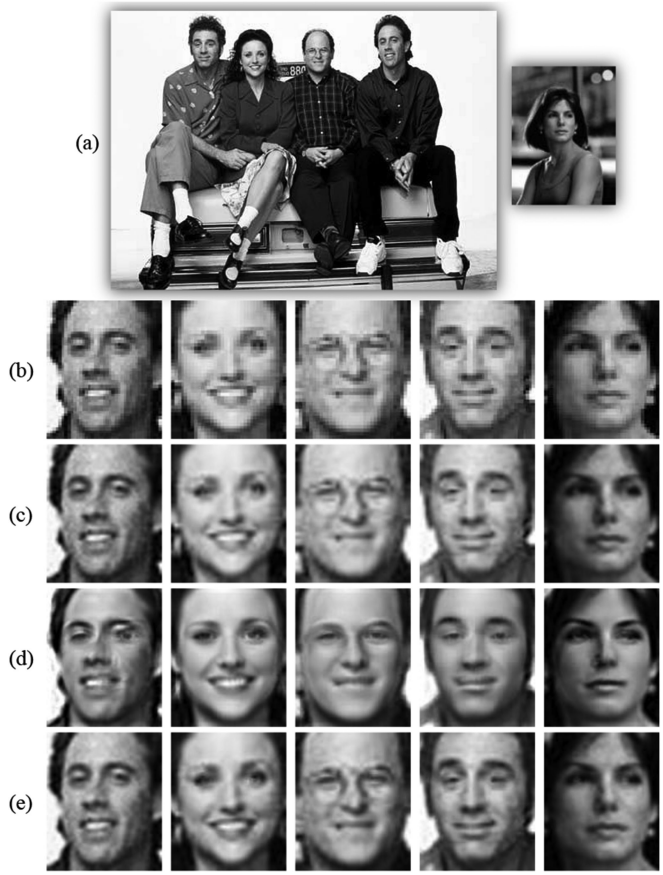


Fig. 12. Super-resolved results with some real-world images from CMU+MIT face database.

TABLE III
PSNR (dB) AND SSIM RESULTS OF VERY LOW-RESOLUTION FACES

Methods	16×14 pixels(8X)		8×7 pixels(16X)	
	PSNR	SSIM	PSNR	SSIM
Bicubic	17.11	0.5068	14.05	0.3776
EigTran	19.15	0.6196	17.75	0.5954
LcR	23.02	0.7947	19.29	0.6690
SRLSP	23.51	0.8190	19.90	0.7148

The input LR face is 16×14 pixels or 8×7 pixels.

examples, the structure prior of human face, and reconstruction constraints, which helps to better pose the obtained SR solution and produce more faithful SR recovery.

F. SR With Real-World Images

In order to further support the effectiveness of our proposed face SR method, we conduct some experiments on real-world images from CMU+MIT face database [63] as shown in Fig. 12(a). Fig. 12(b) is the extracted and aligned LR faces. Fig. 12(c)–(e) are the reconstructed HR faces by three representation methods, Bicubic interpolation, LcR [36], and our method, respectively. 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 (here we use the training faces from the FEI face database). The images reconstructed with Bicubic interpolation are too smooth (see the face contours, nose and eyes) compared with our method. The super-resolved face by LcR has an obvious "ghost effect" and is dissimilar to the input. When compared with the results on the standard face database, our super-resolved results with real-word images are much more worse. This is mainly because the actual imaging process (such as motion and defocus blur, low light, mixed noise [64], [65], and so on) is much more complex than the simple down-sampling. In addition, as one interpolation based method, the noise in input faces may also be maintained in the outputs.

G. Effect of SR on Face Recognition

Although it is logical to believe that super-resolved HR face images should be beneficial to the following face recognition task, recently, there is still doubt whether the reconstruction results measured by PSNR and root mean square error (RMSE) translate to improve face recognition. In [66], Xu *et al.* investigated the problem of how much face SR can improve face recognition. They reached the conclusion that when the resolution of the input LR faces is larger than 32×32 pixels, the super-resolved HR face images can be better recognized than the

TABLE IV
FACE RECOGNITION ACCURACY (%) ASSOCIATED WITH DIFFERENT SR METHODS

Methods	Classifier	
	NNC	SRC
HR	75.52	90.17
Bicubic	72.56	87.36
ANR	72.70	89.03
EigTran	73.43	88.82
LcR	73.62	89.13
SRLSP	74.62	89.65

LR face images; however, when the input faces have very low dimension (e.g., 8×8 pixels), some of the face SR approaches do not work properly.

