Image Super-Resolution Based on Structure-Modulated Sparse Representation

Yongqin Zhang, *Member, IEEE*, Jiaying Liu, *Member, IEEE*, Wenhan Yang, and Zongming Guo, *Member, IEEE*

***Abstract*— Sparse representation has recently attracted enormous interests in the field of image restoration. The conventional sparsity-based methods enforce sparse coding on small image patches with certain constraints. However, they neglected the characteristics of image structures both within the same scale and across the different scales for the image sparse representation. This drawback limits the modeling capability of sparsity-based super-resolution methods, especially for the recovery of the observed low-resolution images. In this paper, we propose a joint super-resolution framework of structure-modulated sparse representations to improve the performance of sparsity-based image super-resolution. The proposed algorithm formulates the constrained optimization problem for high-resolution image recovery. The multistep magnification scheme with the ridge regression is first used to exploit the multiscale redundancy for the initial estimation of the high-resolution image. Then, the gradient histogram preservation is incorporated as a regularization term in sparse modeling of the image super-resolution problem. Finally, the numerical solution is provided to solve the super-resolution problem of model parameter estimation and sparse representation. Extensive experiments on image super-resolution are carried out to validate the generality, effectiveness, and robustness of the proposed algorithm. Experimental results demonstrate that our proposed algorithm, which can recover more fine structures and details from an input low-resolution image, outperforms the state-of-the-art methods both subjectively and objectively in most cases.**

***Index Terms*— Super-resolution, ridge regression, sparse representation, dictionary learning, gradient histogram.**

I. Introduction

IN REAL-WORLD scenarios, the low-resolution (LR) images are generally captured in many imaging applications, such as surveillance video, consumer photographs, remote sensing, magnetic resonance (MR) imaging and video standard conversion [1]. The resolution of images is limited by the image acquisition devices, the optics, the hardware storage and other constraints in digital imaging systems. However, high-resolution (HR) images or videos are usually desired for subsequent image processing and analysis in most real applications. As an effective way to solve this problem, super-resolution (SR) techniques aim to reconstruct HR images from the observed LR images. The super-resolution reconstruction increases high-frequency components and removes the undesirable effects, *e.g.*, the resolution degradation, blur and noise.

The problem of image super-resolution was first studied by Tsai and Huang in 1980s [2]. Subsequently, many SR techniques have been proposed over the last three decades. Early SR studies mainly focus on exploring the shift and aliasing properties of the Fourier transform. Although these approaches are computationally efficient, they have limited abilities to model the complicated image degradation and various image priors. Due to these drawbacks of frequency domain approaches, the spatial domain approaches are very popular recently for their flexibility to model all kinds of image degradations. After a brief review of the development of super-resolution technologies, according to the number of input LR images, the super-resolution approaches [3] can be broadly classified into two major categories: multi-frame super-resolution [4]–[8] and single-image super-resolution [9]–[14]. More specifically, there are two basic groups for multi-frame super-resolution methods. One group is static super-resolution [1], which can be further classified into the frequency domain methods [2], [15], the non-uniform interpolation methods [16]–[18], the statistical methods [4], [5], [19]–[22], and Projection onto Convex Sets (POCS) [23]. The POCS is convenient for incorporating any kind of constraints or priors. However, the POCS, whose solution depends on the initial values, has the drawbacks of heavy computation and slow convergence. The other group is dynamic super-resolution [6], [24]–[26], which utilizes the previous reconstructed HR frames to estimate the current HR frame.

Correspondingly, single-image super-resolution methods can also be further divided into interpolation-based methods [27], [28], reconstruction-based methods [13] and example learning-based methods [9], [10], [29]. The interpolation-based methods usually utilize a base function to construct the unknown data points on the regular grids of HR images. Although they have the advantage of relatively low complexity, the interpolation-based methods tend to

Manuscript received July 21, 2014; revised January 10, 2015 and April 16, 2015; accepted April 21, 2015. Date of publication May 8, 2015; date of current version May 22, 2015. This work was supported in part by the Beijing Natural Science Foundation under Contract 4142021, in part by the China Post-Doctoral Science Foundation under Grant 2013M530481 in part by the National Natural Science Foundation of China under Contract 61201442, and in part by the National High-Tech Technology Research and Development Program (863 Program) of China under Grant 2014AA015205. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xin Li. *(Corresponding author: Jiaying Liu.)*

Y. Zhang is with the School of Information Science and Technology, Northwest University, Xi'an 710127, China (e-mail: zhangyongqin@pku.edu.cn).

J. Liu and W. Yang are with the Institute of Computer Science and Technology, Peking University, Beijing 100871, China (e-mail: liujiaying@pku.edu.cn; yangwenhan@pku.edu.cn).

Z. Guo is with the Institute of Computer Science and Technology, Peking University, Beijing 100871, China, and also with the Cooperative Medianet Innovation Center, Shanghai 201805, China (e-mail: guozongming@pku.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/TIP.2015.2431435

produce considerable edge halos, blurring and aliasing artifacts. Therefore, this class of SR methods is often insufficient for practical applications.

The reconstruction-based methods [30], [31] usually incorporate the reconstruction constraints or the prior knowledge to model a regularized cost function with a data-fidelity term. The typical image priors include the gradient priors [13], [32]–[35], the nonlocal self-similarity priors [36]–[39] and the sparsity priors [39]–[43]. These different priors characterize different and complementary aspects of natural image features. Therefore, the combination of multiple image priors for SR modeling may be beneficial to the improvement of the SR performance. This family of methods has the ability to recover sharp edges and suppress aliasing artifacts. However, the reconstruction-based methods, whose performance depends heavily on the priors imposed on the HR images, are unable to restore the fine structures when the upscaling factor is larger.

The example learning-based methods exploit the information from training images or example images to learn the mapping between the LR and HR image patches for super-resolution reconstruction. Recently, numerous SR methods have appeared to estimate the relationship between the LR and HR image patches with promising results. Some typical methods [9], [11], [12] usually need a large and representative database of the LR and HR image pairs to encompass various images as much as possible that leads to a heavy computational load in the mapping learning process. Glasner *et al.* [10] implies that if the structural patterns of the input LR image do not appear in a general image database, the mapping learned from the database may not be able to restore the faithful high-frequency details in the HR image. Yang *et al.* [11] employed sparse dictionary learning on the LR and HR image patches from a general image database, and then utilized sparse representations of the LR input to generate the output HR image. Dong *et al.* [44] proposed a deep learning method that learns an end-to-end mapping between the LR and HR images for single image super-resolution. Michaeli and Irani [45] exploited the inherent recurrence property of small natural image patches to estimate the optimal blur kernel for blind super-resolution. Timofte *et al.* [46] introduced the anchored neighborhood regression (ANR) that learns sparse dictionaries and regressors anchored to the dictionary atoms for fast super-resolution. Subsequently, they [47] proposed an improved variant of ANR that achieves substantially less complexity and better performance. Similarly, Perez-Pellitero *et al.* [48] presented an improved training strategy for SR linear regressors and an inverse-search approach for the speedup of the regression-based SR method. In this paper, we mainly focus on the study of the example learning-based SR methods with multiple image priors for further improvements of single image super-resolution. The optimized example learning-based SR method will build a suitable training set and make full use of image priors to reduce edge halos, blurring and aliasing artifacts effectively.

It is observed that the local patterns in natural images tend to redundantly repeat both within the same scale and across different scales [10]. Inspired by multi-scale self-similarities, sparse representation and structural distribution similarities of natural images, we propose a novel joint framework of the structure-modulated sparse representation (SMSR) for single image super-resolution. The multi-scale similarity redundancy is investigated and exploited for the initial estimation of the target HR image. The image gradient histogram of a LR input is incorporated as a gradient regularization term of the image sparse representation model. The proposed SMSR algorithm employs the gradient prior and nonlocally centralized sparsity to design the constrained optimization problem for dictionary training and HR image reconstruction. The main contributions of our work can be summarized as follows:

- The multi-step magnification scheme with the ridge regression is proposed to initialize the target HR image for the solution of image SR problem;
- The novel sparsity-based super-resolution model is proposed with the combination of multiple image priors on the structural self-similarity, the gradient histogram and the nonlocal sparsity;
- The gradient histogram preservation (GHP) is theoretically deduced for image SR reconstruction and also incorporated as the regularization term for the sparse modeling of HR image recovery.

The remainder of this paper is organized as follows. Related work is reviewed in Section II. Section III provides the detailed descriptions of the proposed SMSR algorithm. The experimental results and analysis are given in Section IV. The conclusions and future work are drawn in Section V.

