MR image super-resolution reconstruction using sparse representation, nonlocal similarity and sparse derivative prior

Di Zhang [a,c,*], Jiazhong He [b], Yun Zhao [a], Minghui Du [c]

[a] School of Information Engineering, Guangdong Medical College, Dongguan, China
[b] Department of Physics, Shaoguan University, Shaoguan, China
[c] School of Electronics and Information, South China University of Technology, Guangzhou, China

ARTICLE INFO

ABSTRACT

In magnetic resonance (MR) imaging, image spatial resolution is determined by various instrumental limitations and physical considerations. This paper presents a new algorithm for producing a high-resolution version of a low-resolution MR image. The proposed method consists of two consecutive steps: (1) reconstructs a high-resolution MR image from a given low-resolution observation via solving a joint sparse representation and nonlocal similarity $L1$-norm minimization problem; and (2) applies a sparse derivative prior based post-processing to suppress blurring effects. Extensive experiments on simulated brain MR images and two real clinical MR image datasets validate that the proposed method achieves much better results than many state-of-the-art algorithms in terms of both quantitative measures and visual perception.

1. Introduction

Compared with other medical imaging techniques, MR imaging uses non-ionizing radiation and provides distinct microscopic chemical and physical information of molecules. However, the spatial resolution of MR images is limited by various instrumental limitations (e.g., gradients' intensity, filter bandwidth) and physical considerations. The resolution limitation could result in partial volume effect (PVE), a phenomenon that each pixel in the MR images could contain more than one material or tissue type. To reduce PVEs, a common practice is to magnify the images using standard interpolation techniques. However, interpolation techniques usually do not take into account the fact that a low-resolution (LR) pixel is actually a weighted average of the high-resolution (HR) pixels inside it, thus the magnified HR images are typically featured with blurred edges and tissues.

To overcome this problem, various methods have been proposed [1,5–8,11,12,17–20] so far. Among them, super-resolution (SR) is one of the most promising methods and receives much attention in the research community. SR image reconstruction is the process of recovering a HR image from a single (e.g., [2]) or a set of LR images (e.g., [3]). The essential difference between single-frame and multi-frame SR image reconstruction is that new high-frequency information could also be recovered from different LR frames [4]. In MR image analysis, SR was first used to reconstruct a HR image by merging multiple LR acquisitions with subpixel displacements. In [5], Herment et al. reduced total data acquisition time by merging multiple k-space data. Shilling et al. [6] improved the resolution and contrast of MR images by fusing multiple 2D slices with different slice directions. Greenspan et al. [7] used iterative back-projection and Islam et al. [8] used a wavelet-based deblurring approach to improve the resolution of 3D MR images.

Though multi-frame SR image reconstruction is theoretically more promising than single-frame SR image reconstruction, it suffers many difficulties in real applications, such as subpixel image registration/acquisition, the increase of computational complexity as frame number increases. On the other hand, many researches [9,10] have demonstrated that, given a proper prior image model, single-frame SR image reconstruction can also be as effective as multi-frame SR image reconstruction. To reconstruct a HR image from a single MR image, Rousseau [11] used the idea presented in [9] and proposed a patch-based nonlocal regularization framework for brain MR image reconstruction. Manjón et al. [12] extended the patch-based nonlocal regularization framework to include a coherence constraint.

Recently, a powerful statistical image modeling technique, sparse representation [13,14], has been successfully applied in natural image SR applications [15,16,22]. In MR image analysis,

* Corresponding author at: Guangdong Medical College, School of Information Engineering, Song Shan Hu, Dongguan, China. Tel./fax: +86 20 3876 9529.
E-mail address: haihaiwenqi@163.com (D. Zhang).

however, many efforts have been focused on applying compressed sensing (CS) [17–20], a technique through which a perfect MR image reconstruction is possible by only a small subset of k-space samples that is far less than the Nyquist sampling theorem. Since these CS-based SR methods [17–20] focus on manipulating k-space samples, they do have many advantages, e.g., theoretical simplicity and low computational cost. Nevertheless, they also present some important drawbacks. For example, recovering high frequency information in the k-space will inevitably cause visual artifacts in the image space, thus they need to work in the frequency and the image space in turn to suppress artifacts [20,21]. On the other hand, sparse representation based techniques manipulate image patches in the image space, thus provide much more interpretable information to human eyes, and more importantly, facilitate experts to adopt a much more flexible observation model (e.g., local motion) and incorporate various image priors. Based on these considerations, Andrea et al. [21] proposed to reconstruct HR 3D brain MR image from LR 3D image volumes using overcomplete dictionaries.

Motivated by the ideas presented in Refs. [11,12,21,36,37], in this paper, we propose a new algorithm for reconstructing a HR MR image from a single LR image. The proposed method consists of two consecutive steps: (1) reconstructs a HR MR image from a given LR observation via solving a joint sparse representation and nonlocal similarity $L1$-norm minimization problem; and (2) applies a sparse derivative prior based post-processing on the reconstructed HR image to suppress blurring effects.

The rest of the paper is organized as follows. In Section 2, we give a brief review of SR image reconstruction based on sparse representation. Sections 3 and 4 present the nonlocal similarity and sparse derivative prior image model, respectively. Section 5 presents the new algorithm. Extensive experiments on simulated brain MR images and two real clinical MR image datasets are conducted in Section 6 to verify the efficiency of our method. Finally, we provide discussion in Section 7.

2. SR image reconstruction based on sparse representation

In SR image reconstruction, the LR image can be modeled as a down-sampled version of the HR image which has been blurred, i.e.,

$$\mathbf{Y} = \mathbf{WZ} \tag{1}$$

where $\mathbf{Y}$ is the observed LR image, $\mathbf{Z}$ is the original HR image, and $\mathbf{W}$ is a degradation operator representing the blur and down-sampling operator which operates on $\mathbf{Z}$ to yield $\mathbf{Y}$ (geometric shift is not included in $\mathbf{W}$ since we focus on the single-frame SR reconstruction). A *maximum a posterior* (MAP) estimate of the unknown HR image $\mathbf{Z}$ can be computed as

$$\hat{\mathbf{Z}} = \arg \max_{\mathbf{Z}} \{ \log \Pr(\mathbf{Z}|\mathbf{Y}) \}$$
$$= \arg \min_{\mathbf{Z}} \{ \|\mathbf{Y} - \mathbf{WZ}\|_2^2 - \log \Pr(\mathbf{Z}) \} \tag{2}$$

where $\Pr(\mathbf{Z})$ is the prior image model. Many works have been contributed to find a good prior image model, and total variation (TV) [23–26] is one of the most commonly used models. Sparse representation has also been successfully applied as a prior image model as well. Given an image $\mathbf{Z}$, the sparse representation assumes that there exists a sparse vector $\mathbf{\Lambda}$ and a proper learned dictionary $\mathbf{\Psi}$ (each column in $\mathbf{\Psi}$ is referred to as an atom), such that

$$\mathbf{Z} \approx \mathbf{\Psi\Lambda}, \quad \text{s.t.} \ \|\mathbf{\Lambda}\|_0 \leq \varepsilon \tag{3}$$

where ε is a predefined threshold to control the sparsity of $\mathbf{\Lambda}$ and $L0$-norm $\| \cdot \|_0$ counts the number of nonzero elements in a vector.

