Noise Robust Face Image Super-Resolution Through Smooth Sparse Representation

Junjun Jiang, *Member, IEEE*, Jiayi Ma, *Member, IEEE*, Chen Chen, Xinwei Jiang, and Zheng Wang

***Abstract*—Face image super-resolution has attracted much attention in recent years. Many algorithms have been proposed. Among them, sparse representation (SR)-based face image super-resolution approaches are able to achieve competitive performance. However, these SR-based approaches only perform well under the condition that the input is noiseless or has small noise. When the input is corrupted by large noise, the reconstruction weights (or coefficients) of the input low-resolution (LR) patches using SR-based approaches will be seriously unstable, thus leading to poor reconstruction results. To this end, in this paper, we propose a novel SR-based face image super-resolution approach that incorporates smooth priors to enforce similar training patches having similar sparse coding coefficients. Specifically, we introduce the fused least absolute shrinkage and selection operator-based smooth constraint and locality-based smooth constraint to the least squares representation-based patch representation in order to obtain stable reconstruction weights, especially when the noise level of the input LR image is high. Experiments are carried out on the benchmark FEI face database and CMU+MIT face database. Visual and quantitative comparisons show that the proposed face image super-resolution method yields superior reconstruction results when the input LR face image is contaminated by strong noise.**

***Index Terms*—Face image, fused least absolute shrinkage and selection operator (LASSO), smoothness constraint, sparse representation (SR), super-resolution.**

I. Introduction

WITH the deepening of urbanization process, urban security issues become more and more serious in major cities of many countries around the world. In order to maintain social stability, city surveillance camera system (e.g., closed circuit television system) has been established in most cities. Although the city surveillance camera network is expanding gradually, the role it can play is often limited. Actually, the surveillance cameras are normally installed to cover a large field of view and the captured images are of low-resolution (LR) and low-quality. The interested persons of perpetrators and potential eyewitnesses are often in the form of a few pixels and lack of detailed features, which may not be helpful for image analysis and recognition [1], [2]. Therefore, how to transcend the limitations of surveillance camera system and reconstruct a high-resolution (HR) and high-quality face from an LR surveillance image is a problem that is exigent to be solved.

Super-resolution reconstruction is of great importance for vision applications, and numerous algorithms have been proposed in recent years. Generally speaking, they can be categorized based on their tasks, i.e., generic image super-resolution with self-similarity prior [3]–[7], sparsity prior [8]–[10], optical flow prior [11] or gradient prior [12], and domain-specific image super-resolution focus on specific classes of images such as faces [13]–[16] and text images [2], [17], [18]. In this paper, we mainly focus on face image super-resolution (also called face hallucination), which refers to inducing an HR face image from an LR face image by learning the relationship between the HR and LR training pairs, and thus providing more facial details for the following face synthesis and analysis [19], [20]. It is a hot research topic in image processing and computer vision [13]. Baker and Kanade [21] proposed "face hallucination" to infer the HR face image from an input LR one based on a parent structure with the assistance of LR and HR training samples. Rather than using the whole or parts of a face, the super-resolution is established based on training images (pixel by pixel) using Gaussian, Laplacian, and feature pyramids. Since the introduction of this paper, a number of different methods and models have been introduced for estimating the image information lost in the image degradation process [22], [23]. Liu *et al.* [24] described a two-step approach integrating a global parametric Gaussian model and a local nonparametric Markov random field (MRF) model. The first step is to learn a global linear model to construct the relationship between HR face images and the corresponding smoothed and down-sampled LR ones. The second step is to model the residue between an original HR image and the reconstructed HR image by a nonparametric MRF model.

Following [21] and [24], progress has been made in estimating an HR face image from a single LR face image with a

Manuscript received January 17, 2016; revised May 5, 2016 and July 1, 2016; accepted July 10, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61501413, Grant 61503288, and Grant 61402424, in part by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant CUGL160412, and in part by the China Post-Doctoral Science Foundation under Grant 2015M570665 and Grant 2016T90725. This paper was recommended by Associate Editor S. Zafeiriou. *(Corresponding author: Jiayi Ma.)*

J. Jiang and X. Jiang are with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: junjun0595@163.com; ysjxw@hotmail.com).

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

C. Chen is with the Center for Research in Computer Vision, University of Central Florida, Orlando, FL 32816-2365 USA (e-mail: chenchen870713@gmail.com).

Z. Wang is with the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan 430072, China (e-mail: wangzwhu@whu.edu.cn).

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

Digital Object Identifier 10.1109/TCYB.2016.2594184

training set of HR and LR image pairs. They usually construct the target HR face either globally using holistic face image or locally using patches.

Global face-based super-resolution methods utilize different face representation models such as principal component analysis (PCA) [25], kernel PCA [26], canonical correlation analysis (CCA) [27], locality preserving projections [28], and non-negative matrix factorization [8], to model an input LR face image using a linear combination of LR prototypes in the training set. Then, the target HR face image is reconstructed by replacing the LR training images with the corresponding HR ones, while using the same coefficients. However, global face-based super-resolution methods cannot well recover the fine individual details of an input face which are essential for the following face recognition task.

By decomposing a holistic face image into small patches, local patch-based face image super-resolution methods have strong synthesis ability and can capture more facial details. Therefore, in this paper, we mainly focus on the local patch-based methods. For a comprehensive overview of face image super-resolution techniques, readers are referred to [13].

The basic assumption of local patch-based face image super-resolution methods is that if two LR face patches are similar, then their corresponding HR ones are also similar, i.e., the high-frequency details lost in an LR image can be learned from a training set of LR and HR image pairs. Therefore, once obtaining the relationship between the LR and HR patches, we can use the learned relationship (explicit regressions or implicit representations) to predict the target HR patch of the observation LR patch. As for the explicit regression-based methods, they aim at learning the direct mapping function from the LR space to the corresponding HR space [29]–[31]. For the implicit representation-based methods, Freeman *et al.* [32] presented an example-based super-resolution approach to select the nearest patch by modeling the relationship between local regions of images and scenes using the patch-wise Markov network learned from the training set. This approach is computationally intensive and sensitive to training set. To increase the flexibility of the nearest patch selection, Chang *et al.* [33] introduced the neighbor embedding (NE) algorithm which is inspired by the idea of locally linear embedding [34]. It predicts the HR patch by combining the HR training patches in a linear manner. Recently, the coupled spaces learning techniques have been introduced to model the relationship between the LR and HR images in a common coherent subspace from the perspective of manifold alignment. For example, Li *et al.* [35] performed face image super-resolution on a synthesized common manifold by two explicit mappings. Huang *et al.* [27] used the CCA to project both LR and HR training patches onto a common coherent subspace. An and Bhanu [36] further extended CCA to 2-D CCA. Jiang *et al.* [37] proposed a coupled-layer NE for face super-resolution.

To incorporate more face structure priors, Ma *et al.* [15] proposed a position-patch-based framework. It constructs image patches using all the training patches based on least squares representation (LSR). To overcome the unstable solution of LSR, sparsity constraint is imposed on the patch representation, leading to the sparse representation (SR)-based face image super-resolution methods [8], [38], [39]. Wang *et al.* [40], [41] introduced a weighted SR for face image super-resolution. In [42], a novel SR method was developed by exploiting the support information on the representation coefficients. This method is further extended to the coupled space while preserving the local manifold structure of the HR data space [30]. Gao and Yang [43] proposed to learn the relationship between LR and HR training samples in the SR space rather than in the original data space. In [44], a Cauchy regularized SR method was proposed to represent the input LR patch. Though SR has achieved great success in face image super-resolution problem, one challenge for SR-based methods is their sensitivity to noise. To further explore the relationship between the local pixels, Shi *et al.* [45] proposed to combine the global reconstruction model, the local sparsity model and the pixel correlation model into a unified regularization framework and presented a novel two-phase framework for the face image super-resolution problem. Noise in the LR observation image has a great impact on the reconstruction weights of SR-based methods. To address this problem, Jiang *et al.* [46], [47] introduced a locality constraint to the patch representation model and developed an efficient face image super-resolution algorithm using locality-constrained representation (LcR). By introducing the locality constraint, the impact of noise to face image super-resolution performance can be reduced that has also been certified in [48].