II. Related Work

The task of single image super-resolution is to recover a HR image from an input LR image. For an observed image $\mathbf{y}$, the problem of image super-resolution is generally modeled as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \upsilon, \tag{1}$$

where the degradation matrix $\mathbf{H}$ is a composite operator of blurring and down-sampling, $\mathbf{x}$ is the original image, and υ is the noise term. In the past decades, many works have been reported on single image super-resolution. Due to the ill-posed nature of the SR inverse problem, the regularization is introduced to eliminate the uncertainty of recovery. Several regularization-based techniques have been extensively studied in the recent literatures [7], [49]–[52]. The typical regularization models include the total variation (TV) [7], [49], the nonlocal similarity [52] and the sparsity-based regularization [50], [51]. The TV regularization was introduced in image processing and successfully applied to inverse problems. Since its piecewise constant assumption, the TV regularization [7], [49] tends to over-smooth the images.

To recover solutions which have discontinuities or are spatially inhomogeneous, the sparsity-based regularization has appeared and attracted great attention for image super-resolution problems in recent years [11], [12], [39], [41]. Specifically, for an input LR image $\mathbf{y} \in \mathbb{R}^{\mathrm{M}}$, let $\mathbf{x} \in \mathbb{R}^{\mathrm{N}}$ and $\hat{\mathbf{x}} \in \mathbb{R}^{\mathrm{N}}$ denote the HR and reconstructed HR images, respectively. Correspondingly, $\mathbf{y}_i \in \mathbb{R}^m$, $\mathbf{x}_i \in \mathbb{R}^n$ and $\hat{\mathbf{x}}_i \in \mathbb{R}^n$, $i = 1, \cdots, l$ represent the LR, HR, and reconstructed

TABLE I
PSNR(dB)/SSIM RESULTS ON THE RECONSTRUCTED HR IMAGES OBTAINED BY THE NCSR METHOD WITH DIFFERENT INTERPOLATIONS APPLIED TO THE *'Butterfly'* IMAGE FOR DIFFERENT SCALING RATIOS

Ratios	Oracle	SMSR1	Bicubic	Bilinear	Nearest
×2	**31.32/0.9343**	30.50/0.9342	30.33/0.9273	30.32/0.9272	29.40/0.9206
×3	**29.15/0.9270**	28.26/0.9192	28.09/0.9160	28.12/0.9164	27.22/0.8988

Fig. 1. Visual comparisons of the super-resolution results of the NCSR method with the different interpolations for the *'Butterfly'* image (the scaling factor 3). From left to right: (a) Ground truth, (b) the oracle interpolation, (c) SMSR1, (d) the bicubic interpolation, (e) the bilinear interpolation, and (f) the nearest-neighbor interpolation.

HR image patches, respectively, where l is the total number of overlapped image patches. Assuming that $\mathbf{R}_i \in \mathbb{R}^{n\times \mathrm{N}}$ denotes the extracting matrix, a HR image patch can be written as $\mathbf{x}_i = \mathbf{R}_i\mathbf{x}$. According to the observation model (1), the SR problem is first formulated by sparse coding of $\mathbf{y}$ with respect to Φ as follows [51]:

$$\alpha_{\mathbf{y}} = \arg\min_{\alpha} \left\{ \|\mathbf{y} - \mathbf{H}\Phi\alpha\|_2^2 + \lambda\|\alpha\|_1 \right\}, \tag{2}$$

where the Lagrange multiplier λ is a slack variable that balances the tradeoff between fitting the data perfectly and employing a sparse solution, and $\|\alpha\|_1$ is the sparsity-inducing term. Then the HR image $\mathbf{x}$ is reconstructed by

$$\hat{\mathbf{x}} = \Phi \circ \alpha_{\mathbf{y}} = \left(\sum\nolimits_{i=1}^{l} \mathbf{R}_i^T \mathbf{R}_i \right)^{-1} \sum\nolimits_{i=1}^{l} \left(\mathbf{R}_i^T \Phi \alpha_{\mathbf{y},i} \right). \tag{3}$$

Note that $\Phi \in \mathbb{R}^{\mathrm{N}\times \mathrm{M}}$ ($\mathrm{N} < \mathrm{M}$) is an overcomplete dictionary, and most entries of coding vectors α are zero or close to zero.

The choice of dictionary Φ is a critical issue in sparse representation modeling. There are two main categories for dictionary selection: the analytical dictionaries and the learning dictionaries. The analytical dictionaries, such as DCT, wavelets, curvelets and contourlets, are generally highly structured and efficient. Although such analytical dictionaries can be achieved by a fast transform, they have limited adaptive ability for different types of data due to the fixed data representation. The learning dictionaries trained from image examples can better characterize the image features and thus cause much performance improvements of the sparse representation [53]–[57]. These methods aim at learning a universal overcomplete dictionary to represent various image structures. But sparse decomposition over a highly redundant dictionary is potentially unstable and tends to cause visual artifacts. Recently, sparse coding with adaptive dictionary learning in LR and HR spaces has become the focus of ongoing research on image super-resolution for its high efficiency for signal modeling. Yang *et al.* [11] proposed the coupled dictionaries trained from the LR and HR image patch pairs for single image super-resolution. Kim and Kwon [12] adopted kernel ridge regression (KRR) to learn a map from input LR images to target HR images based on example pairs of input and output images. Although these methods [11], [12] assume that the similarities of geometry and sparse representation exist between the LR and HR spaces, they do not explicitly point out the mapping from the LR input to the related optimal HR counterpart. Dong *et al.* [41] proposed an adaptive sparse domain selection (ASDS) model for image SR recovery, where the piecewise autoregressive (AR) models and a nonlocal self-similarity constraint are formulated as the regularization terms for more effective reconstruction. Considering the sparse coding noise, the local and nonlocal sparsity constraints, they further proposed a nonlocally centralized sparse representation (NCSR) model for image super-resolution with very encouraging performance [39]. Specifically, the NCSR method adopts the bicubic interpolation to amplify the input LR image to estimate the initial value of the target HR image. To the best of our knowledge, the NCSR method is one of the best state-of-the-art methods for image super-resolution. But it does not consider the geometric distribution similarity between the LR and HR images, and its performance heavily depends on the initial estimation of the target HR image. In order to illustrate that problem, we first adopted different interpolation approaches to initialize the NCSR method, and then evaluated the results of one example. Let SMSR1 denote the first stage output of our proposed SMSR method, which is made up of two stages: the gradual magnification and the structured sparse representation. Table I shows the PSNR(dB)/SSIM results of the NCSR method with different initial values obtained by the nearest-neighbor interpolation, the bilinear interpolation, the bicubic interpolation and the oracle interpolation, respectively. For the oracle, the original HR image is assumed to be the initial value of the target HR image in the NCSR method. In practice, the oracle interpolation is not feasible due to the lack of the original HR image. The visual comparison of the results obtained by the NCSR method with different initial values is given in Fig. 1.

As can be seen from the reported results, the NCSR method with the bicubic interpolation has better performance than that with the nearest-neighbor interpolation or the bilinear interpolation in most cases. However, the NCSR method with

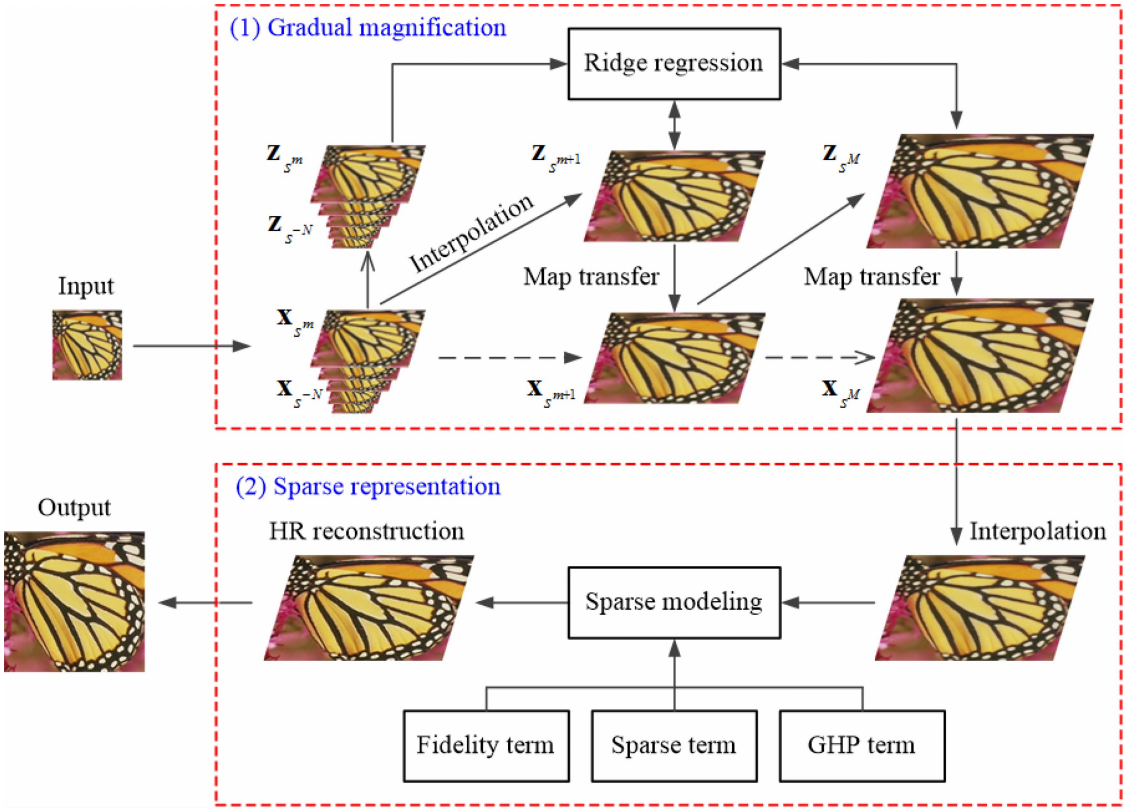


Fig. 2. Overview of the proposed SMSR method. The subgraphs with the solid boxes denote the specific techniques. The subgraph with the dashed box is the algorithm modules: gradual magnification, and sparse representation.

the bicubic interpolation still causes the smooth edges and blurred textures in the reconstructed HR images. The reason is that these learning-based SR methods ignore the geometry constraints between similar distributions of the LR and HR images, and also do not take full advantage of similarity redundancy both within the same scale and across different scales. Specifically, the NCSR method does not consider other image priors, such as the gradient histogram preservation, which may be beneficial to the improvement of image super-resolution. Therefore, there is still much space to further improve the performance of single image SR by exploiting prior knowledge of natural images.