Let $\mathbf{\Psi}_h$ and $\mathbf{\Psi}_l$ are the coupled two dictionaries for the HR and LR images, respectively. For single-frame SR reconstruction, given a HR image $\mathbf{Z}$ and the corresponding LR image $\mathbf{Y}$, there exists a sparse vector $\mathbf{\Lambda}$ simultaneously satisfies [15]

$$\mathbf{Z} \approx \mathbf{\Psi}_h \mathbf{\Lambda} \ \text{and} \ \mathbf{Y} \approx \mathbf{\Psi}_l \mathbf{\Lambda}, \ \text{s.t.} \ \|\mathbf{\Lambda}\|_0 \leq \varepsilon, \tag{4}$$

With the sparsity prior image model defined in (4), finding the solution to (1) is equivalent to finding the representation of $\mathbf{Z}$ over $\mathbf{\Psi}_h$, which can be estimated from its LR observation $\mathbf{Y}$ by solving the following $L0$-norm minimization problem:

$$\hat{\mathbf{\Lambda}} = \arg \min_{\mathbf{\Lambda}} \{ \|\mathbf{Y} - \mathbf{\Psi}_l \mathbf{\Lambda}\|_2^2 + \lambda \|\mathbf{\Lambda}\|_0 \} \tag{5}$$

where λ is a parameter controlling the importance of sparsity prior. Since $L0$-norm is nonconvex and solving (5) is NP-hard, many recent works [13,27] demonstrated that if the coefficients $\mathbf{\Lambda}$ is *sparse enough*, the solution to (5) can be efficiently approximated by solving the following $L1$-norm minimization problem

$$\hat{\mathbf{\Lambda}} = \arg \min_{\mathbf{\Lambda}} \{ \|\mathbf{Y} - \mathbf{\Psi}_l \mathbf{\Lambda}\|_2^2 + \lambda \|\mathbf{\Lambda}\|_1 \} \tag{6}$$

where $L1$-norm $\| \cdot \|_1$ calculates the sum of the absolute of each element in a vector. Notice that (6) is also known as the Lasso in statistical literature [28].

Once $\hat{\mathbf{\Lambda}}$ is obtained, $\mathbf{Z}$ can then be estimated as

$$\hat{\mathbf{Z}} = \mathbf{\Psi}_h \hat{\mathbf{\Lambda}} \tag{7}$$

Traditionally, we divide an image into overlapped or non-overlapped patches, apply (6) and (7) to each image patch and fuse all the reconstructed HR patches to get the final HR image. To avoid confusion, we use letters in lowercase or uppercase to represent an image patch or an entire image throughout this paper unless otherwise stated. For the convenience of later discussion, we write the patch-based version of (6) and (7) as follows:

$$\hat{\mathbf{\alpha}} = \arg \min_{\mathbf{\alpha}} \{ \|\mathbf{y} - \mathbf{\Psi}_l \mathbf{\alpha}\|_2^2 + \lambda \|\mathbf{\alpha}\|_1 \} \tag{8}$$

$$\hat{\mathbf{z}} = \mathbf{\Psi}_h \hat{\mathbf{\alpha}} \tag{9}$$

3. Nonlocal similarity

A critical issue in sparse representation prior is the choice of dictionaries. In general, the more redundant the dictionary is, the better SR result will be. Unfortunately, the computational complexity will also increase as the size of dictionary increases. Many dictionary learning algorithms thus aim at getting a compact overcomplete dictionary to represent various image patches. Nevertheless, due to the diversity of natural image patterns, it is impossible for such compact dictionary to cover all the patterns. As a result, the similar patterns can well be reconstructed while the dissimilar ones cannot. Considering the fact that either similar or dissimilar patterns, there are often many repetitive patterns throughout an image, such nonlocal redundancy is very helpful in preserving edge sharpness and suppressing noise in the reconstructed images [12,14,29,30]. As a supplementary to the sparse representation prior, in this section, we will develop a nonlocal similarity regularization term for SR image reconstruction.

For a given image patch $\mathbf{z}_j$, we search for the similar patches within a sufficiently large area around $\mathbf{z}_j$. Two similarity criteria have been frequently used so far: (1) a patch $\mathbf{z}_j^s$ is selected as a similar patch to $\mathbf{z}_j$ if $d_j^s = ||\mathbf{z}_j^s - \mathbf{z}_j||_2^2 \leq t$, where t is a preset threshold, or (2) one can select the patch if it is within the first L (e.g., $L=15$) closest patches to $\mathbf{z}_j$. However, they are all based on Euclidean norm in the vector space and do not truly reflect the

similarity between two vectors. In this paper, we use the cosine of the angle between two vectors as the distance measurement,

$$d_j^s = \frac{\langle \mathbf{z}_j^s, \mathbf{z}_j \rangle}{(||\mathbf{z}_j^s||_2 \cdot ||\mathbf{z}_j||_2)} \tag{10}$$

Suppose L similar patches ($\mathbf{z}_j^s$, $s = 1, 2, \ldots, L$) have been located for $\mathbf{z}_j$, let z_j^s be the central pixel of $\mathbf{z}_j^s$ and z_j be the central pixel of $\mathbf{z}_j$. We can use the weighted average of z_j^s to predict z_j,

$$\hat{z}_j = \sum_{s=1}^L z_j^s c_j^s \tag{11}$$

where c_j^s is the weight assigned to z_j^s, determined by

$$c_j^s = \frac{\exp(-d_j^s/h)}{\sum_{s=1}^L \exp(-d_j^s/h)} \tag{12}$$

where d_j^s is determined by (10) and h is a controlling factor of the weight. Considering the fact that there are plenty of repetitive patterns throughout a MR image, the mean squared error between the prediction and the ground truth, i.e.,

$$||z_j - \hat{z}_j||_2^2 = ||z_j - \sum_{s=1}^L z_j^s c_j^s||_2^2 \tag{13}$$

should be sufficiently small. Let $\mathbf{c}_j$ be the column vector containing all the weights c_j^s and patch $\mathbf{p}_j$ be the column vector containing all z_j^s. By summing the mean squared prediction error across the whole image patch $\mathbf{z}_j$, we get

$$\sum_{z_j \in \mathbf{z}_j} ||z_j - \sum_{s=1}^L z_j^s c_j^s||_2^2 = \sum_{z_j \in \mathbf{z}_j} ||z_j - \mathbf{c}_j^T \mathbf{p}_j||_2^2 \tag{14}$$

Since (8) calculates the sparse representation coefficients of HR image patch $\mathbf{z}$ only using fidelity constraint and sparsity prior, to incorporate the nonlocal similarity regularization, we revise (8) as follows:

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \left\{ ||\mathbf{y} - \boldsymbol{\Psi}_l \boldsymbol{\alpha}||_2^2 + \lambda ||\boldsymbol{\alpha}||_1 + \eta \sum_{z_j \in \hat{\mathbf{z}}_j} ||z_j - \mathbf{c}_j^T \mathbf{p}_j||_2^2 \right\} \tag{15}$$

where η is a parameter controlling the contribution of nonlocal similarity regularization. If we define a matrix $\mathbf{C}$ as

$$C_{js} = \begin{cases} c_j^s, \text{if} z_j^s \in \mathbf{p}_j, c_j^s \in \mathbf{c}_j \\ 0, \text{otherwise} \end{cases} \tag{16}$$

By substituting (16) and $\mathbf{z} = \boldsymbol{\Psi}_l \boldsymbol{\alpha}$ into (15), we get

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \left\{ ||\mathbf{y} - \boldsymbol{\Psi}_l \boldsymbol{\alpha}||_2^2 + \lambda ||\boldsymbol{\alpha}||_1 + \eta ||(\mathbf{I} - \mathbf{C})\boldsymbol{\Psi}_l \boldsymbol{\alpha}||_2^2 \right\} \tag{17}$$

where $\mathbf{I}$ is an identity matrix. By introducing

$$\mathbf{y}^* = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} \text{ and } \boldsymbol{\Gamma} = \begin{bmatrix} \mathbf{I} \\ \eta(\mathbf{I} - \mathbf{C}) \end{bmatrix} \tag{18}$$

Eq. (17) can further be simplified as

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \left\{ ||\mathbf{y}^* - \boldsymbol{\Gamma} \boldsymbol{\Psi}_l \boldsymbol{\alpha}||_2^2 + \lambda ||\boldsymbol{\alpha}||_1 \right\} \tag{19}$$

Please note that (19) has exactly the same form of (8), which is also a $L1$-norm minimization problem.