The fused least absolute shrinkage and selection operator (LASSO) penalty introduced in [49] can enforce a sparse solution in both the coefficients and the differences between neighboring coefficients through an L_1 norm regularizer, which is desirable for applications such as prostate cancer analysis [49], image denoising [50], [51], and time-varying networks [52]. It encourages the sparsity of the regression coefficients and shrinks the differences between neighboring coefficients toward zero simultaneously. As such, the method achieves both sparseness and smoothness in the regression coefficients.

A. Motivation and Contributions

The aforementioned local patch-based approaches only consider certain aspects of the patch representation, thus the representation is not optimal. For example, SR-based methods focus on the sparsity but neglect the locality prior, while LcR considers the locality but takes no consideration of the smoothness of the reconstruction weights. In addition, for LcR [46], [47], it may assign very different reconstruction weights to two similar patches, which is unreasonable in practice. To obtain smooth reconstruction weights, in this paper we incorporate the fused LASSO term and locality regularization into the patch representation objective function, and propose a novel face image super-resolution method based on smooth SR (SSR). The combination of fused LASSO term and locality regularization is able to enforce similar patches having similar reconstruction weights, while the sparsity constraint makes the reconstruction weights sparse. As a result,

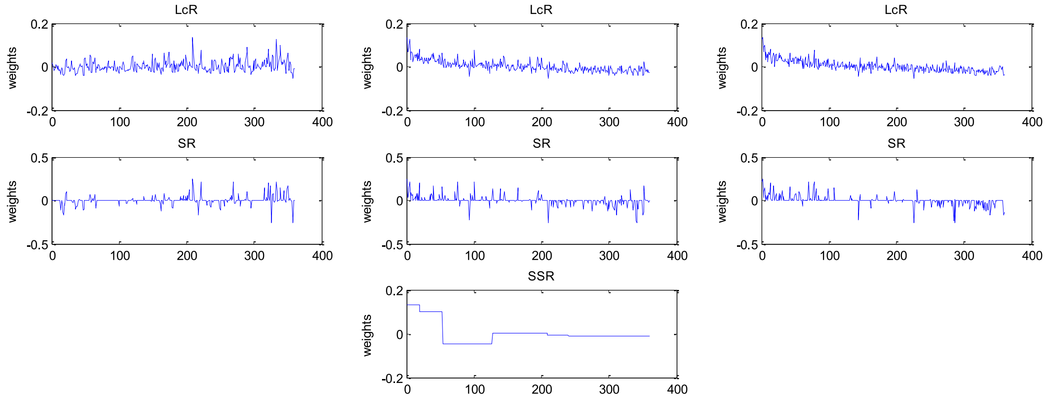


Fig. 1. Plots of reconstruction weights of one randomly selected patch associated with different patch representation methods. The first and the second rows are the plots of LcR [46] and SR [8] methods: the horizontal axis of the first column is the index of the original training images without sorting, while these of the second and the third are the indexes of the original training images with sorting by Euclidean distance and cosine distance, respectively. The third row is the plot of reconstruction weights according to the sorted training patch index.

the reconstruction weights of SSR achieve smoothness and sparsity simultaneously.

Fig. 1 shows the reconstruction weights of an image patch according to the sorted training patch index by the distance (we use Euclidean distance or cosine distance as a similarity measure) to the input patch (for more details about the experimental configuration, please refer to the experiment section). It is worth noting that smaller index means the training patch is closer (or more similar) to the input patch. From the plots in this figure, we can see that SR-based method has no regularity. The reconstruction weights are randomly distributed among the LR training patch samples. Although the general trend of the reconstruction weights of LcR is decreasing with the decreasing similarity between the input patch and the training patch, the reconstruction weights are not smooth [e.g., some training patches that are far away from (or dissimilar to) the input patch still have large reconstruction weights]. Our proposed SSR method can obtain smooth reconstruction weights. Moreover, the training patches that are similar to the input patch are given large reconstruction weights while the training patches that are far away from the input one are assigned with small reconstruction weights. This can be attributed to the incorporating of fused LASSO prior and locality prior simultaneously, which makes similar patches in the training set have similar weights and similar patches to the input one have large weights, thus leading to a smooth representation.

The contributions of this paper are threefold.

1) Based on the assumption that similar training patches have similar coding coefficients, we propose an SSR model which considers the smoothness and sparsity of the reconstruction weights simultaneously.
2) To the best of our knowledge, it is the first time that fused LASSO is introduced to face image super-resolution problem.
3) Extensive experimental results verified the superior performance of our proposed method compared to the state-of-the-art face super-resolution algorithms, especially under the noise condition.

B. *Organization of This Paper*

The remainder of this paper is organized as follows. Section II introduces notations used in this paper and gives the formulations of patch-based face image super-resolution method. Section III first reviews the SR method, and then describes the proposed patch SSR model and shows details of face image super-resolution via SSR. Section IV presents the experimental results and analysis. Finally, Section V concludes this paper.

II. Notations and Formulations

Given an LR observation image, the goal of face image super-resolution is to construct its HR version by learning the relationship between the LR and HR training sets. As for a local patch-based method, we divide the LR observation image I_t^L into M patches, $\{x_t(p,q)|1 \le p \le U, 1 \le q \le V\}$, according to the predefined patch_size and overlap pixels. $x_t(p,q)$ denotes a small patch at the position (p,q) of the LR observation image x_t, U represents the patch number in each column, and V represents the patch number in each row. Therefore, we have $M = UV$. In the same way we divide all the N LR and HR training face image pairs, $I^L = \{I_1^L, I_2^L, \ldots, I_N^L\}$ and $I^H = \{I_1^H, I_2^H, \ldots, I_N^H\}$, into M patches, respectively, $\{x_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$ and $\{y_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$. $x_i(p,q)$ denotes a small patch at the position (p,q) of the ith training sample in the LR training set, while $y_i(p,q)$ denotes a small patch at the position (p,q) of the ith training sample in the HR training set. For more details about the dividing strategy, please refer to [46]. For the LR observation patch $x_t(p,q)$ located at the position (p,q), e.g., the nose patch as shown in Fig. 2, its HR patch $y_t(p,q)$ is estimated using the LR and HR training image patch pairs at the same position [when it does not lead to a misunderstanding, we drop the term (p,q) for convenient from now on]. In particular, they first represent the given LR patch x_t with the LR training image patch set $\{x_i\}_{i=1}^N$, and then transform the representation coefficients w (i.e., the

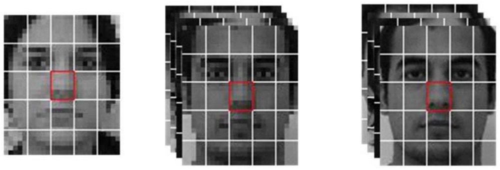


Fig. 2. Face image dividing according to positions. The left column is the input LR face, the middle column is the LR training face image set and the right column is the HR training face image set.

outcome of different representation methods) to faithfully represent each corresponding (unknown) HR patch y_t by replacing the LR training image patch set $\{x_i\}_{i=1}^N$ with its HR counterpart $\{y_i\}_{i=1}^N$. To handle the compatibility problem between adjacent patches, simple averaging in the overlapping regions is performed. From the process presented above, we learn that the key issue of these local patch-based methods is to obtain the optimal representation coefficients w, and this section reviews some existing representation schemes.