III. Proposed SMSR Algorithm

A. Overview of the SMSR Model

The flowchart of our proposed SMSR method is shown in Fig. 2. Given an input LR image, our goal is to produce a suitable HR image such that its underlying high-frequency details are recovered while preserving the intrinsic geometrical structures of original HR image. In brief, our SMSR algorithm consists of two stages: the gradual magnification and the structured sparse representation. The basic procedures of our SMSR method are given as follows. Firstly, for an input LR image $\mathbf{y}$, the HR database $\mathbf{D_x}$ and its corresponding LR database $\mathbf{D_z}$ are separately built for the gradual magnification. Then, the LR image $\mathbf{z}_{s^m}$ at the m-th is estimated from the HR image $\mathbf{x}_{s^{m-1}}$ by the bicubic interpolation method where $m = 1, \cdots, M$. The ridge regression is applied to both each query patch of $\mathbf{z}_{s^m}$ and its k_n nearest patches, which are found from the LR database $\mathbf{D_z}$ by the approximate nearest-neighbor (ANN) searches.[1] Therefore the corresponding HR image $\mathbf{x}_{s^m}$ is reconstructed from the fitted coefficients and its k_n HR nearest patches in the HR database $\mathbf{D_x}$ by the map transfer. At the end of each step, the reconstructed HR image $\mathbf{x}_{s^m}$ and the LR image $\mathbf{z}_{s^m}$ at the m-th scale are added to the image databases $\mathbf{D_x}$ and $\mathbf{D_z}$, respectively. These procedures of the multi-step magnification technique are iterated until the desired HR image $\mathbf{x}_{s^M}$ is achieved. To generate the initial value $\mathbf{x}_H^0$ of the target HR image $\mathbf{x}$, the estimated HR image $\mathbf{x}_{s^M}$ may be blurred and downsampled to achieve the same size of $\mathbf{x}$ by the bicubic interpolation method. Next, an iteration method for solving the problem of the image sparse representation begins with dictionary learning. Our dictionary learning adopts a multi-class and multi-level training framework, which is the same as that of the NCSR method [39]. Subsequently, according to the estimated image signal deviation, the reference histogram of gradients is estimated. After computing the transform function, the HR image is successively updated by the gradient histogram preservation regularization, the data fidelity constraint and the nonlocal means. After that, the HR image is sparsely coded over the trained dictionaries. Each element of the sparse coefficients is further updated with

[1]The ANN package by Mount & Ayra, http://www.cs.umd.edu/~mount/ANN.

its nonlocal means by the shrinkage operator. The iteration proceeds until the convergence condition is reached. Finally, the target HR image $\mathbf{x}_H$ is reconstructed by gathering the small image patches with the weighted averaging method.

B. Gradual Magnification With Ridge Regression

For only a LR image available for SR reconstruction, we need to build a training set of the LR and HR image pairs to restore the high-frequency details lost in the LR input. Since there is the self-similarity redundancy both within the same scale and across different scales, the input LR image and its degraded versions can be used to construct the LR and HR image pairs. Considering the different sizes of the LR and HR images, the LR input $\mathbf{y}$ is enlarged to the same size of the HR image $\mathbf{x}$ by the bicubic interpolation, whereas it is also a blurred and downsampled version of the HR image $\mathbf{x}$. Therefore, the correspondence between the LR and HR images at the same scale is established as follows:

$$\mathbf{z} = (\mathbf{y}) \uparrow s = ((\mathbf{x} * \mathbf{G}) \downarrow s) \uparrow s = \mathbf{E}_s \mathbf{x}, \tag{4}$$

where $*$ is a convolution operator, $\uparrow s$ is an upsampling operator with the scaling factor s, $\downarrow s$ is a downsampling operator with the scaling factor s, $\mathbf{G}$ is a blurring kernel (*e.g.*, an isotropic Gaussian kernel with the standard deviation σ_G), $\mathbf{E}_s$ is a composite operator of blurring, downsampling and then upsampling with the scaling factor s, $\mathbf{y}$ is the LR image, $\mathbf{x}$ is the HR image, and $\mathbf{z}$ is the LR image at the same size of $\mathbf{x}$.

Note that the LR image is more similar to its HR image when the scaling factor is smaller, whereas the high-frequency details tend to be lost when the scaling factor is larger. Therefore, in our example learning-based SR framework, we adopt the multi-step magnification scheme with the ridge regression for the initial estimation of the target HR image. For an input LR image $\mathbf{y}$ with a total scaling factor d, $M = \lceil \log(d) / \log(s) \rceil$ is determined as the number of magnification steps, where s is the upscaling factor of each step. Specifically, the relationship between the HR image $\mathbf{x}_{s^m}$ at the m-th scale and the HR image $\mathbf{x}_{s^{m+1}}$ at the $(m+1)$-th scale can be expressed as follows:

$$\mathbf{x}_{s^m} = \left(\mathbf{x}_{s^{m+1}} * \mathbf{G}_s\right) \downarrow s. \tag{5}$$

And Eq. (4) in the multi-scale case can be written in another form:

$$\mathbf{z}_{s^m} = ((\mathbf{x}_{s^m} * \mathbf{G}_s) \downarrow s) \uparrow s = \mathbf{E}_s \mathbf{x}_{s^m}, \tag{6}$$

where the input LR image $\mathbf{y}$ is regarded as the HR image $\mathbf{x}_{s^0}$ at the scale $m = 0$. Specifically, the HR image $\mathbf{x}_{s^0}$ is blurred and downsampled to generate the HR image $\mathbf{x}_{s^m}$ at the scale $m = -1, \cdots, -N$. The LR image $\mathbf{z}_{s^0}$ is produced by first downsampling the HR image $\mathbf{x}_{s^0}$ and then upsampling the downsampled result by the bicubic interpolation.

In our multi-step magnification scheme, we build two training sets: the LR image database $\mathbf{D_z}$ and the HR image database $\mathbf{D_x}$. Specifically, the HR image $\mathbf{x}_{s^0}$ is blurred and downsampled to produce the HR image $\mathbf{x}_{s^m}$ at the scale $m = -1, \cdots, -N$. The HR image database $\mathbf{D_x}$ is constructed by collecting of these HR images $\mathbf{x}_{s^m}$ with the scale $m = -1, \cdots, -N$. Correspondingly, the LR image $\mathbf{z}_{s^0}$ is produced by first downsampling the HR image $\mathbf{x}_{s^0}$ and then upsampling the downsampled result by the bicubic interpolation. Like the generation of the LR image $\mathbf{z}_{s^0}$, the LR image $\mathbf{z}_{s^m}$ at the scale $m = -1, \cdots, -N$ is also obtained by first downsampling the HR image $\mathbf{x}_{s^m}$ and then upsampling the downsampled result by the bicubic interpolation. The collection of these LR images $\mathbf{z}_{s^m}$ with the scale $m = -1, \cdots, -N$ is used to construct the LR image database $\mathbf{D_z}$. Initially, the image databases $\mathbf{D_z}$ and $\mathbf{D_x}$ consist of all the patches of the LR image $\mathbf{z}_{s^m}$ and the HR image $\mathbf{x}_{s^m}$ at the scale $m = 0, \cdots, -N$, respectively. At each step of the multi-step magnification process, for the upscaling ratio s^m with $m = 1, \cdots, M$, $\mathbf{D_z}$ and $\mathbf{D_x}$ are separately constructed by gathering the LR images $\mathbf{z}_{s^n}$ and the HR images $\mathbf{x}_{s^n}$, where $n = m - 1, \cdots, -N$. First, the LR image $\mathbf{z}_{s^m}$ is estimated from the HR image $\mathbf{x}_{s^{m-1}}$ by the bicubic interpolation. For each patch $\mathbf{y}_{\mathbf{z}_i}$ of the LR image $\mathbf{z}_{s^m}$ at the m-th scale, the ANN method is used to find the k_n similar patches $\mathbf{N}_{k_n^l}$ in the LR database $\mathbf{D_z}$. We adopt the ridge regression to model the near-linear relationships between $\mathbf{y}_{\mathbf{z}_i}$ and $\mathbf{N}_{k_n^l}$ as the constrained optimization problem:

$$\min_{\gamma} \left\| \mathbf{y}_{\mathbf{z}_i} - \mathbf{N}_{k_n^l} \gamma \right\|_2^2 + \tau \| \gamma \|_2, \tag{7}$$

where γ is the coefficient vector, and τ is a regularization parameter that alleviates the singularity problems and stabilizes the solution. By solving this regularized least squares regression problem, the closed-form solution is given by

$$\gamma = \left(\mathbf{N}_{k_n^l}^T \mathbf{N}_{k_n^l} + \tau \mathbf{I}\right)^{-1} \mathbf{N}_{k_n^l}^T \mathbf{y}_{\mathbf{z}_i}, \tag{8}$$

where $\mathbf{I}$ is an identity matrix. Then through transferring the mapping relationship from the LR space to the HR space, the HR patch can be reconstructed by multiplying the same coefficient vector and the corresponding HR patches as follows:

$$\mathbf{y}_{\mathbf{x}_i} = \mathbf{N}_{k_n^h} \gamma, \tag{9}$$

where $\mathbf{y}_{\mathbf{x}_i}$ is the corresponding HR patch of the HR image $\mathbf{x}_{s^m}$ at the m-th scale, and $\mathbf{N}_{k_n^h}$ denotes the k_n similar HR patches in the HR image database $\mathbf{D_x}$. Next we compute all the HR patches $\mathbf{y}_{\mathbf{x}_i}$ and then reconstruct the HR image $\mathbf{x}_{s^m}$ by the weighted averaging method. The LR image $\mathbf{z}_{s^m}$ and the HR image $\mathbf{x}_{s^m}$ are separately added to the databases $\mathbf{D_z}$ and $\mathbf{D_x}$. Subsequently, the multi-step magnification scheme proceeds to the next step, which enlarges $\mathbf{x}_{s^m}$ to $\mathbf{x}_{s^{m+1}}$. Finally, after the M steps of magnification, the HR image $\mathbf{x}_{s^M}$ is obtained for subsequent processing procedures.

C. Structured Sparse Representation

1) SR Model: Since the HR image $\mathbf{x}_{s^M}$ reconstructed by the multi-step magnification scheme may be larger than the target HR image, the degraded version $\mathbf{x}_H^0$ of the right size is obtained by first blurring and then downsampling $\mathbf{x}_{s^M}$ with the bicubic interpolation. We use $\mathbf{x}_H^0$ as the initial value of the target HR image for sparse modeling of SR problem. Meanwhile, image gradients conveying abundant semantic

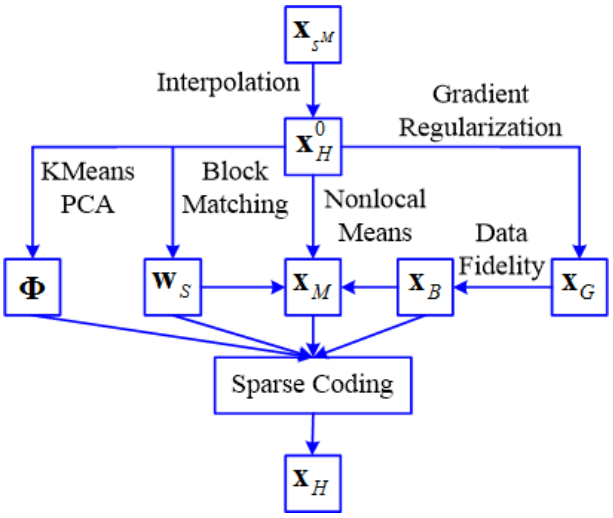


Fig. 3. The flowchart of our sparse representation of single image SR model.

features are crucial to the subjective visual image quality. Therefore, the histogram of image gradients can be used as a feature descriptor to constrain new image SR model. Besides that the gradual magnification is used as a preprocessing process for structured sparse representation, we propose the gradient histogram preservation regularization for single image SR modeling. That is, the gradient histogram of the reconstructed HR image should be close to that of the original HR image, which is estimated from the given LR image. Fig. 3 shows the flowchart of the proposed sparse representation of image SR model. Mathematically, our proposed structured sparse coding model of single image SR with multiple image priors is given as follows:

$$\alpha_{\mathbf{y}} = \arg\min_{\Phi,\alpha,f} \left\{ \begin{array}{l} \|y - \mathbf{H}\Phi \circ \alpha\|_2^2 \\ +\lambda_1 \sum_i \|\alpha_i - \beta_i\|_1 \\ +\lambda_2 \|f\,(\nabla\mathbf{x}|\nabla\mathbf{y}) - \nabla\mathbf{x}\|^2 \end{array} \right\}, \quad \text{s.t. } \mathbf{h}_f = \mathbf{h}_r, \tag{10}$$

where λ_1 and λ_2 are the regularization parameters, α_i is the coding coefficients of each patch $\mathbf{x}_i$ over the dictionary Φ, α denotes the concatenation of all α_i, β_i is the nonlocal means of α_i in the sparse coding domain, ∇ denotes the gradient operator, f is the transform function, $\mathbf{h}_r$ is the reference histogram of $\mathbf{x}$, and $\mathbf{h}_f$ is the histogram of the transformed gradient image $|f\,(\nabla\mathbf{x}|\nabla\mathbf{y})|$. Note that f is an odd function that is monotonically non-descending in the domain $(0, +\infty)$. On the right side of (10), the first term is the data fidelity of the solution, the second term is the sparse nonlocal regularization [39] and the third term is our proposed gradient regularization. Considering the natural images usually contain repetitive patterns [36], the nonlocal similar patches to the given patch $\mathbf{x}_i$ centered at pixel i are searched not only in the image spatial domain but also across different scales [10]. For the current estimate $\hat{\mathbf{x}}$, the similar patches of $\hat{\mathbf{x}}_i$ are denoted by $\hat{\mathbf{x}}_i^\ell$, whose coding coefficients are α_i^ℓ. Then β_i can be computed as the weighted average of the sparse codes of the associated nonlocal similar patches [39]:

$$\beta_i = \sum\nolimits_\ell \mathbf{w}_i^\ell \alpha_i^\ell, \tag{11}$$

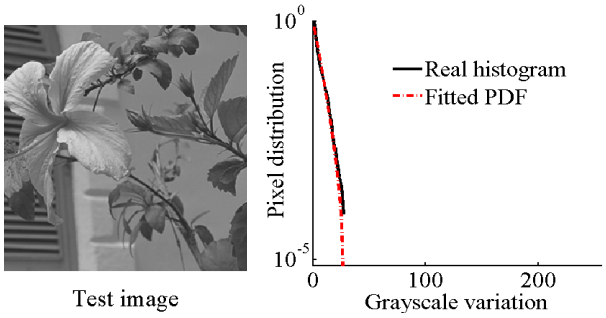


Fig. 4. One example and the gradient histogram of $\sum_i b_i \nabla\mathbf{x}_i$. (a) Test image, (b) the comparison between the real histogram and the fitted PDF for the test image.

where the weight $\mathbf{w}_i^\ell$ is defined as

$$\mathbf{w}_i^\ell = \frac{1}{W} \exp\left(-\left\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_i^\ell\right\|_2^2 / h\right), \tag{12}$$

where h is a control parameter to adjust the decay rate and W is a normalized factor to insure that $\sum_i \mathbf{w}_i^\ell = 1$.