4. Sparse derivative prior image model

Compared with the TV model, using sparsity prior or nonlocal similarity can improve the quality of the final HR image. However, we

will demonstrate in Section 6 that sparsity prior or nonlocal similarity based methods also produce blurry effects along strong edges or in soft tissue areas in the reconstructed MR images. In general, the blurry effect is caused by the average operation in these methods. For sparsity prior based methods, overlapped HR patches are separately reconstructed and the pixels in the overlapped regions are averaged via various strategies to maintain compatibility between adjacent patches, and as a result, the selected averaging strategy largely determines the blurry degree. On the other hand, for nonlocal similarity based methods, L similar patches (not identical patches) are averaged to predict the central pixel of the target patch, and certainly the higher the degree of similarity, the less blurry effect in the final HR MR image. Since balancing the blurry effect of sparsity prior and nonlocal similarity under a unified $L1$-norm minimization framework is difficult and complicated, in this paper, we adopt the sparse derivative prior image model to suppress blurry effects.

For lots of natural images, researchers have found that after applying a localized, oriented, and bandpass filtering operation, the histogram of the filtered images are exponential with high fourth-order statistics (kurtosis) [38–40]. For example, Fig. 1(a) shows a natural image, the histogram of the filtered image via a 1-D filter $[-1, 0, 1]$ is shown in Fig. 1(b). We can model the distribution in Fig. 1(b) via

$$Pr(x) \propto \exp \left(-\frac{1}{2} \left(\frac{|x|}{\sigma} \right)^\alpha \right) \tag{20}$$

where σ is the standard deviation, and α is the exponent parameter. Fig. 1(b) reveals the fact that the gradients of an image are largely zeros, i.e., strong derivatives are sparse in a given natural image. Eq. (20) is also known as the natural image prior when $0 < \alpha < 1$, for it can model lots of natural images [38–40]. Unlike the TV model assumes local smoothness, the natural image prior tends to encourage a single strong derivative, which would appear as a sharp edge in local regions. The natural image prior has been successively applied in natural image noise reduction [41], compression [42], deconvolution [43,44], and SR as well [36,37]. In [37], Tappen et al. simultaneously used fidelity constraint and natural image prior to perform SR. Kwang and Younghee [36] first used kernel ridge regression and then applied natural image prior based post-processing to suppress ringing artifacts. Nevertheless, to the authors' best knowledge, by far there are many works reported on natural image processing utilizing natural image prior, little similar job was done on MR images or other medical images. To validate whether the natural image prior can be applied to MR images, we filter a brain MR image (shown in Fig. 2(a)) with the same 1-D filter used in Fig. 1 and show the histogram of the filtered image in Fig. 2(b). By comparing Fig. 1 with Fig. 2, it is clear that though the two images shown in Fig. 1(a) and Fig. 2(a) are quite different in nature, the distributions are almost the same, which implies that the natural image prior can also be used in MR image processing.

In [37], Tappen et al. estimated the HR patch $\mathbf{z}$ by maximizing the following joint posterior probability distribution

$$Pr(\{\mathbf{z}\}|\{y\}) = \frac{1}{C} \prod_{j, i \in N_8(j)} \exp\left[-\left(\frac{||\mathbf{z}_j - \mathbf{z}_i||_1}{\sigma_N} \right)^\alpha \right] \prod_j \exp\left[-\left(\frac{||\mathbf{W}\mathbf{z}_j - y_j||_2}{\sigma_R} \right)^2 \right] \tag{21}$$

where $\{y\}$ denotes the observed variables corresponding to the pixels in LR image $\mathbf{Y}$, $\{z\}$ represents the latent variables corresponding to the patch in HR image $\mathbf{Z}$ which satisfies $y = \mathbf{W}\mathbf{z}$, $N_8(j)$ denotes for 8-connected neighbors of the pixel at location j, C is the normalization constant, σ_N and σ_R are standard deviation parameters. The first product in (21) is the natural image prior term and the second one is the fidelity constraint term. Considering that the data fidelity constraint has already been incorporated in the reconstruction process via the coupled two dictionaries $\boldsymbol{\Psi}_h$
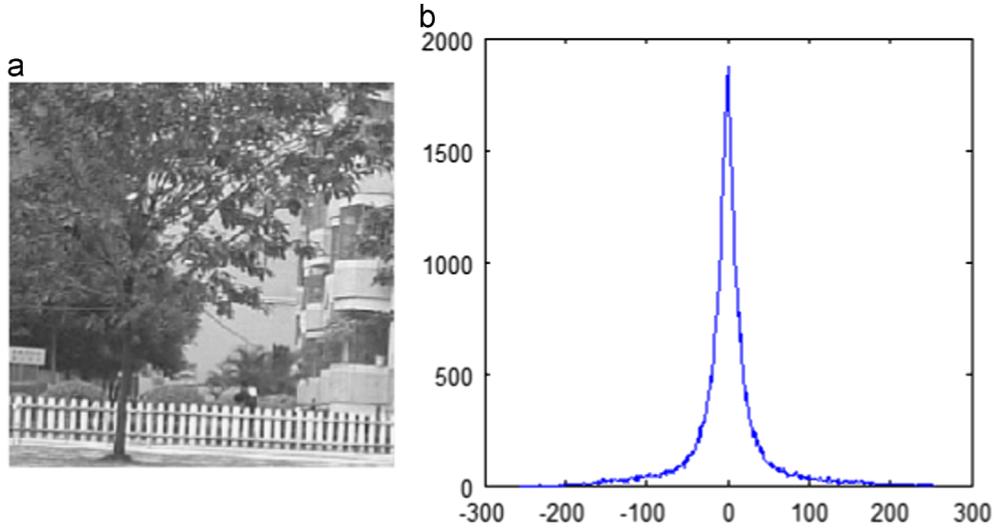
Fig. 1. (a) A natural image, (b) the histogram of the filtered image via 1-D filter $[-1, 0, 1]$.
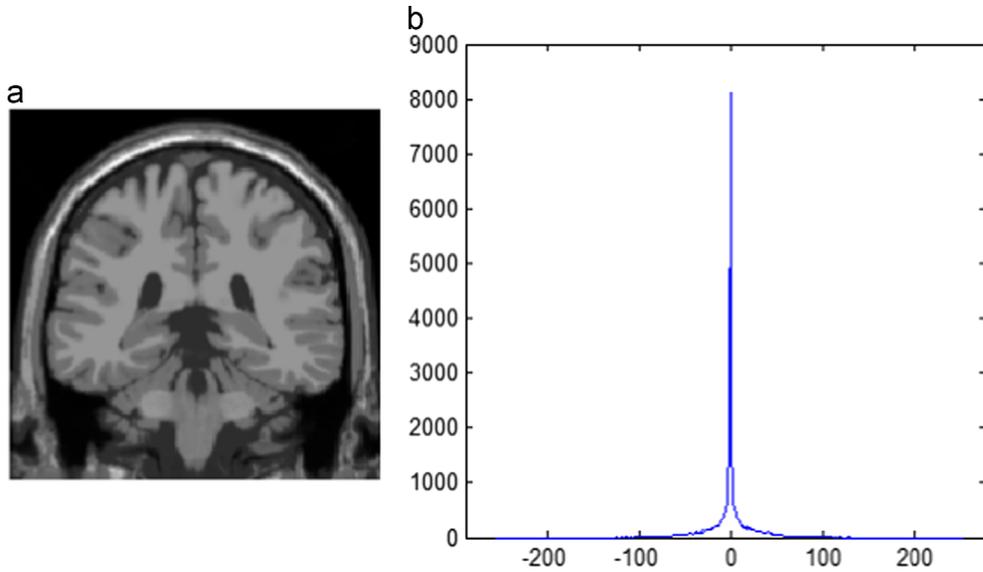
Fig. 2. (a) A brain MR image, (b) the histogram of the filtered image via 1-D filter $[-1, 0, 1]$.