Given an LR observation patch x_t on the input LR face image, LSR uses patches from all training samples at the same position to represent it [15]

$$x_t = \sum_{i=1}^{N} w_i x_i + \epsilon \tag{1}$$

where ϵ is the reconstruction error. The optimal weight can be solved by the following constrained least square fitting problem:

$$\min \left\| x_t - \sum_{i=1}^{N} w_i x_i \right\|_2^2, \quad \text{s.t. } \sum\nolimits_{i=1}^{N} w_i = 1 \tag{2}$$

where $w = [w_1, w_2, \ldots, w_N]^T$ is the optimal N-dimensional weight vector for the LR observation patch x_t. The least square estimation can produce biased solutions when the dimension of the patch is smaller than the size of the training image set [38].

III. Proposed Smooth Sparse Representation-Based Face Image Super-Resolution

A. Sparse Representation

To solve the biased patch representation problem, a possible solution is to impose some regularization terms onto it. Inspired by the compressed sensing theory [53]–[55], which has been proved to be effective in many applications such as feature extraction [56], [57], and classification [58]–[60]. Yang *et al.* [8] introduced the sparsity constraint and used a small subset of patches to represent LR observation patch x_t instead of performing collaboratively over the whole training samples set

$$J_{\text{SR}}(w) = \left\| x_t - \sum_{i=1}^{N} w_i x_i \right\|_2^2, \quad \text{s.t. } \|w\|_1 < \epsilon. \tag{3}$$

Lagrange multipliers offer an equivalent formulation of the above equation

$$J_{\text{SR}}(w) = \left\| x_t - \sum_{i=1}^{N} w_i x_i \right\|_2^2 + \lambda \|w\|_1 \tag{4}$$

where $\| \bullet \|_1$ denotes the ℓ_1-norm, and λ is the regularization parameter that balances the contribution of the reconstruction error and the sparsity of the reconstruction weights. This sparsity constraint not only ensures that the under-determined equation has an exact solution but also allows the learned representation for each patch to capture the salient properties.

B. Smooth Sparse Representation

SR-based methods emphasize that strong sparsity of the reconstruction weights is important in representing the input patch. However, there are two drawbacks of SR in the present context.

1) One is the fact that it ignores smoothness (or flatness) of the sparse reconstruction weights, which tends to set neighbor penalties exactly equal to each other.
2) The other is that it neglects a locality constraint, which states that the training patches that are most similar to the input patch should be given larger reconstruction weights than those that are most dissimilar.

Our proposed SSR method aims to overcome the two limitations of SR-based methods. In addition, by incorporating the locality constraint, it makes our proposed patch representation method smoother (see Section III-E for the smoothness measurement of our proposed method).

1) Fused LASSO-Based Smooth Constraint: For the first problem, we introduce the fused LASSO penalty $\sum_{i=2}^{N} |w_i - w_{i-1}|_1$, which has found its applications in prostate cancer analysis [49], image denoising [61], and time-varying networks [52], to the patch representation objective function

$$J_{\text{SSR}}(w) = \left\| x_t - \sum_{i=1}^{N} w_i x_i \right\|_2^2$$
$$\text{s.t. } \|w\|_1 < \epsilon_1 \text{ and } \sum_{i=2}^{N} \|w_i - w_{i-1}\|_1 < \epsilon_2. \tag{5}$$

Similar to (4), an equivalent formulation of (5) can be given by using the Lagrange multipliers, that is

$$J_{\text{SSR}}(w) = \left\| x_t - \sum_{i=1}^{N} w_i x_i \right\|_2^2 + \lambda_1 \|w\|_1 + \lambda_2 \sum_{i=2}^{N} \|w_i - w_{i-1}\|_1 \tag{6}$$

where the non-negative parameters λ_1 and λ_2 are used to control the contributions of the sparsity constraint and the sparsity difference constraint, respectively. The first constraint encourages sparsity in the reconstruction weights and the second constraint encourages sparsity in their differences, i.e., flatness of the reconstruction weight profiles w_i as a function of i.

2) Locality-Based Smooth Constraint: For the second problem, we sort the LR training patch samples (each column of $X = [x_1, x_2, \ldots, x_N]$) according to the similarity between the input LR patch and each LR training patch sample, which is simply determined by the squared Euclidean distance

$$\text{dist} = \left\{ \|x_t - x_i\|_2^2 \middle| 1 \leq i \leq N \right\}. \tag{7}$$

We then sort the distance dist in ascending order and obtain the index vector, $idx = [[1], [2], \ldots, [N]]$. Then, the sorted LR training patch samples can be defined by $X^{idx} = [x_{[1]}, x_{[2]}, \ldots, x_{[N]}]$. Thus, $x_{[i]}$ is the ith element of the reordered LR training patch samples. Finally, the objective function of our proposed SSR can be written as

$$\arg\min_{w} \left\{ \left\| x_t - \sum_{i=1}^{N} w_i x_{[i]} \right\|_2^2 + \lambda_1 \|w\|_1 + \lambda_2 \sum_{i=2}^{N} \|w_i - w_{i-1}\|_1 \right\}. \tag{8}$$

By sorting the LR training patch samples, we can expect to obtain large reconstruction weights for the training samples that are close to the input LR patch.

C. Optimization of SSR

To get the optimal weights w^* from SSR (8), we employ the fast iterative shrinkage thresholding algorithm [62]. It is known that a gradient method can be used to optimize a smooth objective function[1] $l(w)$ (e.g., in the SSR objective function, $l(w) = \|x_t - \sum_{i=1}^{N} w_i x_{[i]}\|_2^2$).

We denote the regularization terms in (8) as

$$\Omega(w) = \lambda_1 \|w\|_1 + \lambda_2 \sum_{i=2}^{N} \|w_i - w_{i-1}\|_1. \tag{9}$$

Here, $\Omega(w)$ is nonsmooth. Following fast iterative shrinkage and thresholding algorithm, the updating rule of w in each iteration is:

$$w_{(t+1)} = \arg\min_{w} \left\{ l(w_{(t)}) + \langle w - w_{(t)}, \nabla l(w_{(t)}) \rangle + \frac{L}{2} \left\| w - w_{(t)} \right\|_2^2 + \Omega(w) \right\} \tag{10}$$

where $L > 0$ is the Lipschitz constant of the gradient $\nabla l(w)$, and the minimization admits a unique solution. By ignoring some constant terms of $w_{(t)}$, we have

$$w_{(t+1)} = \arg\min_{w} \left\{ \Omega(w) + \frac{L}{2} \left\| w - \left(w_{(t)} - \frac{1}{L} \nabla l(w_{(t)}) \right) \right\|_2^2 \right\}. \tag{11}$$

Thus, the key to solve (8) is how efficiently we can solve (11). Minimization of the approximate objective function (11) can be rewritten as

$$\arg\min_{w} \frac{L}{2} \left\| z_{(t)} - w \right\|_2^2 + \lambda_1 \|w\|_1 + \lambda_2 \sum_{i=2}^{N} \|w_i - w_{i-1}\|_1 \tag{12}$$

where $z = w_{(t)} - (1/L)\nabla l(w_{(t)})$. Equation (12) is the standard fused LASSO signal approximator (FLSA) introduced by Friedman *et al.* [50].