2) Reference Histogram of Gradients: To solve the sparse coding problem in (10), we first need to know the reference histogram of gradients that is assumed to be the gradient histogram of the original HR image $\mathbf{x}$. As there is only one LR input $\mathbf{y}$ available, we use the gradient histogram of $\mathbf{y}$ to infer that of the original HR image $\mathbf{x}$. Different from the GHP for image denoising [34], we have extended the GHP regularization and provided the theoretical deduction for image super-resolution as follows. Let $\mathbf{z}$ denote the upsampled version of the LR image $\mathbf{y}$ that has the same size as the HR image $\mathbf{x}$. For the observation model of image SR problem, Eq. (1) can be simply rewritten in the following formula:

$$\mathbf{z} = \mathbf{B} * \mathbf{x}, \tag{13}$$

where $\mathbf{B}$ is a blurring operator. Thus we have

$$\nabla\mathbf{z} = \mathbf{B} * \nabla\mathbf{x} = b_0 \nabla\mathbf{x} + \sum\nolimits_i b_i \nabla\mathbf{x_i}, \tag{14}$$

where b_0 and b_i denotes the center coefficient and its surrounding neighbors of the blur kernel $\mathbf{B}$, respectively. Assume that each pixel in the gradient image $\nabla\mathbf{x}$ can be regarded as the value of a scalar random variable. The normalized histogram of $\nabla\mathbf{x}$ is seen as a discrete approximation of the probability density function (PDF) of the random variable x. Since the PDF of x can be well modeled by a generalized Gaussian distribution [58], the assumption holds for the distribution of the gradient image $\nabla\mathbf{x}$. According to the Lyapunov central limit theorem in probability theory, the sum of independent random variables, *i.e.*, $\sum_i b_i \nabla\mathbf{x_i}$, can be approximated by the normal distribution. Fig. 4 gives the real histogram distribution of $\sum_i b_i \nabla\mathbf{x}_i$ for one example and the fitted PDF distribution of the random Gaussian variable g. The fitted PDF of the random Gaussian variable g is very close to the real histogram distribution of $\sum_i b_i \nabla\mathbf{x}_i$ with 95% confidence bounds. The experiments verify that both the theoretical deduction of our SMSR model and these

assumptions are correct and hold in most cases. Then Eq. (14) can be reformulated in another form:

$$z \approx x + g, \tag{15}$$

where g is a random Gaussian variable subject to $N(0, \sigma_g^2)$, and z denotes a random variable for the distribution of the pixels in the gradient image $\nabla\mathbf{z}$.

In our algorithm, the 1D deconvolution model is built to estimate the reference histogram of gradients $\mathbf{h}_r$. Let $\mathbf{h}_g$ be a discrete version of the PDF of a random Gaussian variable g. In fact, the standard deviation σ_g is unknown in many image processing applications. To solve this problem, we use the signal variance of the input LR image $\mathbf{y}$ to estimate the distribution of g. According to Eq. (15), the reference histogram of gradients can be estimated by solving the deconvolution problem as follows:

$$\mathbf{h}_r = \arg\min_{\mathbf{h}_x} \left\{ \left\| \mathbf{h}_z - \mathbf{h}_x * \mathbf{h}_g \right\|^2 \right\}, \tag{16}$$

where $\mathbf{h}_z$ is the gradient histogram of the upsampled version of the LR input $\mathbf{y}$, $\mathbf{h}_x$ is the discrete version of the PDF of x that can be well modeled by a generalized Gaussian distribution [58], and $\mathbf{h}_g$ is the discrete version of the PDF of the independent and identically distributed (i.i.d.) random variable g.

3) Numerical Solution: The mentioned above problem of image super-resolution is non-convex and is hard to solve exactly in a reasonable time. In our algorithm, we propose an alternating minimization method to solve the image SR problem in (10) so that the constrained optimization is carried out with some variables fixed in cyclical fashion. First, the multi-step magnification scheme is used to enlarge the input LR image $\mathbf{y}$ to get the HR image $\mathbf{x}_{s^M}$ at the M-th scale. Then the initial value of the target HR image is acquired by downsampling $\mathbf{x}_{s^M}$ with the bicubic interpolation method. Next, the iterative solution process starts with dictionary learning like that of the NCSR method [39]. Specifically, for the current estimation of HR image $\mathbf{x}$, the k-means clustering is used to separate the patches of its multi-scale images into K clusters from each of which a sub-dictionary is trained by the principal component analysis (PCA). Subsequently, for each patch of the HR image, the PCA sub-dictionary of which cluster it belongs to is automatically selected as the dictionary Φ. For the fixed α_i, β_i and Φ, the image SR problem in (10) is reduced to the sub-problem of the GHP as follows:

$$\min_f \| f\left(\nabla\mathbf{x}|\nabla\mathbf{y}\right) - \nabla\mathbf{x}\|^2, \quad \text{s.t. } \mathbf{h}_f = \mathbf{h}_r. \tag{17}$$

Thus we can update the transform function f by solving the reduced sub-problem in (17). After that, for the fixed Φ and f, the image SR problem in (10) is reduced to the sub-problem in the following form:

$$\arg\min_{\alpha} \begin{Bmatrix} \|\mathbf{y} - \mathbf{H}\Phi \circ \alpha\|_2^2 + \lambda_1 \sum_i \|\alpha_i - \beta_i\|_1 \\ +\lambda_2 \| f\left(\nabla\mathbf{x}|\nabla\mathbf{y}\right) - \nabla\mathbf{x}\|^2 \end{Bmatrix}, \tag{18}$$

where the parameter λ_1 is used to weight the l_1-norm sparsity regularization term. Candes *et al.* [59] pointed out that iteratively reweighting the l_1-norm sparsity regularization term can lead to a better sparse representation. To improve the reconstruction of sparse signals, we adopt an adaptively reweighting method in [39], [41], and [59] that exploits the image nonlocal redundancy to estimate the regularization parameter λ_1. To solve the convex minimization sub-problem in (18), we first update the HR image $\mathbf{x}$ by the gradient descent method as follows:

$$\hat{\mathbf{x}}^{(t+1/2)} = \hat{\mathbf{x}}^{(t)} + \delta \begin{pmatrix} \mathbf{H}^T \left(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}^{(t)}\right) \\ +\lambda_2 \nabla^T \left(\mathbf{f} - \nabla\hat{\mathbf{x}}^{(t)}\right) \end{pmatrix}, \tag{19}$$

where δ is a small constant. The update process of HR image in (19) can be divided into two stages: the gradient regularization and the data fidelity constraint. Therefore, the sparse coding coefficients α_i are updated as follows:

$$\alpha_i^{(t+1/2)} = \Phi_k^T \mathbf{R}_i \hat{\mathbf{x}}^{(t+1/2)}, \tag{20}$$

where $\Phi_k, k = 1, \cdots, K$ is the PCA sub-dictionary of a cluster that the patch $\hat{\mathbf{x}}_i$ belongs to. The nonlocal means β_i of α_i can be estimated by Eq. (11). By employing the iterative shrinkage operator [50] applied to each element of α_i, we can further update the coding coefficients α_i in the following form:

$$\alpha_i^{(t+1)} = S_{\lambda/c} \begin{pmatrix} \Phi^T \circ \mathbf{H}^T \left(\mathbf{y} - \mathbf{H}\Phi \circ \alpha_i^{(t+1/2)}\right)/c \\ +\alpha_i^{(t+1/2)} - \beta_i \end{pmatrix} + \beta_i, \tag{21}$$

where $S_{\lambda/c}$ is the soft thresholding function, and c is a regulatory parameter to ensure the convexity of the shrinkage function. Finally, the whole HR image is reconstructed as:

$$\begin{aligned} \hat{\mathbf{x}}^{(t+1)} &= \Phi^{(t+1)} \circ \alpha^{(t+1)} \\ &= \left(\sum\nolimits_{i=1}^{l} \mathbf{R}_i^T \mathbf{R}_i \right)^{-1} \sum\nolimits_{i=1}^{l} \left(\mathbf{R}_i^T \Phi_k^{(t+1)} \alpha_i^{(t+1)} \right). \end{aligned} \tag{22}$$

Note that our proposed image SR model in (10) is similar in mathematical form to the one discussed by Attouch *et al.* [60], [61]. They proved that the minimization of the nonsmooth nonconvex objective function in the form of (10) can reach an abstract convergence result with descent methods under certain conditions. These conditions satisfy a sufficient-decrease assumption and allow a relative error tolerance. The above iterative procedures are executed repeatedly until the convergence condition is achieved [60], [61]. The theoretical convergence analysis of the proposed algorithm is left for our future research. More specifically, we select a certain number of iterations or a preset error tolerance as the convergence conditions of our algorithm for the simplicity.

D. Method Summary

To give further clarification of the specific implementation of our proposed SMSR algorithm, it is summarized in **Algorithm 1**.

IV. Experimental Results

In this work, numerous experimental studies on image super-resolution were carried out to verify the performance of our proposed SMSR algorithm. In our experiments, the basic parameters of our SMSR algorithm are set as follows: the patch size is 6×6 with the overlap width equal to 4

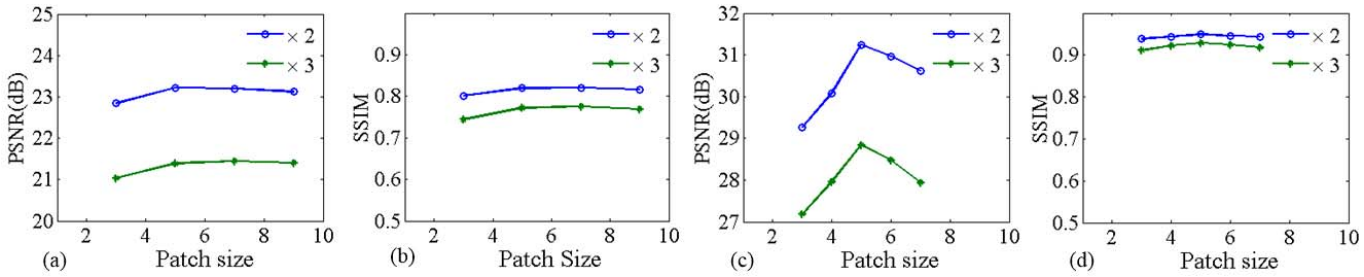


Fig. 5. The PSNR(dB)/SSIM results as a function of the patch size in the case of the different upscaling factors for the *'Butterfly'* image. From left to right and top to bottom: (a) PSNR(dB) results of our SMSR1, (b) SSIM results of our SMSR1, (c) PSNR(dB) results of SMSR2 and (d) SSIM results of SMSR2.