and Ψ_l, directly using (21) in our case is unsuitable. In this paper, a modified version of (21) is thus introduced as follows

$$Pr(\{\mathbf{z}\}\,|\,\{\hat{\mathbf{z}}\}) = \frac{1}{C} \prod_{j,\,i\in N_8(j)} \exp\left[-\left(\frac{\|\mathbf{z}_j-\mathbf{z}_i\|_1}{\sigma_N}\right)^{\alpha}\right] \prod_j \exp\left[-\left(\frac{\|\mathbf{z}_j-\hat{\mathbf{z}}_j\|_2}{\sigma_R}\right)^2\right] \tag{22}$$

where $\{\hat{\mathbf{z}}\}$ denotes the reconstructed HR patches via $\hat{\mathbf{z}} = \Psi_h\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\alpha}}$ is the solution to (19). The motivation for the replacement of the second product term is intuitive: we want that the final result is not deviate far from the output yielded by (19).

To find the solution through which the joint posterior probability distribution (22) is maximized, we first convert (22) into a factor graph (shown in Fig. 3, the complete derivation of the factor graph can be traced back to [37]) and then use a max-sum-type belief propagation (BP) algorithm to compute the exact maximum posterior probability at each variable node. To facilitate the optimization, unlike Tappen et al. did in [37] where 16 learned interpolators are used to generate 16 h candidates for each latent variable, we use $L+1$ patches (the source patch got by (19) plus the corresponding L patches got in the step of searching for nonlocal similar patches) as candidates.
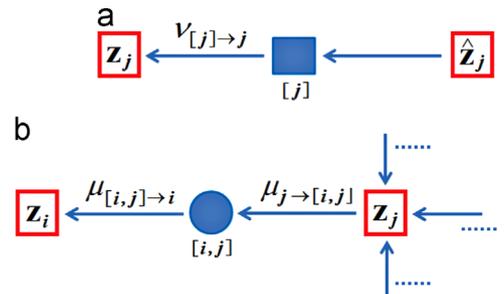
Fig. 3. Factor graph for the optimization of (22). (a) deviation penalty term between latent node j and observed node j, (b) natural image prior based messages propagate from latent node j to node i.

Based on (22), the logarithm of the required message-passing equations for the factor graph can be summarized as

$$\nu_{[j]\to j}(\mathbf{z}_j) = -\frac{1}{2}\left(\|\mathbf{z}_j-\hat{\mathbf{z}}_j\|_2/\sigma_R\right)^2$$

$$\mu_{[i,j]\to i}(\mathbf{z}_i) = \max_{\mathbf{z}_j}\left[\mu_{j\to[i,j]}(\mathbf{z}_j) - \frac{1}{2}\left(\|\mathbf{z}_j-\mathbf{z}_i\|_1/\sigma_N\right)^{\alpha}\right]$$

$$\mu_{j\to[i,j]}(\mathbf{z}_j) = \nu_{[j]\to j}(\mathbf{z}_j) + \sum_{k \in N_8(j)\setminus i} \mu_{[j,k]\to j}(\mathbf{z}_j) \qquad (23)$$

where $\nu_{[j]\to j}$ calculates the message sent from constraint node [j] to variable node $\mathbf{z}_j$, $\mu_{[i,j]\to i}$ represents the message sent from constraint node [i,j] to variable node $\mathbf{z}_i$, and $\mu_{j\to[i,j]}$ is the message sent from variable node j to constraint node [i,j]. In (23), $\mu_{j\to[i,j]}$ and $\mu_{[i,j]\to i}$ are natural image prior based messages propagate from latent node j to node i, while $\nu_{[j]\to j}$ is the deviation penalty term between latent node j and observed node j.

With $L+1$ candidate HR patches in hand, the sparse derivative prior based post-processing is summarized in Algorithm 1.

Algorithm 1.

(1) Run BP algorithm X times using the three message-passing equations defined in (23).
(2) For each variable node $\mathbf{z}_j$, select the candidate HR patch with the highest belief.
(3) Insert the selected candidate HR patches into the corresponding position to form the final output HR image.

5. The proposed method

In this section, we first discuss how to construct the coupled HR and LR dictionaries, and then propose a new SR algorithm to reconstruct a HR image from a single LR MR image.

5.1. Dictionary construction

In this paper, we use a modified version of the method presented in [15,21] to construct the coupled two dictionaries $\mathbf{\Psi}_h$ and $\mathbf{\Psi}_l$ (as illustrated in Fig. 4). Each HR training image in the training set $\{\mathbf{Z}_j, j=1,2,\dots,N\}$ is first blurred and down-sampled by a factor of q to produce the corresponding LR image set $\{\mathbf{Y}_j, j=1,2,\dots,N\}$ (i.e., $\mathbf{Y}_j=\mathbf{WZ}_j$, where $\mathbf{W}$ is the degradation operator defined in (1)). Then a upsampled set $\{\mathbf{Y}_j^U, j=1,2,\dots,N\}$ is obtained by scaling each image in $\{\mathbf{Y}_j\}$ by the same factor q via bicubic interpolation (i.e., $\mathbf{Y}_j^U=\mathbf{U}_q\mathbf{Y}_j$, where $\mathbf{U}_q$ is the bicubic interpolator with a magnification factor of q). Since we want dictionary $\mathbf{\Psi}_h$ contains useful discriminative high-frequency information, the original HR training set $\{\mathbf{Z}_j\}$ is further processed to obtain an updated HR training set $\{\mathbf{Z}_j^U, j=1,2,\dots,N\}$ via $\mathbf{Z}_j^U=\mathbf{Z}_j-\mathbf{Y}_j^U$. For dictionary $\mathbf{\Psi}_l$, many previous work demonstrated that building it in feature space is more suitable than in image space [15,16,21]. Considering the fact that the high-frequency information of the LR image is critical for predicting the lost high-frequency information in the target HR image, people often choose the feature space as some kind of high-pass filtered image. For example, Yang et al. [15] used the first- and second-order derivatives as the feature due to their simplicity and effectiveness, Rueda et al. [21] applied a multi-scale edge analysis, where a series of 6 different filters (Sobel kernels, size $3\times3\times3$ and $5\times5\times5$, in x, y and z directions) are used. To
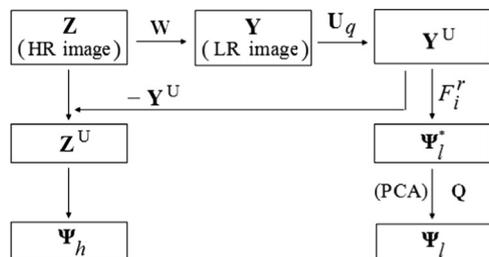
combine the merits of both multi-scale analysis and the first- and second-order derivatives, in this paper, we use a multi-scale (size 3×3 and 5×5) first- and second-order derivative analysis to extract features from the upsampled image set $\{\mathbf{Y}_j^U\}$.