It can be shown that $w_i(\lambda_1, \lambda_2)$, the optimal solution for the standard FLSA for regularization parameters (λ_1, λ_2), can be obtained from $w_i(0, \lambda_2)$ (set $\lambda_1 = 0$) by soft-thresholding

$$w_i(\lambda_1, \lambda_2) = \text{sign}(w_i(0, \lambda_2)) \cdot \max(|w_i(0, \lambda_2)| - \lambda_1, 0) \quad \text{for } i = 1, 2, \ldots, N. \tag{13}$$

Hence, we can set $\lambda_1 = 0$ without loss of generality and focus on the regularization path by varying only λ_2

$$\arg\min_{w} \frac{1}{2} \left\| z_{(t)} - w \right\|_2^2 + \lambda_2 \sum_{i=2}^{N} \|w_i - w_{i-1}\|_1. \tag{14}$$

Let D be a finite different matrix with dimension $(N-1) \times N$ and its entries are zeros everywhere except -1 in the diagonal and 1 in the superdiagonal

$$D = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix}.$$

By introducing a dual variable u, we can reformulate (14) as the following equivalent min–max problem:

$$\min_{w} \max_{\|u\|_\infty \leq \lambda_2} \frac{1}{2} \left\| z_{(t)} - w \right\|_2^2 + \langle Dw, u \rangle. \tag{15}$$

Exchanging min and max and setting the derivative of (15) with respect to w to zero, solution to (15) is obtained from the primal-dual relationship

$$w = z - D^T u. \tag{16}$$

Plugging (16) into (15), we get the dual problem

$$\min_{\|u\|_\infty \leq \lambda_2} \frac{1}{2} u^T D D^T u - u^T D z. \tag{17}$$

Since (17) is a box constrained optimization problem, it can be efficiently solved by the following projected gradient method [63], [64]:

$$u_{(t+1)} = P_{\lambda_2}\left(u_{(t)} - \alpha \left(D D^T u_{(t)} - D z_{(t)} \right) \right) \tag{18}$$

where P_{λ_2} denotes the projection operator onto an l_∞ ball $\{u : \|u\|_\infty \leq \lambda_2\}$, and α is the reciprocal of the largest eigenvalue of the matrix DD^T.

We solve the FLSA problem (14) through an iterative algorithm by alternating between the primal and the dual optimization as follows:

$$\begin{cases} \text{Primal: } w_{(t+1)} = z_{(t)} - D^T u_{(t)} \\ \text{Dual: } u_{(t+1)} = P_{\lambda_2}\left(u_{(t)} - \left(D D^T u_{(t)} - D z_{(t)} \right) \right) \end{cases} \tag{19}$$

where the first step updates the primal variable based on the current estimate of $u_{(t)}$, and the second step updates the dual variable based on the current estimate of the primal variable $w_{(t)}$. Note that $z_{(t)}$ can be obtained by $z_{(t)} = w_{(t)} - (1/L)\nabla l(w_{(t)})$.

[1] Smooth means that l has a Lipschitz continuous gradient.

Algorithm 1 Face Image Super-Resolution via SSR

```
Input: LR and HR training set $I^L = \{I_1^L, I_2^L, \cdots, I_N^L\}$ and $I^H = \{I_1^H, I_2^H, \cdots, I_N^H\}$, and a LR observation face image $I_t^L$. The parameters: $patch\_size$, $overlap$, $\lambda_1$, and $\lambda_2$.
Compute $U$ and $V$:
    $U = ceil((imrow - overlap)/(patch\_size - overlap))$
    $V = ceil((imcol - overlap)/(patch\_size - overlap))$
Divide each of the LR and HR training images and the input LR image into $M$ small patches according to the same location of face, $\{x_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$, $\{y_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$ and $\{x_t(p,q)|1 \le p \le U, 1 \le q \le V\}$, respectively.
for $p = 1 : U$ do
    for $q = 1 : V$ do
        $dist(p,q) = \left\{||x_t(p,q) - x_i(p,q)||_2^2\right\}_{i=1}^N$.
        Sort $dist(p,q)$ in ascending order to get $idx$.
        Obtain the sorted LR and HR training patch samples, $X^{idx} = [x_{[1]}, x_{[2]}, \ldots, x_{[N]}]$ and $Y^{idx} = [y_{[1]}, y_{[2]}, \ldots, y_{[N]}]$.
        Calculate $w^*(p,q)$ according to (8).
        $y_t(p,q) = \sum_i w_i^*(p,q) y_i(p,q)$.
    end for
end for
Integrating all the obtained HR patches $\{y_t(p,q)|1 \le p \le U, 1 \le q \le V\}$ above according to the position $(p,q)$. The final HR image can be generated by averaging pixel values in the overlapping regions.
Output: HR super-resolved face image $I_t^H$.
```

1) Convergence Analysis: To solve the objective function (15), i.e., the reconstruction weights w and the dual variable u, we adopt an iterative strategy and solve one variable by fixing the other. We decompose our objective function (15) into two subproblems: 1) the primal problem and 2) the dual problem. By setting the derivative of the primal problem with respect to w to zero (16), it can enable us to find the global optimum by the primal iteration. To optimize (18), the dual iteration has been proved to weakly converge to a fixed point [63], [64]. That is to say, it can theoretically ensure that a near optimal solution could be achieved in each step during the iteration. Therefore, its value will not be increased in the iteration process, which guarantees the convergence of the iterative algorithm. Actually, the procedure in (19) can be seen as the alternating direction method of multipliers algorithm which has a fast convergence speed [65].

D. Face Image Super-Resolution Through SSR

As a position patch-based face image super-resolution method, SSR consists of the following steps. First, we divide all the LR and HR training face images into patches according to the positions. Then, for each LR test image patch x_t, we can calculate its optimal patch representation weight vector w^*, and transform it to the HR training patch space by

$$y_t = \sum_i w_i^* y_{[i]}. \tag{20}$$

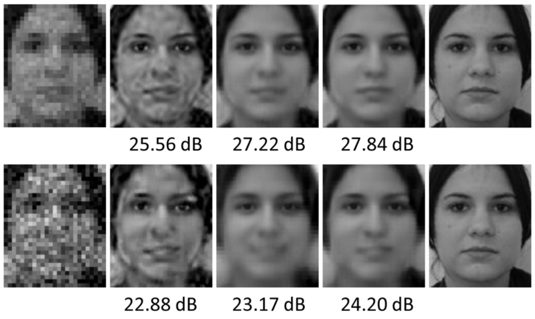


Fig. 3. Super-resolved results and peak signal-to-noise ratio (PSNR) values by SR (second column), fused LASSO (third column) and our SSR (fourth column). The first and last columns are the input LR face and ground truth HR face, respectively. The values under the second to the fifth columns are the PSNR values.

TABLE I
SI Indexes of One Subject (Corresponding to Fig. 3) of Different Methods Under Different Noise Levels

Noise levels	SR	Fused Lasso	SSR
σ=10	0.0062	0.8689	0.9206
σ=30	0.1184	0.9572	0.9698

Similarly, $[y_{[1]}, y_{[2]}, \ldots, y_{[N]}]$ are the sorted HR training patch samples. All the LR patches in the input LR face image are processed in raster-scan order, from left to right and top to bottom. Lastly, we enforce compatibility between adjacent patches (the values in the overlapped regions are simply averaged) following [8]. The entire process of the proposed SSR-based face image super-resolution method is summarized in Algorithm 1. ceil(x) is the function that rounds the elements of x to the nearest integers toward infinity.

E. Smoothness Measurement

In recent years, many image super-resolution methods have been proposed that use penalties on the regression (coding) coefficients in order to achieve sparseness or shrink them toward zero. They may overemphasize on sparsity to reconstruct the input image patch. As a result, very distinct patches may be chosen. In addition, the SR solution is usually not stable, especially when the noise is strong, due to the limitation of sparse recovery. Based on the assumption that similar training patches have similar coding coefficients, it is possible that there is some natural ordering of the coefficients. In other words, neighboring coefficients should not change fast but should change smoothly. By incorporating the smoothness constraint, the most similar training patches will be chosen in the sparse coding process, thus the input noise will be averaged. In addition, it can also lead to the exact solution of the under-determined least squares problem.

In order to measure the smoothness of different patch representation approaches, e.g., SR, fused LASSO, and the proposed SSR, we define an evaluation metric of smoothness for the patch representation called smooth index (SI). Let $w = [w_1, w_2, \ldots, w_N]$ denote the representation of one patch, then SI is defined as

$$\mathrm{SI} = e^{-\sum_{i=2}^N |w_i - w_{i-1}|}. \tag{21}$$

Fig. 4. Some training faces in the FEI face database.