In order to evaluate the effect of our proposed face SR method on the subsequent face recognition task, we conduct a face recognition experiment using the super-resolved faces from different SR methods. As in many face recognition methods [67]–[68], in this paper we conduct the face recognition experiment on the Extended Yale-B face database [69], which contains 2414 frontal images of 38 subjects under various lighting conditions. All the face images are manually aligned and cropped to 128×128 pixels, with 256 gray levels per pixel. We randomly select one fifth of the data (491 images) for training, and leave the rest (1923 images) for testing. Note that all the 1923 test images are down-sampled to 64×64 pixels. We employ five representative face SR methods mentioned above (i.e., Bicubic interpolation, ANR [20], EigTran [25], LcR [36] and our proposed SRLSP) to super-resolve the test images to HR level with 128×128 pixels. We utilize two popular classification algorithms, i.e., nearest neighbor classifier (NNC) and sparse representation classifier (SRC) [67], to query the identity of the super-resolved HR face image.

Table IV shows the recognition accuracies associated with different SR methods. In addition, the performance of directly comparing HR version of probe images against HR gallery images is given as a baseline for comparison, and is denoted by bold "**HR**". From the recognition rates we learn that the HR face images reconstructed by our proposed method lead to a better recognition result compared with the other methods. Our recognition rate is very close to the ideal case **HR**, which indicates the effectiveness of the proposed face SR method on the subsequent face recognition task. We attribute this superiority of SRLSP to its ability to maintain the input information (inherited by the interpolation method) and to learn high-frequency information from the training samples.

H. Discussion

1) General Prior Versus Domain-Specific Prior: Prior information learned from the face training samples is much more effective than information learned from the general training samples (such as houses, plants, animals, etc.). Bicubic interpolation, NEDI [11], SAI [12] and GPR [59] are general image SR methods designed for general images, and the remaining methods (EigTran [25], LLT [31], LSR [38], SC [33], LcR [36]) are approaches that learn prior information from face training samples. The performance of the former is much worse than that of the latter.

2) Global Versus Local Modeling: Position-patch based methods are better than global face methods. Wang *et al*'s global EigTran [25] can capture the global structure of face by modeling the entire face image as a whole through PCA decomposition; however, it will also result in low reconstruction precision and unsatisfactory results around the facial contour. By decomposing a complete face image into smaller patches according to the positions, the position-patch based models have higher reconstruction precision than the global model.

3) Why Smooth Weighting?: The smooth weighting strategy is important for modeling the relationship between the LR and HR training set. This can be justified by our previously proposed LcR [36] and our proposed SRLSP method. LcR [36] utilizes the smooth weighting strategy for patch representation and gives different freedom (i.e., by weighting the training samples based on how close they are to the test sample) to the training samples, while SRLSP penalizes the regression measurement by a weight that changes more smoothly with the distance. Note that NE [34] and ANR [20] also consider neighborhood information, and these two methods are essentially the simple 0/1 hard threshold weighting. SRLSP ($\alpha = 0$) does not take the smooth weighting into consideration, and its performance is worse than that of SRLSP. This also demonstrates the advantage of the smooth weighting strategy.

4) Face Structure Prior Is Very Important: In addition to incorporating the smooth weighting strategy, which can be viewed as statistical properties learned from the training set, exploiting the structure information is also crucial to the face image SR problem. As a highly structured object, human faces have a significant local similarity with each other. Therefore, structure information can be used to guide the reconstruction of face images leading to more plausible and reliable reconstruction results. This is demonstrated by LLT [31] and SRLSP ($\alpha = 0$). By exploiting the structure information, SRLSP ($\alpha = 0$) is better than LLT [31].

5) Computational Complexity: Generally speaking, learning- or example-based image SR methods can be roughly divided into two categories: regression-based methods and coding-based methods. The former can learn and save the mapping functions off-line. Therefore, we can expect faster image SR reconstruction than with the coding-based methods, which need complex coding strategies, e.g., sparse coding, for each input LR patch. Our proposed regression method is data driven. In other words, the learned mapping functions depend on the input data, e.g., the Euclidean distance between the input LR patch and the dictionary atoms. Thus, the mapping functions vary according to the observation LR patch, and we cannot learn and save the mapping functions off-line. It will take some time to calculate the Euclidean distance between the input LR patch and the dictionary atoms and the mapping function. In particular, in the Windows platform with MATLAB 7.14 (R2012a) on an Intel Core i3 CPU with 3.20 GHz and 4G memory PC, the average CPU time for each test image is around 6.9 seconds. However, thanks to the independence of the reconstruction of

each target HR patch, we can easily accelerate our method via parallel computation.