The eight 1-D filters used to extract multi-scale first- and second-order derivatives are

$$F_1^1 = [-1, 0, 1]; F_2^1 = [-1, 0, 1]^T; F_3^1 = [-1, -2, 0, 2, 1];$$
$$F_4^1 = [-1, -2, 0, 2, 1]^T$$
$$F_1^2 = [1, -2, 1]; F_2^2 = [1, -2, 1]^T; F_3^2 = [1, 0, -2, 0, 1];$$
$$F_4^2 = [1, 0, -2, 0, 1]^T \qquad (24)$$

where F_i^1 and F_i^2 ($i=1$ to 4) are the first- and second-order derivative filters, respectively. For each image in the upsampled image set $\{\mathbf{Y}_j^U\}$, applying these eight filters we obtain eight different filtered images $\{F_i^r\mathbf{Y}_j^U, r=1,2$ and $i=1$ to 4$\}$.

Following the preprocessing steps described hereinbefore, for each updated HR training image $\mathbf{Z}_j^U$, we now have eight different filtered LR images $\{F_i^r\mathbf{Y}_j^U\}$. The procedure for constructing dictionaries $\mathbf{\Psi}_h$ and $\mathbf{\Psi}_l$ are thus summarized as follows:

1. At each location d of the updated HR training image $\mathbf{Z}_j^U$, extract a patch $\mathbf{p}_\mathbf{Z}^d$ of size $m\times m$.
2. Extract the corresponding LR patches of the same size from the eight filtered images $\{F_i^r\mathbf{Y}_j^U, r=1,2$ and $i=1$ to 4$\}$ at the same location. Then concatenate all the eight LR patches to form a single vector $\overline{\mathbf{p}_\mathbf{Y}^d}$ of length 8 m².
3. Construct the HR dictionary $\mathbf{\Psi}_h$ and a temporary LR dictionary $\mathbf{\Psi}_l^*$ by gathering all patches $\{\mathbf{p}_\mathbf{Z}^d\}$ and $\{\overline{\mathbf{p}_\mathbf{Y}^d}\}$, respectively.
4. Apply Principal Component Analysis (PCA) to $\mathbf{\Psi}_l^*$, build the corresponding orthogonal transformation matrix $\mathbf{Q}$ by collecting the eigenvectors of the covariance matrix that represents at least 90% of the original variance.
5. Construct the LR dictionary $\mathbf{\Psi}_l$ via $\mathbf{Q\Psi}_l^*$.

The reason why we use $\mathbf{\Psi}_l$ instead of $\mathbf{\Psi}_l^*$ as the final LR dictionary is the redundancy of multi-scale first- and second-order derivative analysis, since eight different filters are applied to the same image, resulting in complementary but redundant information.

5.2. Global regularization by back-projection

Since LR dictionary $\mathbf{\Psi}_l$ is not constructed in the original LR image space, thus the original fidelity constraint $\|\mathbf{y}-\mathbf{\Psi}_l\boldsymbol{\alpha}\|_2^2$ in the image space must be replaced by a corresponding constraint in the feature space, which does not demand exact equality between the LR patch $\mathbf{y}$ and its estimation $\mathbf{\Psi}_l\boldsymbol{\alpha}$. On the other hand, because no continuity conditions are imposed along the boundaries between patches, the reconstructed HR image $\mathbf{Z}^\wedge$ should thus be further refined to satisfy the SR model (1). To this end, we simply project $\mathbf{Z}^\wedge$ onto the solution space (i.e., $\mathbf{Y}=\mathbf{WZ}$[15,21]), computing

$$\mathbf{Z}^* = \arg\min_\mathbf{Z}\left\{\|\mathbf{Y}-\mathbf{WZ}\|_2^2 + \xi\|\mathbf{Z}-\hat{\mathbf{Z}}\|_2^2\right\} \qquad (25)$$

where ξ is a controlling parameter. Instead of using standard gradient descent method to solve (25), we can iteratively calculate the difference $\mathbf{Y}-\mathbf{WZ}$, convolve it with a back-projection kernel, then warp back into the HR image space to update the estimated HR image. This process can be written as [15,21,48,49]

$$\mathbf{Z}^{t+1} = \mathbf{Z}^t + (\mathbf{U}_q(\mathbf{Y}-\mathbf{WZ}^t))*g \qquad (26)$$

where $\mathbf{Z}^t$ is the estimate of the HR image after the tth iteration, $\mathbf{U}_q$ the bicubic interpolator with a magnification factor of q, g is the back-projection filter and $*$ is the convolution operator. This updating process is iteratively repeated until the difference between two consecutive images is less than a given threshold.

Fig. 4. Illustration of low- and high-resolution dictionary construction.
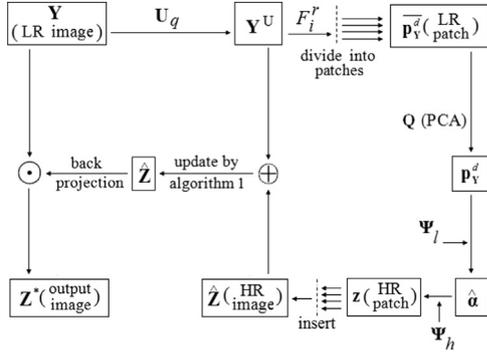
Fig. 5. Illustration of the proposed SR image reconstruction algorithm.

5.3. The proposed SR algorithm

With dictionaries $\mathbf{\Psi}_h$ and $\mathbf{\Psi}_l$, to reconstruct a HR image $\mathbf{Z}$ from a given LR image $\mathbf{Y}$, our proposed SR image reconstruction algorithm (as depicted in Fig. 5) is outlined in Algorithm 2.

Algorithm 2.