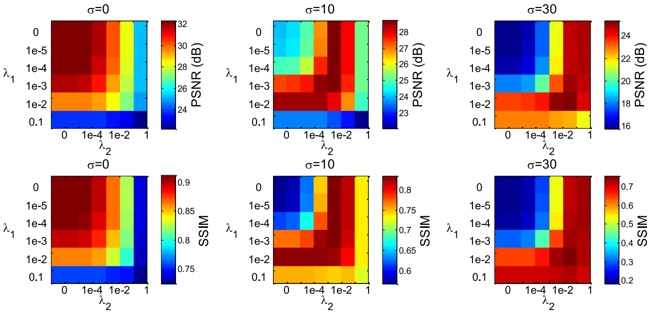


Fig. 5. PSNR (dB) and structural similarity index measure (SSIM) performance of our proposed SSR method under different levels of noise using different values of λ_1 and λ_2.

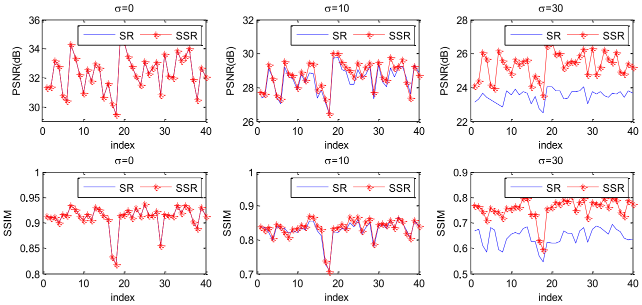


Fig. 6. PSNR (dB) and SSIM comparisons of all the 40 test images reconstructed by SR and our proposed SSR under different noise levels.

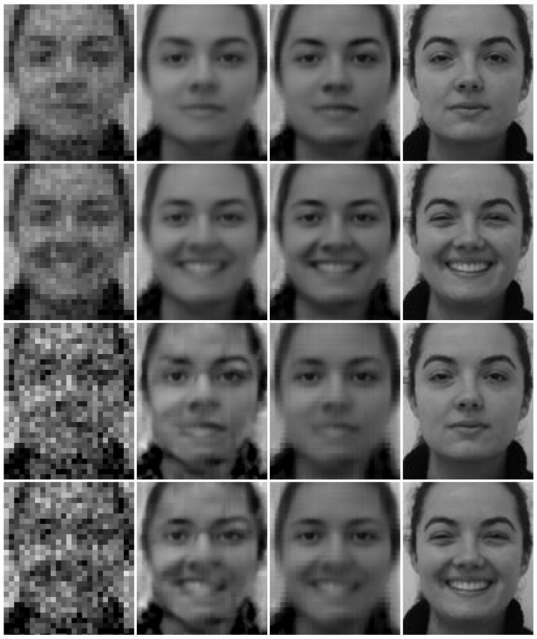


Fig. 7. Visual reconstruction results of one subject with different noise levels [from top to bottom (two rows as a group), the noise levels are $\sigma = 0, 10, 30$, respectively] using SR method (the second column) and the proposed SSR method (the third column). Note that the first column are the input LR faces, while the last column are the ground truth HR faces.

TABLE II
AVERAGE PSNR (dB) AND SSIM COMPARISONS OF DIFFERENT METHODS UNDER DIFFERENT NOISE LEVELS. RED INDICATES THE BEST AND BLUE INDICATES THE SECOND BEST PERFORMANCE

Methods	$\sigma = 10$		$\sigma = 30$		Implus noise	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	25.41	0.6856	19.44	0.3269	23.65	0.6948
Wang [25]	26.68	0.7495	24.60	0.7252	25.75	0.7319
NE [33]	28.13	0.8178	24.57	0.7180	26.78	0.7995
LSR [15]	23.60	0.5206	15.26	0.1557	20.85	0.5816
SR [8]	28.44	0.8273	23.51	0.6446	26.76	0.8045
WSR [40]	28.15	0.8067	24.15	0.6857	26.61	0.7896
LcR [46]	28.18	0.8323	25.06	0.7458	27.02	0.7976
SSR	28.69	0.8334	25.69	0.7590	27.20	0.8148
Gains	0.25	0.0061	0.61	0.0132	0.18	0.0103

From the definition of SI, the smoothness of a representation approach is measured by the inverse of the accumulated absolute difference between two adjacent coefficients. The smoother of *w*, the larger of SI. The maximum value of SI is 1. Fig. 3 presents the reconstruction HR faces and their PSNRs (dB) values by SR, fused LASSO and our proposed SSR. Table I reports the average[2] SI indexes of one subject in Fig. 3 of different methods under different noise levels. Note that we have tuned the parameters λ in SR as well as λ_1 and λ_2 in fused LASSO and SSR to achieve the best performances. From Fig. 3, we can learn that our proposed SSR is better than SR and fused LASSO in terms of visual quality and quantitative measure. In addition, the representation of fused LASSO and SSR are smoother than that of the SR method from the results in Table I. This implies that fused LASSO-based smooth constraint is very important for the patch representation. Moreover, with the locality-based smooth constraint, the representation can be smoother.

IV. Experimental Results

In this section, we report the results of extensive experiments performed to evaluate the effectiveness of the proposed SSR approach for face image super-resolution. The experiments are conducted on two public face databases namely, the FEI face database [66] and the CMU+MIT face database [67]. All the face images are aligned by some recently proposed automatic alignment methods [68] and robust feature matching technology [69].

A. Database

The FEI face dataset [66] contains 200 persons, 100 men, and 100 women. Each person has two images (one with normal

[2] SI is a measurement of the representation of one image patch, to measure SI index of the entire face image, we calculate the average SI of all the patches in a face image. Here, the SI is calculated with the optimal reconstruction weights, which is obtained when the method achieves the best performance in terms of PSNR.