Algorithm 1 Pseudocodes of the SMSR-Based Image Super-Resolution

Input: a LR image $\mathbf{y}$ and a total scaling factor d.
Output: a HR image $\mathbf{x}_H$.

I. Initialization

- Set the initial parameters λ_1, λ_2, δ and c;
- Through exploiting the multi-scale similarity redundancy, the input LR image $\mathbf{y}$ is enlarged to obtain $\mathbf{x}_{s^M}$ by the multi-step magnification scheme;
- Set the initial value of the target HR image that is d times the size of $\mathbf{y}$ by dowmsampling the enlarged result $\mathbf{x}_{s^M}$;

II. Outer loop (dictionary learning and sparse coding): for each iteration $t = 1$ to T_1 do

- Update the dictionaries $\{\mathbf{\Phi}_k\}$ by means of k-means clustering and PCA;
- Compute the transform function f with the reference gradient histogram $\mathbf{h}_r$, and update the HR image $\hat{\mathbf{x}}^{(t)}$ by the gradient regularization;
- Inner loop: for each iteration $j = 1$ to T_2 do
  1) Update $\hat{\mathbf{x}}^{(t+1/2)}$ by the fidelity constraint;
  2) Compute the sparse coding coefficients of each patch $\alpha_i^{(t+1/2)} = \mathbf{\Phi}_k^T \mathbf{R}_i \hat{\mathbf{x}}^{(t+1/2)}$, where $\mathbf{\Phi}_k$ is the dictionary assigned to the patch $\hat{\mathbf{x}}_i = \mathbf{R}_i \hat{\mathbf{x}}^{(t+1/2)}$;
  3) Compute the regularization parameter λ_1 and the nonlocal means β_i of $\alpha_i^{(t+1/2)}$;
  4) Update the coding coefficients $\alpha_i^{(t+1)}$ again by the iterative shrinkage operator using (21);
  5) Reconstruct the estimate $\hat{\mathbf{x}}^{(t+1)}$ using (22);
- Update the HR image $\mathbf{x}_H = \hat{\mathbf{x}}^{(t+1)}$.

between the adjacent patches, $K = 64$, $\lambda_2 = 0.1$, $\delta = 7$, $\varepsilon = 0.3$, $c = 0.35$, $s = 1.25$, and $\tau = 0.1$. Two cases of magnification ratios $d = 2$ and $d = 3$ were separately implemented for the test images. For $d = 2$, $T_1 = 7$, $T_2 = 40$ and $M = 4$, whereas for $d = 3$, $T_1 = 6$, $T_2 = 160$ and $M = 5$. The proposed SMSR algorithm was also compared with the bicubic interpolation method and the state-of-the-art methods published recently [39], [41] for verifying its validity both subjectively and objectively. Indeed, our SMSR algorithm consists of two stages: the gradual magnification and the structured sparse representation (See Fig. 2). In this paper, for the convenience of our description, we refer to these two output stages as 'SMSR1' and 'SMSR2', respectively, where the output result of 'SMSR2' is also the final output of our proposed SMSR algorithm.

Note that both the ScSR method [11] and the ANR method [46] adopt the bicubic interpolation for the generation of the simulated LR images in their reported experimental results. However, for a fair comparision, just like the other SR methods [39], [41], the Gaussian blurring and downsampling operator has been used to generate the simulated LR images in our experiments. That is, a HR image is first blurred with a 7×7 Gaussian kernel with standard deviation 1.6 and then downsampled by a total scaling factor d in both horizontal and vertical directions. Therefore, for test images, the simulated LR images in our experiments are different from those obtained by the bicubic interpolation in the experimental reports [11], [46]. For the color images, since the human visual system presents more sensitivity to the luminance changes, the image SR methods are only applied to the luminance component, whereas the chromatic components are zoomed in by the simple bicubic interpolation method. In our SMSR model, we have also researched on the patch size as a function of the upscaling factor. To assess the impact of the patch size in the first stage, we have evaluated our SMSR1 with various size of image patch in the experiments. For the test *'Butterfly'* image, the PSNR(dB)/SSIM results are shown in Fig. 5 (a) and (b), where our SMSR1 achieves the best performance with the patch size of 5×5 for the total upscaling factor 2 and with the patch size of 7×7 for the total upscaling factor 3, respectively. Considering the balance between the computational complexity and the performance, the patch size is selected as 5×5 pixels in the first stage. Furthermore, we have assessed various size of image patch in the second stage of our SMSR algorithm. For the test *'Butterfly'* image, the PSNR(dB)/SSIM results are shown in Fig. 5 (c) and (d), where our SMSR2 achieves the best performance with the patch size of 5×5 pixels for both the total upscaling factors 2 and 3. Considering that the patch size is 6×6 pixels in the NCSR method [39], the patch size is also chosen as 6×6 pixels in the second stage of our SMSR algorithm for a fair comparison in the experiments.

To assess the impact of the first stage in our SMSR framework, we have tested the different initialization methods. These initialization methods include the nearest interpolation, the bilinear interpolation, the bicubic interpolation, our proposed SMSR1 and the oracle interpolation method. The compared results of one example for objective and

TABLE II

PSNR(dB)/SSIM RESULTS ON THE RECONSTRUCTED HR IMAGES OBTAINED BY OUR SMSR FRAMEWORK WITH DIFFERENT INTERPOLATIONS APPLIED TO THE *'Butterfly'* IMAGE FOR DIFFERENT SCALING RATIOS

Ratios	Oracle	SMSR1	Bicubic	Bilinear	Nearest
×2	**31.62/0.9367**	30.96/0.9464	30.57/0.9295	30.54/0.9295	28.04/0.9084
×3	**29.39/0.9300**	28.48/0.9251	28.29/0.9194	28.26/0.9188	26.91/0.8950

Fig. 6. Visual comparisons of the super-resolution results of our SMSR framework with the different interpolations for the *'Butterfly'* image (the total scaling factor 3). From left to right: (a) Ground truth, (b) the oracle interpolation, (c) SMSR1, (d) the bicubic interpolation, (e) the bilinear interpolation, and (f) the nearest-neighbor interpolation.

TABLE III

PSNR(dB)/SSIM RESULTS ON THE RECONSTRUCTED HR IMAGES WITH THE TOTAL SCALING FACTOR $d = 2$

Images	SMSR2	NCSR [39]	ASDS [41]	Bicubic	ANR [46]	SMSR1	SMSR_ISS	SMSR_GHP
Bike	**27.34/0.8820**	27.17/0.8729	26.95/0.8722	21.90/0.6478	22.19/0.6771	22.27/0.6774	27.29/0.8738	27.26/0.8751
Brain	31.69/0.8931	31.52/0.8970	30.78/0.8764	26.17/0.7289	26.49/0.7480	26.63/0.7513	**31.75/0.8978**	31.60/0.8977
Butterfly	**30.96/0.9464**	30.33/0.9273	29.65/0.9371	22.42/0.7802	22.80/0.8034	23.21/0.8195	30.50/0.9342	30.57/0.9295
Comic	**27.39/0.8844**	27.34/0.8775	27.10/0.8763	22.15/0.6415	22.40/0.6707	22.48/0.6701	27.34/0.8774	27.38/0.8788
Flower	**31.75/0.9075**	31.52/0.8925	31.34/0.9008	26.25/0.7421	26.41/0.7642	26.67/0.7672	31.56/0.8990	31.58/0.8933
Hat	**33.39/0.9049**	32.74/0.8755	32.94/0.9061	28.30/0.8081	28.32/0.8215	28.85/0.8270	33.02/0.8935	32.81/0.8763
Leaves	**31.56/0.9639**	30.85/0.9511	30.47/0.9552	21.62/0.7378	21.87/0.7674	22.38/0.7903	31.04/0.9569	31.05/0.9528
Parrot	**32.86/0.9335**	32.37/0.9045	32.25/0.9322	26.87/0.8588	26.90/0.8697	27.42/0.8713	32.56/0.9188	32.45/0.9048
Parthenon	**28.93/0.8217**	28.73/0.8079	28.49/0.8058	25.26/0.6689	25.45/0.6868	25.57/0.6885	28.82/0.8179	28.77/0.8088
Plants	**36.23/0.9409**	35.35/0.9172	35.68/0.9383	29.57/0.8353	29.66/0.8503	30.04/0.8529	35.54/0.9243	35.41/0.9176
Raccoon	**31.24/0.8552**	31.12/0.8537	30.96/0.8450	27.50/0.6826	27.59/0.7019	27.78/0.7018	31.06/0.8509	31.15/0.8538
SunFlower	**30.53/0.8997**	30.40/0.8931	30.31/0.8954	24.85/0.7233	24.96/0.7456	25.20/0.7455	30.29/0.8890	30.44/0.8938
Average	**31.16/0.9028**	30.79/0.8892	30.58/0.8951	25.24/0.7379	25.42/0.7589	25.71/0.7636	30.90/0.8944	30.87/0.8902