1) Upsample the LR image $\mathbf{Y}$ using $\mathbf{Y}^U = \mathbf{U}_q\mathbf{Y}$
2) Perform multi-scale first- and second-order derivative analysis (i.e., apply the eight 1-D filters defined in (24) to $\mathbf{Y}^U$ to get eight different filtered images $\{F_i^r\mathbf{Y}^U, r=1,2 \text{ and } i=1 \text{ to } 4\}$)
3) Divide $\mathbf{Y}^U$ into a grid of nonoverlapping patches with size of $m \times m$. For each patch $\mathbf{y}$ in the image $\mathbf{Y}^U$, do
   - Concatenate the patches of the eight filtered images $\{F_i^r\mathbf{Y}^U\}$ that correspond to the same location of $\mathbf{y}$ to form a patch vector $\overline{\mathbf{p}_\mathbf{Y}^d}$
   - Reduce the dimensionality of $\overline{\mathbf{p}_\mathbf{Y}^d}$ via $\mathbf{p}_\mathbf{Y}^d = \mathbf{Q}\overline{\mathbf{p}_\mathbf{Y}^d}$
   - Substitute $\mathbf{p}_\mathbf{Y}^d$ for $\mathbf{y}$ in (18) and solve the optimization problem defined in (19)
   - Generate the HR patch $\mathbf{z}$ via $\mathbf{z} = \mathbf{\Psi}_h\hat{\boldsymbol{\alpha}}$
   - Insert patch $\mathbf{z}$ into the corresponding location of the HR image $\mathbf{Z}^\wedge$
4) Update $\hat{\mathbf{Z}}$ using $\hat{\mathbf{Z}} = \hat{\mathbf{Z}} + \mathbf{Y}^U$
5) Use Algorithm 1 to update $\hat{\mathbf{Z}}$
6) Use (26), find the image $\mathbf{Z}^*$, which is the closest image to $\hat{\mathbf{Z}}$ that satisfies the global reconstruction constraint (25)
7) Output $\mathbf{Z}^*$ as the final result of SR image reconstruction

6. Experimental results

6.1. Tested methods

To examine more comprehensively the proposed approach, we test two versions of the proposed method: one using sparse representation prior and nonlocal similarity (denoted by SRNL, step 5 is thus excluded from Algorithm 2); the other one using sparse representation prior, nonlocal similarity and sparse derivative prior (denoted by SRNINL). To examine the effect of sparse representation prior and nonlocal similarity separately, we test the sparse representation prior based method proposed by Rueda et al. [21] (denoted by SRA) and nonlocal similarity based upsampling [12] (denoted by NLUP). Table 1 clearly shows the relation among these four methods. On the other hand, two state-of-the-art methods, adaptive sparse domain selection and adaptive regularization (denoted by ASAR) [16], sparse regression and natural image prior (denoted by SRNI) [36], are also tested. Moreover, for a clear comparison among the above mentioned six methods, we also provide the results yielded by two baseline techniques, i.e., bicubic interpolation (denoted by SBI) and TV prior [23] (denoted by TV).

Table 1
The close related four algorithms.

Method	Description
SRNINL (the proposed method)	Use sparse representation prior, nonlocal similarity, and sparse derivative prior
SRNL	Use sparse representation prior and nonlocal similarity
SRA	Use only sparse representation prior [21]
NLUP	Use only nonlocal similarity [12]

Since SRA and NLUP are originally designed for 3D MR image reconstruction, here we implement a 2D version of SRA and simply draw out 2D slices from the 3D HR volume reconstructed by NLUP for comparison. On the other hand, considering that the original NLUP includes an additional image denoising step using MNLM3D [34] while the other tested approaches do not, to do a fair comparison, we exclude denoising operation from NLUP in all our experiments.

6.2. Implementation details

For dictionary construction, each HR training image is first blurred with a Gaussian kernel of size 3×3 and standard deviation 1, then down-sampled by a factor of 2 to produce the corresponding LR image. Finally, the LR image is magnified by a factor of 2 via bicubic interpolation to produce the upsampled LR image.

For the proposed algorithm, the observed LR image $\mathbf{Y}$ is also upsampled using bicubic interpolation with a magnification factor of 2. In our experiments, the magnification factor is also set to 2, with a patch size of 3×3 in LR image and 6×6 in HR image. Accordingly, the image patch size m at step 3 of Algorithm 2 is also set to 6. To search for nonlocal similar patches, the searching radius is set to 7 which represents a 15×15 searching window and the first 15 (i.e., $L=15$) closest patches are chosen to calculate the nonlocal similarity regularization term.

To evaluate the performance of SR algorithms we use four different MR data sets: one dictionary dataset and three evaluation datasets. Dictionary dataset contains twenty T1-weighted brain MR images of normal person or patients suffering from mild cognitive impairment (MCI) and Alzheimer's disease, all downloaded from Internet. For each image in the dictionary dataset, the slice thickness is 1.0 mm, slice dimension is 512×512 and the pixel size is 0.469 mm $\times$ 0.469 mm. The number of slices per volume varies between 144 and 168. Three evaluation datasets are Brainweb [33] dataset, ADNI dataset [45] and Cardiac MRI dataset [46].

To build the dictionaries, four slices per volume in the dictionary dataset are selected as the original training HR images. In order to improve the representative ability of dictionaries, we preprocess these images by cropping out the texture and edge regions and discarding the smooth parts. Dictionaries $\mathbf{\Psi}_h$ and $\mathbf{\Psi}_l$ are then constructed from the preprocessed HR and LR images, respectively. For all the tested sparse representation based methods (i.e., SRNL, SRNINL, SRA and ASAR), the size of final dictionaries is reduced to 1024 atoms.

In this paper, two quantitative measures are used to perform comparison between the reconstructed image $\mathbf{A}$ and the original image $\mathbf{B}$:

- Peak signal-to-noise ratio (PSNR):

$$\mathrm{PSNR}(\mathbf{A}, \mathbf{B}) = 10 \times \log_{10}\left(\frac{255^2}{||\mathbf{A}-\mathbf{B}||_2^2/C}\right) \qquad (27)$$

where C is the dimension of $\mathbf{A}$ or $\mathbf{B}$.

- Structural Similarity Index (SSIM) [32]:

$$\text{SSIM}(\mathbf{A}, \mathbf{B}) = \frac{(2\mu_{\mathbf{A}}\mu_{\mathbf{B}} + c_1)(2\sigma_{\mathbf{AB}} + c_2)}{(\mu_{\mathbf{A}}^2 + \mu_{\mathbf{B}}^2 + c_1)(\sigma_{\mathbf{A}}^2 + \sigma_{\mathbf{B}}^2 + c_2)} \qquad (28)$$

where $\mu_{\mathbf{A}}$ and $\mu_{\mathbf{B}}$ are the mean value of images $\mathbf{A}$ and $\mathbf{B}$, $\sigma_{\mathbf{A}}$ and $\sigma_{\mathbf{B}}$ are the standard deviation of images $\mathbf{A}$ and $\mathbf{B}$, $\sigma_{\mathbf{AB}}$ is the covariance of $\mathbf{A}$ and $\mathbf{B}$, $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ (L is the dynamic range, $k_1 = 0.01$ and $k_2 = 0.03$). To evaluate the computational complexity, the exact running time (denoted by CPU time) of each algorithm is used. For NLUP, since it reconstructs 3D MR images, we use the ratio of running time to the total slice number for comparison.