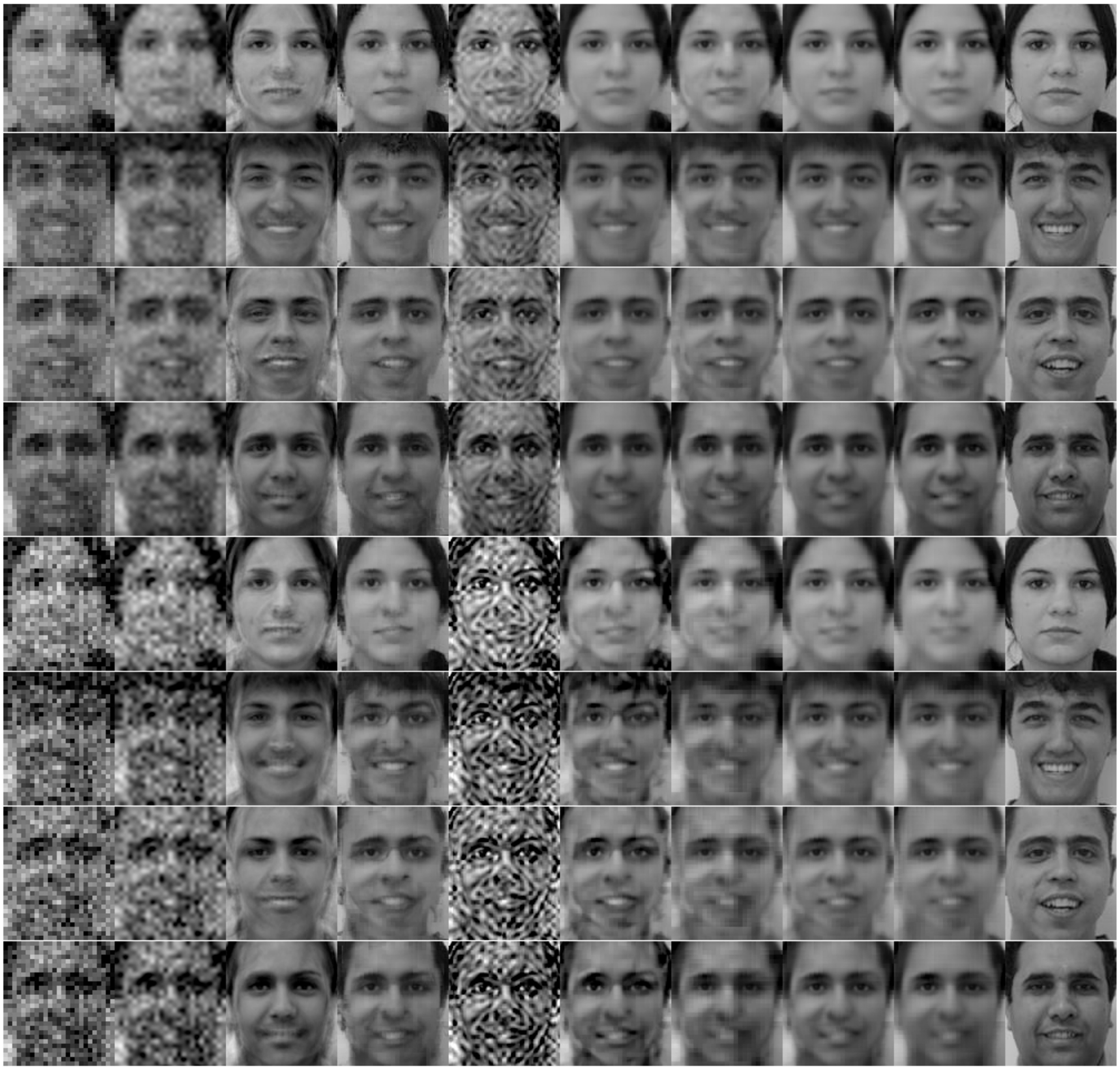


Fig. 8. Visual reconstruction results of four subjects with different noise levels [from top to bottom (four rows as a group), the noise levels are $\sigma = 10, 30$, respectively] using different methods (from left to right, they are the input LR image, results of bicubic interpolation, Wang and Tang's method [25], NE [33], LSR [15], SR [8], WSR [40], LcR [46], and our proposed SSR method, and the ground truth HR face images).

expression and the other with smile expression). Some samples from this dataset are illustrated in Fig. 4. A face alignment process has been done [66] per image enforcing the central points of the eyes at fixed locations. We further crop and scale the images into 120×100 normalized face images manually. We randomly choose 180 persons (360 face images) for training and leave the rest 20 persons (360 face images) for testing. In the experiment, the HR images are first smoothed and down-sampled to LR 30×25 images (by a factor of four), and then the additive white Gaussian noise (AWGN) with different noise levels (denoted by σ, e.g., $\sigma = 0, 10, 30$) is added. The LR patch size is 4×4 while the HR counterpart is 16×16. The overlap between neighbor patches is 3 pixels for LR patches and 12 pixels for HR patches.

B. Parameter Settings

In our proposed method, the parameters λ_1 and λ_1 are set experimentally. As shown in Fig. 5, we report the PSNR and SSIM[3] [70] performance under different noise levels according to various values of λ_1 and λ_2. Experimentally, we set $\lambda_1 = 1e - 4$ and $\lambda_2 = 0$ ($\sigma = 0$), $\lambda_1 = 1e - 4$ and $\lambda_2 = 1e - 2$ ($\sigma = 10$), and $\lambda_1 = 1e - 2$ and $\lambda_2 = 0.1$ ($\sigma = 30$) to achieve the best performance. From the objective function of SSR (6), we learn that λ_1 and λ_2 are used to control the contributions of the sparsity constraint and the sparsity difference constraint, respectively. When $\sigma = 0$, λ_2 is set to 0 to obtain the best performance, which indicates that SSR will degenerate to SR method [8]. With the increase of noise level, λ_2 should be set to larger values. This indicates that the sparsity difference constraint plays an important role in removing the noise when reconstructing the target HR face image. It is worth noting that we set λ_1 and λ_2 to obtain the best performance in terms

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

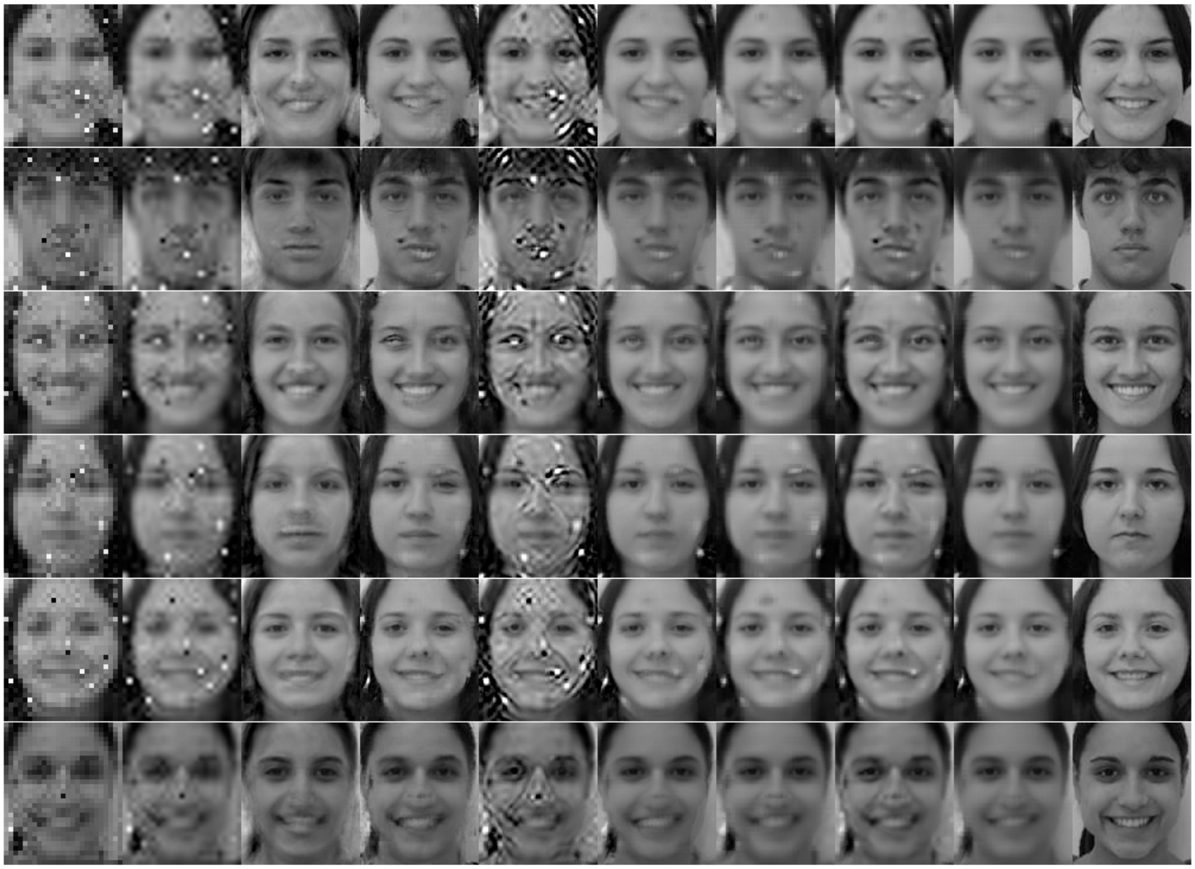


Fig. 9. Visual reconstruction results of six subjects with impulse noise using different methods (from left to right, they are the input LR image, results of bicubic interpolation, Wang and Tang's method [25], NE [33], LSR [15], SR [8], WSR [40], LcR [46], and our proposed SSR method, and the ground truth HR face images).

of average PSNR and SSIM of all 40 test images. Therefore, the best parameter settings are used for all test images, rather than setting up separate parameters for each one.