TABLE IV

PSNR(dB)/SSIM RESULTS ON THE RECONSTRUCTED HR IMAGES WITH THE TOTAL SCALING FACTOR $d = 3$

Images	SMSR2	NCSR [39]	ASDS [41]	Bicubic	ANR [46]	SMSR1	SMSR_ISS	SMSR_GHP
Bike	**24.95/0.8107**	24.72/0.8026	24.60/0.7959	20.80/0.5759	21.11/0.6170	21.05/0.6148	24.77/0.8034	24.79/0.8054
Brain	**29.56/0.8434**	29.36/0.8407	28.81/0.8251	24.75/0.6660	24.95/0.6892	24.84/0.6833	29.42/0.8415	29.43/0.8420
Butterfly	**28.48/0.9251**	28.09/0.9160	27.30/0.9049	20.78/0.7175	21.19/0.7567	21.39/0.7723	28.26/0.9192	28.29/0.9193
Comic	**24.72/0.7950**	24.65/0.7909	24.48/0.7819	20.87/0.5573	21.03/0.5930	21.04/0.5879	24.67/0.7912	24.68/0.7926
Flower	**29.68/0.8618**	29.51/0.8567	29.17/0.8463	24.83/0.6753	24.98/0.7051	24.82/0.7021	29.59/0.8594	29.56/0.8578
Hat	**31.60/0.8768**	31.26/0.8701	30.99/0.8714	27.20/0.7778	27.27/0.7951	27.51/0.7958	31.40/0.8738	31.37/0.8719
Leaves	**27.77/0.9302**	27.46/0.9217	26.94/0.9101	19.83/0.6411	20.05/0.6841	20.06/0.7039	27.69/0.9271	27.62/0.9252
Parrot	**30.76/0.9186**	30.53/0.9148	30.09/0.9101	25.58/0.8261	25.84/0.8414	26.19/0.8422	30.60/0.9170	30.56/0.9156
Parthenon	**27.24/0.7550**	27.18/0.7508	26.89/0.7369	24.12/0.6205	24.25/0.6392	24.18/0.6320	27.21/0.7528	27.21/0.7521
Plants	**34.23/0.9211**	34.03/0.9189	33.41/0.9072	27.83/0.7875	27.91/0.8061	27.92/0.8034	34.09/0.9202	34.15/0.9201
Raccoon	**29.33/0.7717**	29.28/0.7711	29.23/0.7655	26.38/0.6280	26.24/0.6453	26.49/0.6452	29.28/0.7711	29.29/0.7710
SunFlower	**28.05/0.8405**	27.93/0.8377	27.89/0.8331	23.22/0.6466	23.25/0.6711	23.42/0.6713	27.96/0.8380	27.98/0.8389
Average	**28.86/0.8542**	28.67/0.8494	28.32/0.8407	23.85/0.6766	24.00/0.7036	24.08/0.7045	28.74/0.8512	28.74/0.8510

subjective evaluations are shown in Table II and Fig. 6, respectively. As can be seen from the results, our SMSR1 for the initial estimation of the target HR image is better than current competing interpolation methods and is very close to the performance of the oracle interpolation.

The proposed SMSR algorithm and the current competing methods were separately applied to a set of test images from standard image databases. To evaluate the objective quality of the restored HR images, the PSNR and SSIM [62] were calculated for the comparisons of our SMSR algorithm and the state-of-the-art methods. For a fair comparison, we have implemented the qualitative and quantitative evaluation on the reconstructed HR images obtained by our proposed SMSR algorithm, the bicubic interpolation method, ANR [46], ASDS [41] and NCSR [39]. Considering the inter-scale similarity, the gradual magnification scheme instead of the interpolation method is used for the initial estimation of the target HR image in our SMSR model. For convenience, let SMSR_ISS denote the united framework of NCSR [39] with the initialization by our multi-step gradual magnification scheme, and let SMSR_GHP denote the GHP regularization combined with NCSR [39]. Both SMSR_ISS and SMSR_GHP

Fig. 7. Visual comparisons of the super-resolution results of the proposed SMSR algorithm and other state-of-the-art methods for test images (the total scaling factor 2). These test images include *'Brain'*, *'Butterfly'*, *'Comic'* and *'Hat'*. From left to right and top to bottom: (a) LR input, (b) ground truth, (c) our SMSR2, (d) NCSR [39], (e) ASDS [41], (f) the bicubic interpolation, (g) ANR [46], (h) our SMSR1, (i) our SMSR_ISS and (j) our SMSR_GHP. Zoom into the pdf file for a detailed view.

have been separately implemented and compared with the current competing methods to evaluate the impact of the first stage and the GHP regularization of our SMSR model. Note that the bicubic interpolation is used for the initial estimation of the target HR image in the SMSR_GHP model, whereas SMSR1 is used for the initial estimation of the target

Fig. 8. Visual comparisons of the super-resolution results of the proposed SMSR algorithm and other state-of-the-art methods for test images (the total scaling factor 3). These test images include *'Brain'*, *'Butterfly'*, *'Comic'* and *'Hat'*. From left to right and top to bottom: (a) LR input, (b) ground truth, (c) our SMSR2, (d) NCSR [39], (e) ASDS [41], (f) the bicubic interpolation, (g) ANR [46], (h) our SMSR1, (i) our SMSR_ISS and (j) our SMSR_GHP. Zoom into the pdf file for a detailed view.

HR image in our SMSR model. For the set of test images, the PSNR/SSIM results of these different methods are shown in Table III for the total scaling factor $d = 2$ and Table IV for the total scaling factor $d = 3$, respectively. Both Table I and Fig. 1 show that the bicubic interpolation with the NCSR method is usually superior to the

nearest-neighbor interpolation or the bilinear interpolation with the NCSR method. Thus only the bicubic interpolation with the NCSR method is compared with our SMSR algorithm for performance evaluation. As can be seen from Table III to Table IV, it is found that the average PSNR gains of our SMSR algorithm over the second best method, *i.e.*, NCSR [39], are about 0.4dB and 0.2dB for these test LR images, respectively. Moreover, for a comparison of the initialization of NCSR [39] and our SMSR algorithm, the average PSNR gains of our SMSR1 algorithm over the bicubic interpolation method are separately about 0.5dB and 0.2dB in the cases of $d = 2$ and $d = 3$. From the PSNR/SSIM results, our proposed SMSR algorithm is generally superior to the current state-of-the-art methods. To further inspect the effectiveness of our SMSR algorithm, the detailed reconstructed HR results of our SMSR algorithm and other competing methods [39], [41] for the test images are shown in Fig. 7 for the case $d = 2$ and Fig. 8 for the case $d = 3$, respectively. The visual comparisons between our SMSR algorithm and the baseline methods [39], [41] demonstrate that the proposed SMSR algorithm can recover finer structures and sharper edges and also has less color distortion for color image super-resolution.

It is also found that our proposed SMSR algorithm has achieved noticeable performance gains in great part due to the knowledge of the blur kernel. Unlike Timofte *et al.* [46] and followers assuming sharp input LR images, the proposed SMSR algorithm assumes that the blur kernel is known beforehand. As the initialization of our SMSR model, the example-based SMSR1 can achieve better results, especially for sharp input LR images, whereas our SMSR2 reaches its best performance with accurate blur kernel for image super-resolution, which is especially useful when the input image is severely blurred. To the best of our knowledge, our proposed SMSR algorithm obviously outperforms the best existing state-of-the-art methods, *e.g.*, NCSR [39], not only in the objective assessment but also in the visual comparisons. As seen from the experimental results, the proposed SMSR algorithm works well for a wide variety of images, and can reach better super-resolution results than the state-of-the-art methods both subjectively and objectively.

V. Conclusions and Future Work

In this paper, we proposed a solution to the single image super-resolution problem with an SMSR method. Since there is abundant similarity redundancy both within the same scale and across the different scales, the multi-scale magnification scheme with the ridge regression is first used to compute the initial estimation of the target HR image. Then the sparse modeling of single image super-resolution is designed with a gradient regularization term that preserves the gradient histogram of the target HR image. Another centralized sparse constraint that exploits the image local and nonlocal redundancy is also incorporated to improve the performance of the image sparse representation. To approximate the global optimization result, which is nonconvex and hard to solve directly, an alternating minimization method with an iteratively reweighted regularization parameter is used to solve the structure-constrained optimization problem of single image super-resolution. The sparse coefficients of the estimated HR image are further corrected by an efficient iterative shrinkage function. We have conducted extensive experiments on image super-resolution and evaluated the results of both the proposed algorithm and the popular SR methods. Experimental results demonstrated that our SMSR algorithm that can produce sharper edges and suppress aliasing artifacts is promising and competitive to the state-of-the-art methods, and outperforms other leading SR methods both visually and quantitatively in most cases. To further improve the performance and the efficiency of super-resolution, we will study on accurate estimation of blur kernel and noisy image super-resolution in the future.