For the proposed SRNINL algorithm, there are total six parameters, i.e., the parameters $\sigma_N, \sigma_R, \alpha, X$ in Algorithm 1 and λ, η in Algorithm 2. Since it is very difficult to determine these parameters at the same time, a stepwise selection strategy is more feasible and thus is adopted here [47]. Specifically, we fix the parameters $\sigma_N, \sigma_R, \alpha, X$ in Algorithm 1 in advance and try to find the optimal values for λ and η. To this end, we choose the Basis Pursuit Solver provided in SparseLab library [31] to solve the $L1$-norm optimization problem defined in (19). We fix the first parameter λ in advance and try to find the optimal η, then the optimal λ is determined based on the chosen η. Finally, based on the chosen λ and η, the four parameters $\sigma_N, \sigma_R, \alpha, X$ in Algorithm 1 are determined using the same strategy. Fig. 6 shows the performance of SRNINL (in PSNR) over the variation of one parameter with other parameters fixed. From Fig. 6, we empirically set $\lambda = 0.01, \eta = 0.18, \alpha = 0.8, \sigma_R = 1, \sigma_N = 100$, and $X = 8$.

All the experiments are implemented in MATLAB R2010b, running on a personal computer with Intel(R) Core(TM)2 Duo CPU P8700 @ 2.53 GHz, 4 GB memory.

6.3. Tests on Brainweb dataset

In this section, we use T1-weighted normal brain MR images with a slice dimension of 181×217 (pixel size is $1\text{ mm} \times 1\text{ mm}$) and an interslice distance of 1 mm, generated by Brainweb [33] digital brain phantom.

6.3.1. SR results on clean data

We test the clean (here clean means with 0% intensity non-uniformity and 0% noise) data first. We randomly pick out 10 slices, downsample them by a factor of two to generate the corresponding LR images, and then perform SR image reconstruction using SRNL, SRNINL, SBI, TV, SRA, ASAR, and SRNI. For NLUP, we downsample the original volume data by a factor of two to get the LR volume data and pick out the corresponding reconstructed 2D HR slices for comparison. Fig. 7 presents one of the reconstructed slices by all the methods. From Fig. 7, we can see that comparing with the result got by SRNL, TV and SRA produce jaggy artifacts and NLUP produces more blurry edges in the center (marked by the red box). TV and SRA also blur the details on the bottom left (marked by the white box). The results got by SRNI and ASAR are as better as SRNL. Nevertheless, by jointly using sparse representation prior, nonlocal similarity and sparse derivative prior, the proposed SRNINL algorithm produces the best result among all the tested methods. The averages of two quantitative measures across 10 slices are reported in Table 2. Note that the two quantitative measures show good consistence with the visual results in Fig. 7. To compare the computational complexity, Table 3 lists the corresponding CPU time. Except SBI, the fastest algorithm is NLUP. This is mainly because NLUP is an iterative filtering process that converges very fast. Its main drawback is that it must reconstruct the whole 3D image completely. The actual running time of NLUP in this case is $1.106\text{ s} \times 180\text{ slices} = 199.08\text{ s}$.

Since the basic principle of SRA and SRNL is the same, the computational complexity of SRA is comparable with that of SRNL. By adding a filtering operation to SRNL, the resulted SRNINL is a little bit slower than SRNL, but still faster than ASAR.

6.3.2. SR results on noisy data

In the second test, we test the algorithms' robustness to noise. We download the noisy data (noise: 9%, intensity non-uniformity: 0%) and repeat the same procedure as we did in the first experiment. Fig. 8 presents one of the reconstructed slices by all the methods. From Fig. 8, we see that unlike methods using either local smooth assumption (e.g., TV) or nonlocal similarity (e.g., NLUP, SRNL, ASAR, SRNINL) to suppress noise, SRA and SRNI are more sensitive to noise and there are many obvious noise-caused artifacts in the center. On the other hand, NLUP, ASAR, and SRNL produce blurry details in the center (marked by the red box), and the structures on the bottom left (marked by the white box) produced by NLUP and ASAR is over-smoothed. In contrast, the proposed SRNINL shows good robustness to noise: not only the noise is effectively suppressed, but also the weak edges are well reconstructed. This is mainly because the noise can be more effectively removed and the edges can be better preserved via jointly using sparse representation prior, nonlocal similarity and sparse derivative prior. Table 4 lists the corresponding average quantitative measures. An interesting observation is that, although SRNINL generates visually more appealing image than NLUP and ASAR do, its quantitative measure SSIM is actually lower than that of NLUP and ASAR. Since intensive noise is included in this experiment and the PSNR of SRNINL is much higher than that of other methods, we believe that SSIM may not be a good measure under noisy cases (as we will demonstrate later in other experiments that the SSIM of SRNINL on other noiseless data is the best among all the test methods).

6.3.3. Effect of intensity non-uniformity

To study the effect of intensity non-uniformity, we download the data with 40% intensity non-uniformity (0% noise) and repeat the same procedure as we did in the previous two experiments. Fig. 9 presents the reconstructed results of the same slice in Fig. 7 by all the tested methods and Table 5 lists the corresponding quantitative measures. By comparing Fig. 9 with Fig. 7, as well as Table 5 with Table 2, we can easily see that

1) Intensity non-uniformity makes almost invisible structure changes to the original clean slice.
2) The performance of SBI, TV, SRA, SRNL, ASAR, SRNI and SRNINL are almost the same as they are in Section 6.3.1, since they all build a 2D HR image from the given 2D LR image. Moreover, SRNINL still produces the best result among all the tested methods.
3) The performance of NLUP degrades dramatically in Fig. 9 and Table 5. This is mainly because that though intensity non-uniformity does not make great visible structure changes inside slice plane, it causes interslice structure changes, which will hamper NLUP to extract useful sub-pixel information via interslice correspondence.