C. Comparisons Between SR and SSR

To verify the effectiveness of the proposed SSR strategy, in this section we compare the PSNR and SSIM values to evaluate SR[4] and SSR methods. Fig. 6 shows the PSNR (dB) and SSIM comparisons of all the 40 test images reconstructed by SR and our proposed SSR under different noise levels ($\sigma = 10, 30$). When $\sigma = 0$, SSR and SR [8] have the same performance. With the increase of noise level, the gain of SSR over SR becomes more obvious, e.g., 0.25 dB and 0.0061 in terms of PSNR and SSIM when $\sigma = 10$, 2.18 dB, and 0.0144 in terms of PSNR and SSIM when $\sigma = 30$. This certifies the effectiveness of introducing the smoothness and locality priors for SR.

In Fig. 7, we show the reconstructed results of one subject using SR and SSR with different noise levels (when $\sigma = 0$, SR and SSR have the same results). From these results, we learn that with the increase of noise level, the reconstructed faces by SR become dirtier and appear distorted. In contrast, our proposed SSR method can maintain the main facial structure and preserve more texture.

D. Compare With the State-of-the-Art Approaches

In order to prove the superiority of our proposed SSR-based face image super-resolution method, we further compare it with the state-of-the-art methods, such as Wang and Tang [25], NE [33], LSR [15], SR [8], weighted sparse regularization (WSR) [40], and LcR [46]. The Bicubic interpolation method is given as a baseline for comparison. To make a fair comparison among the comparison methods, we have carefully tuned the parameters of the competing methods to achieve their best performances.

The average PSNR and SSIM values are shown in Table II when the input images are corrupted by AWGN (with the standard deviation $\sigma = 10, 30$) or non-Gaussian noise (e.g., impulse noise by the random value with 2%). It can be seen from the results that the proposed SSR method is the best when the LR input is contaminated by noise. The PSNR and SSIM gains of SSR over the second best method (marked by blue) are 0.25 dB and 0.0061 when $\sigma = 10$ and 0.61 dB and 0.0132 when $\sigma = 30$, and 0.18 dB and 0.0103 when the input is contaminated by impulse noise.

Figs. 8 and 9 show the reconstructed results with AWGN LR face and impulse noisy LR faces, respectively. It is worth noting that impulse noise is one of the most frequently

[4] It should be noted that we use the sparse learning with efficient projections toolbox to solve the sparse coding problem of (4) to facilitate a fair comparison.

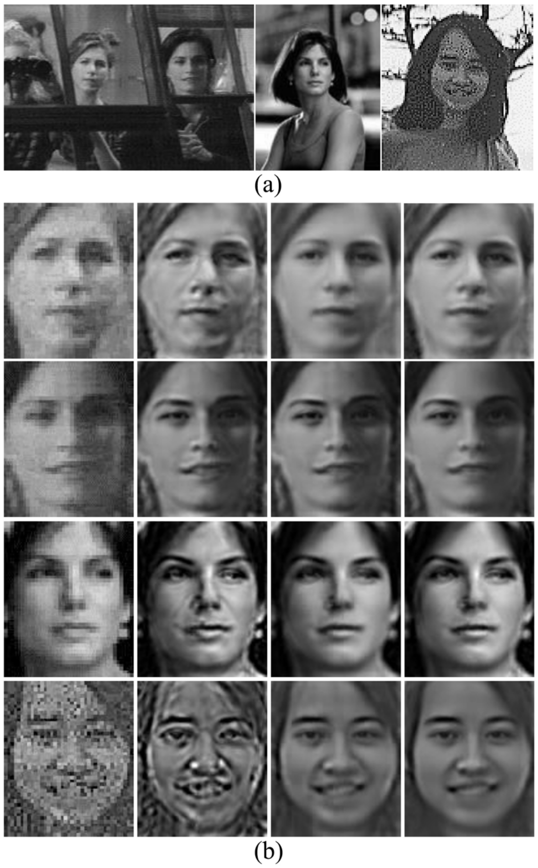


Fig. 10. Visual reconstruction results of four subjects from the CMU+MIT face database using SR [8], LcR [46], and our proposed SSR. (a) Captured pictures. (b) Reconstructed results.

encountered noise in real conditions [71], [72]. From Fig. 8, we can learn that, due to the smoothness and locality priors in our proposed method, the results have the least ringing effects and most detailed features (e.g., edges and corners) are quite close to the ground truth HR face images. The results of bicubic interpolation and LSR [15] are smooth (when the noise level is low), dirty and noisy (when the noise level is high). The results of NE [33] and SR [8] lack of clear contours and have serious blocking artifacts. When compared to LcR [46], our results are very competitive and have relative clear edges (see the mouths and face contours in the seventh and eighth columns of Fig. 8). As for the super-resolved results with impulse noise (Fig. 9), it is difficult for these traditional patch-based methods to remove added impulse noise. NE method [33] introduces some dirty high-frequency information, while LSR [15], SR [8], WSR [40], and LcR [46] still maintain some noise. Wang's global face method can remove most of the noise, but have obvious ghost effect.

E. Super-Resolution Results With Real-World Images

In the above experiments, the input LR face to be reconstructed are obtained by simply smoothing, downsampling and adding AWGN. However, in reality, the image degradation process is much more complex. Face image super-resolution in real-world is an extreme complex and difficult problem [73]. To certify the effectiveness of the proposed method, we give some super-resolved results with LR inputs in real-world conditions. Note that the training samples are all from the FEI face database. Fig. 10(a) shows three pictures with LR and noise from the CMU+MIT face database [67]. Fig. 10(b) compares the reconstructed results of our proposed SSR method, SR [8] method, and the state-of-the-art LcR method [46]. The first column in Fig. 10(b) represents the input LR face images, and the second to the fourth column represent the reconstructed HR faces by SR [8], LcR [46], and our proposed SSR, respectively. Although the input LR face images are of low-quality (contaminated with heavy noise) and different from the training samples, the proposed SSR method performs quite well for the super-resolution task. SR method [8] has obvious artifacts, while LcR method [46] produced varying degrees of "ghost effect" at face contours.

V. Conclusion

In this paper, we propose a new model for face image super-resolution based on SSR. By combining the strengths of fused LASSO and the locality constrained representation, it can achieve considerable improvement over several existing state-of-the-art face image super-resolution models both quantitatively and qualitatively when the input LR face is contaminated by high level of noise. In future work, we will develop a multiresolution face SR method [74]. In addition, how to learn a dictionary that is beneficial to smoothness and sparsity will also be investigated. Lastly, in the experiments, we found that the strategy of overlap patch representation and reconstruction is very time consuming, which hinders the practical applications. Thanks to the independence of different input image patches, we can move forward to accelerate the algorithm via parallel computation [75], [76].

References

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

[2] J. Jiang, X. Ma, Z. Cai, and R. Hu, "Sparse support regression for image super-resolution," *IEEE Photon. J.*, vol. 7, no. 5, pp. 1–11, Oct. 2015.

[3] C. Chen and J. E. Fowler, "Single-image super-resolution using multihypothesis prediction," in *Proc. ASILOMAR*, Pacific Grove, CA, USA, Nov. 2012, pp. 608–612.

[4] 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.

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

[6] X. Li, H. He, R. Wang, and J. Cheng, "Superpixel-guided nonlocal means for image denoising and super-resolution," *Signal Process.*, vol. 124, pp. 173–183, Jul. 2016.

[7] Y. Hu, N. Wang, D. Tao, X. Gao, and X. Li, "SERF: A simple, effective, robust, and fast image super-resolver from cascaded linear regression," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4091–4102, Sep. 2016.