VI. Conclusion

In this paper, we have proposed a novel approach for efficient and effective facial image interpolation method, namely smooth regression with local structure prior (SRLSP). It combines the advantages of three different classes of methods, i.e., interpolation based methods, reconstruction based methods and learning-based methods, thus leading to promising SR reconstruction results. The developed SRLSP method divides each face image into small image patches, and then learns the relationship between the LR image patch and the missing HR pixel information, which can be regarded as a local structure prior (LSP). This LSP is then used to predict the missing HR pixel information of the LR observation patch. Experimental results on the FEI face database and the CAS-PEAL-R1 face database demonstrated the effectiveness of the proposed approach. Moreover, face recognition results also validate the advantages of our proposed SR method over the state-of-the-art SR methods in a face recognition application scenario.

However, there are several problems that need to be investigated in the future: Note that the overlap patch mapping and reconstruction is time consuming, which hinders our method in certain practical applications, e.g., real time face recognition and 3D face synthesis. Thanks to the independence of the reconstruction of each target HR patch, we can accelerate the algorithm via parallel computation [70]–[72]. In this article, we have focused on developing a frontal face image SR method. However, when the LR observation face is in the wild, i.e., arbitrary pose, various skin colors, and extreme ambient illumination, how can we super-resolve face image in the wild is another open question.

Acknowledgment

The authors are very grateful to Dr. M. Larson for carefully revising this paper and giving many helpful suggestions. The authors would also like to thank the anonymous reviewers for their invaluable comments and constructive suggestions.

References

[1] L. Wang, K. Lu, and P. Liu, "Compressed sensing of a remote sensing image based on the priors of the reference image," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 736–740, Apr. 2015.

[2] H. Lu *et al.*, "Reference information based remote sensing image reconstruction with generalized nonconvex low-rank approximation," *Remote Sensing*, vol. 8, no. 6, 2016, Art. no. 499.

[3] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, 2016.

[4] H. Greenspan, "Super-resolution in medical imaging," *Comput. J.*, vol. 52, no. 1, pp. 43–63, 2009.

[5] 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. 9–30, 2014.

[6] Y. Yao, B. R. Abidi, N. D. Kalka, N. A. Schmid, and M. A. Abidi, "Improving long range and high magnification face recognition: Database acquisition, evaluation, and enhancement," *Comput. Vis. Image Understand.*, vol. 111, no. 2, pp. 111–125, 2008.

[7] Z. Wang *et al.*, "Zero-shot person re-identification via cross-view consistency," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 260–272, Feb. 2016.

[8] L. An and B. Bhanu, "Face image super-resolution using 2d CCA," *Signal Process.*, vol. 103, pp. 184–194, 2014.

[9] M.-C. Yang, C.-P. Wei, Y.-R. Yeh, and Y.-C. F. Wang, "Recognition at a long distance: Very low resolution face recognition and hallucination," in *Proc. IEEE Int. Conf. Biometrics*, May 2015, pp. 237–242.

[10] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 372–386.

[11] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.

[12] X. Zhang and X. Wu, "Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 887–896, Jun. 2008.

[13] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.

[14] Y. Zhu, K. Li, and J. Jiang, "Video super-resolution based on automatic key-frame selection and feature-guided variational optical flow," *Signal Process.: Image Commun.*, vol. 29, no. 8, pp. 875–886, 2014.

[15] K. Li, Y. Zhu, J. Yang, and J. Jiang, "Video super-resolution using an adaptive superpixel-guided auto-regressive model," *Pattern Recog.*, vol. 51, pp. 59–71, 2016.

[16] 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.

[17] W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, pp. 25–47, 2000.

[18] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010.

[19] L. An and B. Bhanu, "Image super-resolution by extreme learning machine," in *Proc. IEEE Int. Conf. Image Process*, Sep.-Oct. 2012, pp. 2209–2212.