References

[1] P. Milanfar, Ed., *Super-resolution Imaging* (Digital Imaging and Computer Vision). New York, NY, USA: CRC Press, Sep. 2010.

[2] R. Y. Tsai and T. S. Huang, "Multiple frame image restoration and registration," in *Advances in Computer Vision and Image Processing*, vol. 1. Greenwich, CT, USA: JAI Press, 1984.

[3] W.-C. Siu and K.-W. Hung, "Review of image interpolation and super-resolution," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, Dec. 2012, pp. 1–10.

[4] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP, Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, 1991.

[5] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1621–1633, Dec. 1997.

[6] S. Farsiu, M. Elad, and P. Milanfar, "Video-to-video dynamic super-resolution for grayscale and color sequences," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–15, Feb. 2006, Art. ID 061859.

[7] A. Marquina and S. J. Osher, "Image super-resolution by TV-regularization and Bregman iteration," *J. Sci. Comput.*, vol. 37, no. 3, pp. 367–382, 2008.

[8] X. Zhang, J. Jiang, and S. Peng, "Commutability of blur and affine warping in super-resolution with application to joint estimation of triple-coupled variables," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1796–1808, Apr. 2012.

[9] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar./Apr. 2002.

[10] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 349–356.

[11] 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.

[12] 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.

[13] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Gradient profile prior and its applications in image super-resolution and enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1529–1542, Jun. 2011.

[14] X. Gao, K. Zhang, D. Tao, and X. Li, "Joint learning for single-image super-resolution via a coupled constraint," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 469–480, Feb. 2012.

[15] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 6, pp. 1013–1027, Jun. 1990.

[16] H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," *CVGIP, Graph. Models Image Process.*, vol. 54, no. 2, pp. 181–186, 1992.

[17] M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda, "Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 5, pp. 915–923, Oct. 2000.

[18] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 349–366, Feb. 2007.

[19] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.

[20] D. Capel and A. Zisserman, "Computer vision applied to super resolution," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 75–86, May 2003.

[21] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian methods for image super-resolution," *Comput. J.*, vol. 52, no. 1, pp. 101–113, 2009.

[22] L. J. Karam, N. G. Sadaka, R. Ferzli, and Z. A. Ivanovski, "An efficient selective perceptual-based super-resolution estimator," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3470–3482, Dec. 2011.

[23] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. Opt. Soc. Amer. A-Opt. Image Sci. Vis.*, vol. 6, no. 11, pp. 1715–1726, 1989.

[24] X. Zhang, M. Tang, and R. Tong, "Robust super resolution of compressed video," *Vis. Comput.*, vol. 28, no. 12, pp. 1167–1180, 2012.

[25] S. Izadpanahi and H. Demirel, "Motion based video super resolution using edge directed interpolation and complex wavelet transform," *Signal Process.*, vol. 93, no. 7, pp. 2076–2086, 2013.

[26] J. Salvador, A. Kochale, and S. Schweidler, "Patch-based spatio-temporal super-resolution for video with non-rigid motion," *Signal Process., Image Commun.*, vol. 28, no. 5, pp. 483–493, 2013.

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

[28] Z. Wei and K.-K. Ma, "Contrast-guided image interpolation," *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4271–4285, Nov. 2013.

[29] K. Jia, X. Wang, and X. Tang, "Image transformation based on learning dictionaries across image spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 367–380, Feb. 2013.

[30] 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.

[31] 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.

[32] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[33] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, 2006.

[34] W. Zuo, L. Zhang, C. Song, and D. Zhang, "Texture enhanced image denoising via gradient histogram preservation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1203–1210.

[35] Y. Zhang, J. Liu, M. Li, and Z. Guo, "Joint image denoising using adaptive principal component analysis and self-similarity," *Inf. Sci.*, vol. 259, pp. 128–141, Feb. 2014.

[36] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[37] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Comput. Vis.*, vol. 86, no. 1, pp. 1–32, 2010.

[38] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2272–2279.

[39] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[40] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[41] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

[42] Y.-Q. Zhang, Y. Ding, J.-S. Xiao, J. Liu, and Z. Guo, "Visibility enhancement using an image filtering approach," *EURASIP J. Adv. Signal Process.*, vol. 2012, pp. 1–6, Oct. 2012, Art. ID 220.

[43] Y.-Q. Zhang, Y. Ding, J. Liu, and Z. Guo, "Guided image filtering using signal subspace projection," *IET Image Process.*, vol. 7, no. 3, pp. 270–279, Apr. 2013.

[44] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 184–199.

[45] T. Michaeli and M. Irani, "Nonparametric blind super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 945–952.

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

[47] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vis.*, Nov. 2014, pp. 111–126.

[48] E. Perez-Pellitero, J. Salvador, I. Torres-Xirau, J. Ruiz-Hidalgo, and B. Rosenhahn, "Fast super-resolution via dense local training and inverse regressor search," in *Proc. Asian Conf. Comput. Vis.*, Nov. 2014, pp. 346–359.

[49] H. A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1647–1659, Oct. 2005.

[50] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[51] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[52] G. Peyre, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," *Inverse Problems Imag.*, vol. 5, no. 2, pp. 511–530, 2011.

[53] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[54] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *SIAM Multiscale Model. Simul.*, vol. 7, no. 1, pp. 214–241, 2008.

[55] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, Mar. 2010.

[56] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.

[57] Y. Zhang, J. Xiao, S. Li, C. Shi, and G. Xie, "Learning block-structured incoherent dictionaries for sparse representation," *Sci. China Inf. Sci.*, vol. 58, pp. 102302-1–102302-15, 2015.

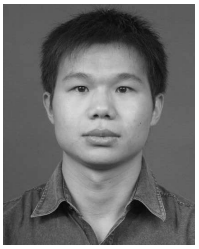
[58] T. S. Cho, C. L. Zitnick, N. Joshi, S. B. Kang, R. Szeliski, and W. T. Freeman, "Image restoration by matching gradient distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 683–694, Apr. 2012.

[59] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 877–905, 2008.

[60] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, 2010.

[61] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods," *Math. Program.*, vol. 137, nos. 1–2, pp. 91–129, 2013.

[62] 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.

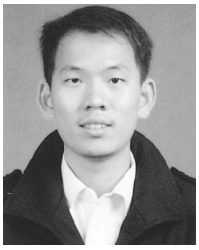


Yongqin Zhang (M'13) received the B.S.E. degree in electronics science and technology from Zhengzhou University, China, and the Ph.D. degree in communication and information systems from Wuhan University, China, in 2005 and 2010, respectively.

He was an Assistant Research Fellow with the Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China, from 2011 to 2012. From 2012 to 2014, he was a Post-Doctoral Researcher with the Institute of Computer Science and Technology, Peking University, China. He is currently a Lecturer with the School of Information Science and Technology, Northwest University, China. His research interests include sparse representation, image restoration, image enhancement, magnetic resonance imaging, and computational imaging.

Jiaying Liu (S'09–M'10) received the B.E. degree in computer science from Northwestern Polytechnic University, Xi'an, China, and the Ph.D. (Hons.) degree in computer science from Peking University, Beijing, China, in 2005 and 2010, respectively.

She was a Visiting Scholar with the University of Southern California, Los Angeles, from 2007 to 2008. She has visited Microsoft Research Asia since 2015 and was supported by Star Track for Young Faculties. She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. She has authored or co-authored over 60 papers and eight granted patents. Her current research interests include image processing, computer vision, and video compression. She has also served as a Technical Committee Member at the Asia-Pacific Signal and Information Processing Association Image, Video, and Multimedia since 2015.



Wenhan Yang received the B.S. degree in computer science from Peking University, Beijing, China, in 2008.

He is currently pursuing the Ph.D. degree with the Institute of Computer Science and Technology, Peking University. His current research interests include image processing, sparse representation, and image restoration.

Zongming Guo (M'09) received the B.S. degree in mathematics, and the M.S. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990 and 1994, respectively.

He is currently a Professor with the Institute of Computer Science and Technology, Peking University. His current research interests include video coding, processing, and communication. He is the Executive Member of the China Society of Motion Picture and Television Engineers. He was a recipient of the First Prize of the State Administration of Radio Film and Television Award in 2004, the First Prize of the Ministry of Education Science and Technology Progress Award in 2006, the Second Prize of the National Science and Technology Award in 2007, the Wang Xuan News Technology Award and the Chia Tai Teaching Award in 2008, and the Government Allowance granted by the State Council in 2009.