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

[9] X. Li, H. He, R. Wang, and D. Tao, "Single image superresolution via directional group sparsity and directional features," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2874–2888, Sep. 2015.

[10] Y. Tang and Y. Yuan, "Learning from errors in super-resolution," *IEEE Trans. Cybern.*, vol. 44, no. 11, pp. 2143–2154, Nov. 2014.

[11] 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.

[12] 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, Sep. 2016.

[13] 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.

[14] J. Yang, H. Tang, Y. Ma, and T. Huang, "Face hallucination VIA sparse coding," in *Proc. ICIP*, San Diego, CA, USA, Oct. 2008, pp. 1264–1267.

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

[16] R. Farrugia and C. Guillemot, "Face hallucination using linear models of coupled sparse support," *arXiv preprint arXiv:1512.06009*, 2015.

[17] C. Peyrard, M. Baccouche, F. Mamalet, and C. Garcia, "ICDAR2015 competition on text image super-resolution," in *Proc. ICDAR*, Tunis, Tunisia, 2015, pp. 1201–1205.

[18] C. Dong, X. Zhu, Y. Deng, C. C. Loy, and Y. Qiao, "Boosting optical character recognition: A super-resolution approach," *arXiv preprint arXiv:1506.02211*, 2015.

[19] N. Wang, J. Li, D. Tao, X. Li, and X. Gao, "Heterogeneous image transformation," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 77–84, 2013.

[20] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "Transductive face sketch-photo synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1364–1376, Sep. 2013.

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

[22] 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.

[23] Y. Li, C. Cai, G. Qiu, and K.-M. Lam, "Face hallucination based on sparse local-pixel structure," *Pattern Recognit.*, vol. 47, no. 3, pp. 1261–1270, 2014.

[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. CVPR*, vol. 1. Kauai, HI, USA, 2001, pp. I-192–I-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] H. Huang, H. He, X. Fan, and J. Zhang, "Super-resolution of human face image using canonical correlation analysis," *Pattern Recognit.*, vol. 43, no. 7, pp. 2532–2543, 2010.

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

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

[30] J. Jiang, R. Hu, C. Liang, Z. Han, and C. Zhang, "Face image super-resolution through locality-induced support regression," *Signal Process.*, vol. 103, pp. 168–183, Oct. 2014.

[31] Y. Li and C. Qi, "Position constraint based face image super-resolution by learning multiple local linear projections," *Signal Process. Image Commun.*, vol. 32, pp. 1–15, Mar. 2015.

[32] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.

[33] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. CVPR*, vol. 1. Washington, DC, USA, 2004, pp. I275–I282.

[34] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

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

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

[37] 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.*, to be published, doi: 10.1109/TCSVT.2015.2433538.

[38] 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.

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

[40] 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.

[41] Z.-Y. Wang, Z. Han, R.-M. Hu, and J.-J. Jiang, "Noise robust face hallucination employing Gaussian–Laplacian mixture model," *Neurocomputing*, vol. 133, pp. 153–160, Jun. 2014.

[42] J. Jiang, R. Hu, Z. Han, and Z. Wang, "Low-resolution and low-quality face super-resolution in monitoring scene via support-driven sparse coding," *J. Signal Process. Syst.*, vol. 75, no. 3, pp. 245–256, 2014.

[43] G. Gao and J. Yang, "A novel sparse representation based framework for face image super-resolution," *Neurocomputing*, vol. 134, pp. 92–99, Jun. 2014.

[44] S. Qu *et al.*, "Face hallucination via Cauchy regularized sparse representation," in *Proc. ICASSP*, South Brisbane, QLD, Australia, 2015, pp. 1216–1220.

[45] J. Shi, X. Liu, and C. Qi, "Global consistency, local sparsity and pixel correlation: A unified framework for face hallucination," *Pattern Recognit.*, vol. 47, no. 11, pp. 3520–3534, 2014.

[46] 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.

[47] 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.

[48] 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, Nov. 2016.

[49] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 67, no. 1, pp. 91–108, 2005.

[50] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Stat.*, vol. 1, no. 2, pp. 302–332, 2007.

[51] H. Gao and H. Zhao, "Multilevel bioluminescence tomography based on radiative transfer equation part 1: l1 regularization," *Opt. Exp.*, vol. 18, no. 3, pp. 1854–1871, 2010.

[52] A. Ahmed and E. P. Xing, "Recovering time-varying networks of dependencies in social and biological studies," *Proc. Nat. Acad. Sci.*, vol. 106, no. 29, pp. 11878–11883, 2009.

[53] L. Wang, K. Lu, P. Liu, R. Ranjan, and L. Chen, "IK-SVD: Dictionary learning for spatial big data via incremental atom update," *Comput. Sci. Eng.*, vol. 16, no. 4, pp. 41–52, Jul./Aug. 2014.

[54] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[55] L. Wang, W. Song, and P. Liu, "Link the remote sensing big data to the image features via wavelet transformation," *Clust. Comput.*, vol. 19, no. 2, pp. 793–810, 2016.

[56] K. Li, J. Yang, and J. Jiang, "Nonrigid structure from motion via sparse representation," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1401–1413, Aug. 2015.

[57] H. Kong, Z. Lai, X. Wang, and F. Liu, "Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning," *Neurocomputing*, vol. 177, pp. 198–205, Feb. 2016.

[58] C. Zhang *et al.*, "Fine-grained image classification via low-rank sparse coding with general and class-specific codebooks," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2016.2545112.

[59] C. Zhang *et al.*, "Beyond explicit codebook generation: Visual representation using implicitly transferred codebooks," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5777–5788, Dec. 2015.

[60] C. Zhang, Q. Huang, and Q. Tian, "Contextual exemplar classifier based image representation for classification," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2016.2527380.

[61] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[62] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[63] B. Eicke, "Iteration methods for convexly constrained ill-posed problems in hilbert space," *Numer. Funct. Anal. Optim.*, vol. 13, nos. 5–6, pp. 413–429, 1992.

[64] M. Piana and M. Bertero, "Projected landweber method and preconditioning," *Inverse Prob.*, vol. 13, no. 2, pp. 441–463, 1997.

[65] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[66] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vis. Comput.*, vol. 28, no. 6, pp. 902–913, 2010.

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

[68] 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.

[69] J. Ma *et al.*, "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.

[70] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[71] 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.

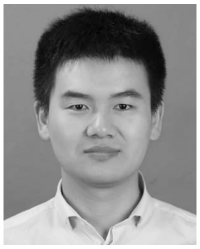
[72] C. L. P. Chen, L. Liu, L. Chen, Y. Y. Tang, and Y. Zhou, "Weighted couple sparse representation with classified regularization for impulse noise removal," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4014–4026, Nov. 2015.

[73] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Learning face hallucination in the wild," in *Proc. AAAI*, Austin, TX, USA, 2015, pp. 3871–3877.

[74] X. Lu and X. Li, "Multiresolution imaging," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 149–160, Jan. 2014.

[75] 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.

[76] 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.



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. He has authored and co-authored over 50 scientific articles and holds eight Chinese patents. His current research interests include image processing and computer vision.

Jiayi Ma (M'16) received the B.S. degree from the Department of Mathematics and the Ph.D. degree from the School of Automation, 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, where he was a Post-Doctoral Researcher from 2014 to 2015. His current research interests include computer vision, machine learning, and pattern recognition.

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 Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX, USA, in 2016.

He is currently a Post-Doctoral Researcher with the Center for Research in Computer Vision, University of Central Florida, Orlando, FL, USA. His current research interests include compressed sensing, signal and image processing, pattern recognition, and computer vision. He has published over 40 papers in refereed journals and conferences in the above areas.

Xinwei Jiang received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2012.

He is currently a Lecturer with the China University of Geosciences, Wuhan. His current research interests include nonparametric statistical models, machine learning, and dimensionality reduction.

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

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