[20] R. Timofte, V. De, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. Int. Conf. Comput. Vis.*, 2013, pp. 1920–1927.

[21] Z. Zhu, F. Guo, H. Yu, and C. Chen, "Fast single image super-resolution via self-example learning and sparse representation," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2178–2190, Dec. 2014.

[22] Y. Zhang, J. Liu, W. Yang, and Z. Guo, "Image super-resolution based on structure-modulated sparse representation," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2797–2810, Sep. 2015.

[23] S. Ye, C. Deng, J. Xu, and X. Gao, "Coupled fisher discrimination dictionary learning for single image super-resolution," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 1196–1200.

[24] 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 Conf. Comput. Vis. Pattern Recog.*, Dec. 2001, vol. 1, pp. 192–198.

[25] 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.

[26] A. Chakrabarti, A. 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.

[27] S. W. Park and M. Savvides, "Breaking the limitation of manifold analysis for super-resolution of facial images," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Apr. 2007, vol. 1, pp. I-573–I-576.

[28] 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 Recog.*, Sep. 2008, pp. 1–8.

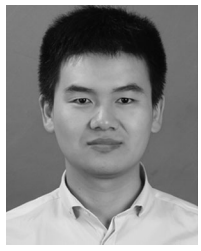
[29] Y. Zhuang, J. Zhang, and F. Wu, "Hallucinating faces: Lph super-resolution and neighbor reconstruction for residue compensation," *Pattern Recog.*, vol. 40, no. 11, pp. 3178–3194, 2007.

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

[31] H. Huang and N. Wu, "Fast facial image super-resolution via local linear transformations for resource-limited applications," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 1363–1377, Oct. 2011.

[32] 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.

[33] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
[34] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun.-Jul. 2004, vol. 1, pp. 275–282.
[35] Y. Hu, K.-M. Lam, G. Qiu, and T. Shen, "From local pixel structure to global image super-resolution: A new face hallucination framework," *IEEE Trans. Image Process.*, vol. 20, no. 2, pp. 433–445, Feb. 2011.
[36] 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.
[37] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4220–4231, Oct. 2014.
[38] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recog.*, vol. 43, no. 6, pp. 2224–2236, 2010.
[39] 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.
[40] G. Gao and J. Yang, "A novel sparse representation based framework for face image super-resolution," *Neurocomputing*, vol. 134, pp. 92–99, 2014.
[41] C.-T. Tu and J.-R. Luo, "Robust face hallucination using ensemble of feature-based regression functions and classifiers," *Image Vis. Comput.*, vol. 44, pp. 59–72, 2015.
[42] J. Jiang, C. Chen, K. Huang, Z. Cai, and R. Hu, "Noise robust position-patch based face super-resolution via Tikhonov regularized neighbor representation," *Inf. Sci.*, vols. 367–368, pp. 354–372, 2016.
[43] Z. Wang, R. Hu, S. Wang, and J. Jiang, "Face hallucination via weighted adaptive sparse regularization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 802–813, May 2014.
[44] X. Ma, H. Song, and X. Qian, "Robust framework of single-frame face superresolution across head pose, facial expression, and illumination variations," *IEEE Trans. Human–Mach. Syst.*, vol. 45, no. 2, pp. 238–250, Apr. 2015.
[45] 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.
[46] J. Jiang, R. Hu, Z. Wang, Z. Han, and J. Ma, "Facial image hallucination through coupled-layer neighbor embedding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1674–1684, Sep. 2016.
[47] K. Su, Q. Tian, Q. Xue, N. Sebe, and J. Ma, "Neighborhood issue in single-frame image super-resolution," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2005, p. 4.
[48] Z. Wang, Z. Miao, Q. J. Wu, Y. Wan, and Z. Tang, "Low-resolution face recognition: A review," *Vis. Comput.*, vol. 30, no. 4, pp. 359–386, 2014.
[49] W. Zou and P. Yuen, "Very low resolution face recognition problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 327–340, Jan. 2012.
[50] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Process. Lett.*, vol. 17, no. 1, pp. 20–23, Jan. 2010.
[51] J. Jiang, R. Hu, Z. Wang, and Z. Cai, "CDMMA: Coupled discriminant multi-manifold analysis for matching low-resolution face images," *Signal Process.*, vol. 124, pp. 162–172, 2016.
[52] P. H. Hennings-Yeomans, S. Baker, and B. V. Kumar, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2008, pp. 1–8.
[53] M. Jian and K. Lam, "Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1761–1772, Nov. 2015.
[54] C. Lan, R. Hu, K. Huang, and Z. Han, "Face hallucination with shape parameters projection constraint," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 883–886.
[55] R. Xing, J. Fu, Y. Shao, and J. You, "Rigid regression for facial image interpolation with local structure prior," in *Proc. 6th Int. Conf. Intell. Human–Mach. Syst. Cybern.*, 2014, vol. 2, pp. 67–70.
[56] 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, Apr. 2004.
[57] C. E. Thomaz and G. A. 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.
[58] 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.
[59] H. He and W. Siu, "Single image super-resolution using gaussian process regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 449–456.
[60] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.
[61] J. Ma *et al.* and "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.
[62] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.
[63] H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 23–38, Jan. 1998.
[64] C. L. P. Chen, L. Liu, L. Chen, Y. Y. Tang, and Y. Zhou, "Weighted couple sparse representation with classified regularization for impulse noise removal," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4014–4026, Nov. 2015.
[65] L. Liu, L. Chen, C. L. P. Chen, Y. Y. Tang, and C. M. Pun, "Weighted joint sparse representation for removing mixed noise in image," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2016.2521428
[66] X. Xu, W. Liu, and L. Li, "Face hallucination: How much it can improve face recognition," in *Proc. Australian Control Conf.*, 2013, pp. 93–98.
[67] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[68] C.-P. Wei and Y.-C. F. Wang, "Undersampled face recognition via robust auxiliary dictionary learning," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1722–1734, Jun. 2015.
[69] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
[70] D. Chen, Y. Hu, C. Cai, K. Zeng, and X. Li, "Brain big data processing with massively parallel computing technology: Challenges and opportunities," *Softw.: Practice Experience*, 2016. doi: 10.1002/spe.2418
[71] D. Chen *et al.*, "Fast and scalable multi-way analysis of massive neural data," *IEEE Trans. Comput.*, vol. 64, no. 3, pp. 707–719, Mar. 2015.
[72] D. Chen *et al.*, "Parallel simulation of complex evacuation scenarios with adaptive agent models," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 3, pp. 847–857, Mar. 2015.
[73] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recog.*, vol. 46, no. 12, pp. 3519–3532, Dec. 2013.
[74] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2016.2594184

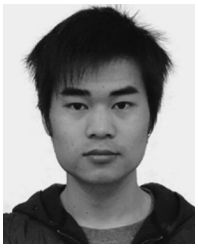


Junjun Jiang (M'15) received the B.S. degree from the School of Mathematical Sciences, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree from the School of Computer, Wuhan University, Wuhan, China, in 2014.

He is currently an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. He has been a Project Researcher with the National Institute of Informatics, Tokyo, Japan, since 2016. He has authored or coauthored more than 60 scientific articles, and has obtained eight Chinese patents. His research interests include image processing and computer vision.

Chen Chen received the B.E. degree in automation from Beijing Forestry University, Beijing, China, in 2009, the M.S. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 2012 and the Ph.D. degree from the University of Texas at Dallas, Richardson, TX, USA, in 2016.

He is currently a Post-Doc with the Center for Research in Computer Vision, University of Central Florida, Orlando, FL, USA. His research interests include compressed sensing, signal and image processing, pattern recognition, computer vision, and hyperspectral image analysis.



Jiayi Ma (M'16) received the B.S. degree in mathematics and the Ph.D. degree in automation from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. He is currently an Associate Professor with the Electronic Information School, Wuhan University, Wuhan, China, where he was a Postdoctoral Researcher from 2014 to 2015. His current research interests include the areas of computer vision, machine learning, and pattern recognition.

Zheng Wang received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 2006 and 2008, respectively, and is currently working toward the Ph.D. degree at the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University.

His research interests include multimedia content analysis and retrieval, computer vision, and pattern recognition.

Zhongyuan Wang (M'13) received the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2008.

He is currently a Professor with the 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.

Ruimin Hu (M'09–SM'09) received the B.S. and M.S. degrees from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1984 and 1990, respectively, and Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1994.

He is the Dean of the School of Computer, Wuhan University, Wuhan, China. He has authored or coauthored two books and more than 100 scientific papers. His research interests include audio/video coding and decoding, video surveillance, and multimedia data processing.