6.3.4. SR results under large magnification factor

In the previous experiments, the scaling factor is fixed to 2. However, it could be interesting to test the algorithms' performance under a larger magnification factor, since more HR patch patterns are associated to a single LR patch. To this end, we set the scaling factor to 4, with a patch size of 3×3 in LR image and 12×12 in HR image. For SRNL, SRNINL, and SRA, we do not use the LR images downsampled by a factor of 4 to construct a new dictionary Ψ_l as Rueda et al. did in [21], we instead adopt the same dictionaries
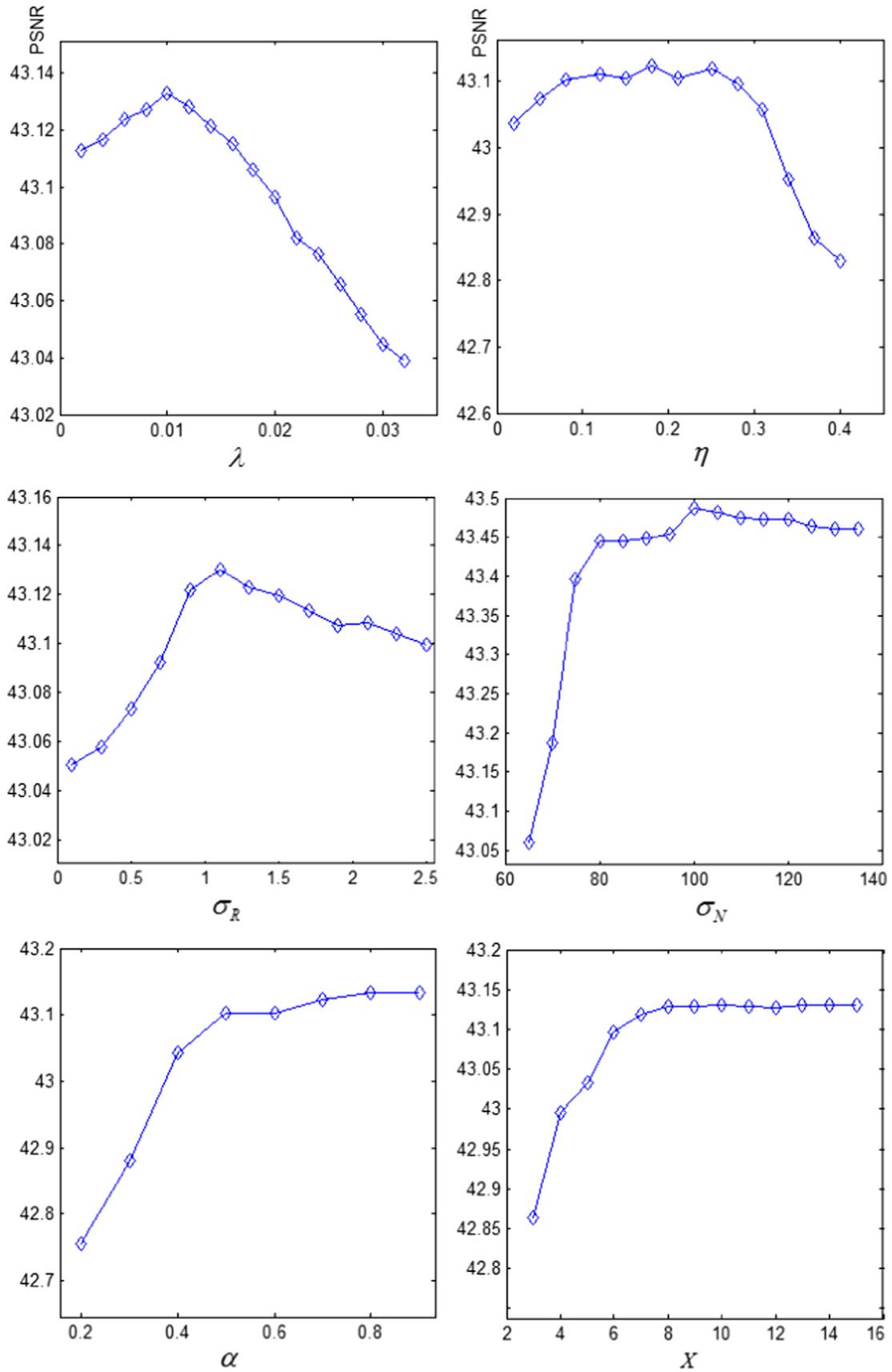
Fig. 6. Illusion of the performance of SRNINL (in PSNR) over the variation of $\lambda, \eta, \sigma_N, \sigma_R, \alpha$, and X.
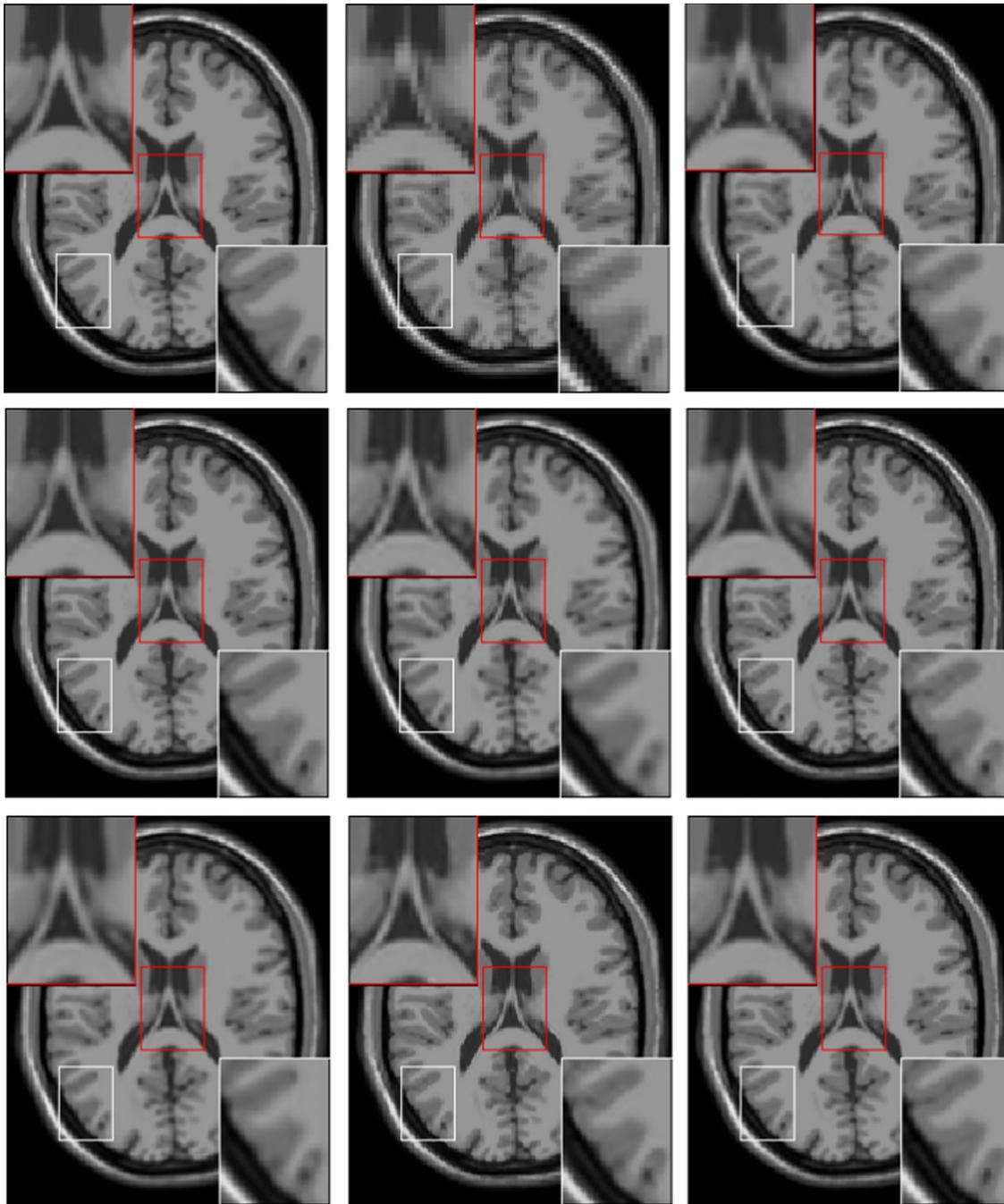
Fig. 7. Comparison of SR results (magnified by a factor of 2). Top row: original, SBI and TV. Middle row: NLUP, SRA and SRNL. Bottom row: ASAR, SRNI and SRNINL. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
The average of PSNR and SSIM on clean data (magnified by a factor of 2).

	SBI	TV	NLUP	SRA	SRNL	ASAR	SRNI	SRNINL
PSNR	33.918	37.496	41.292	39.559	40.144	39.454	39.363	**43.033**
SSIM	0.9088	0.9702	0.9862	0.9774	0.9833	0.9766	0.9802	**0.9952**

Table 3
The average of CPU time on clean data (seconds).

		SBI	TV	NLUP	SRA	SRNL	ASAR	SRNI	SRNINL
CPU time	$\times 2$	0.003	5.726	1.106	107.371	104.974	**138.945**	4.904	107.805
	$\times 4$	0.004	133.967	0.519	145.578	123.318	**150.782**	15.478	134.209
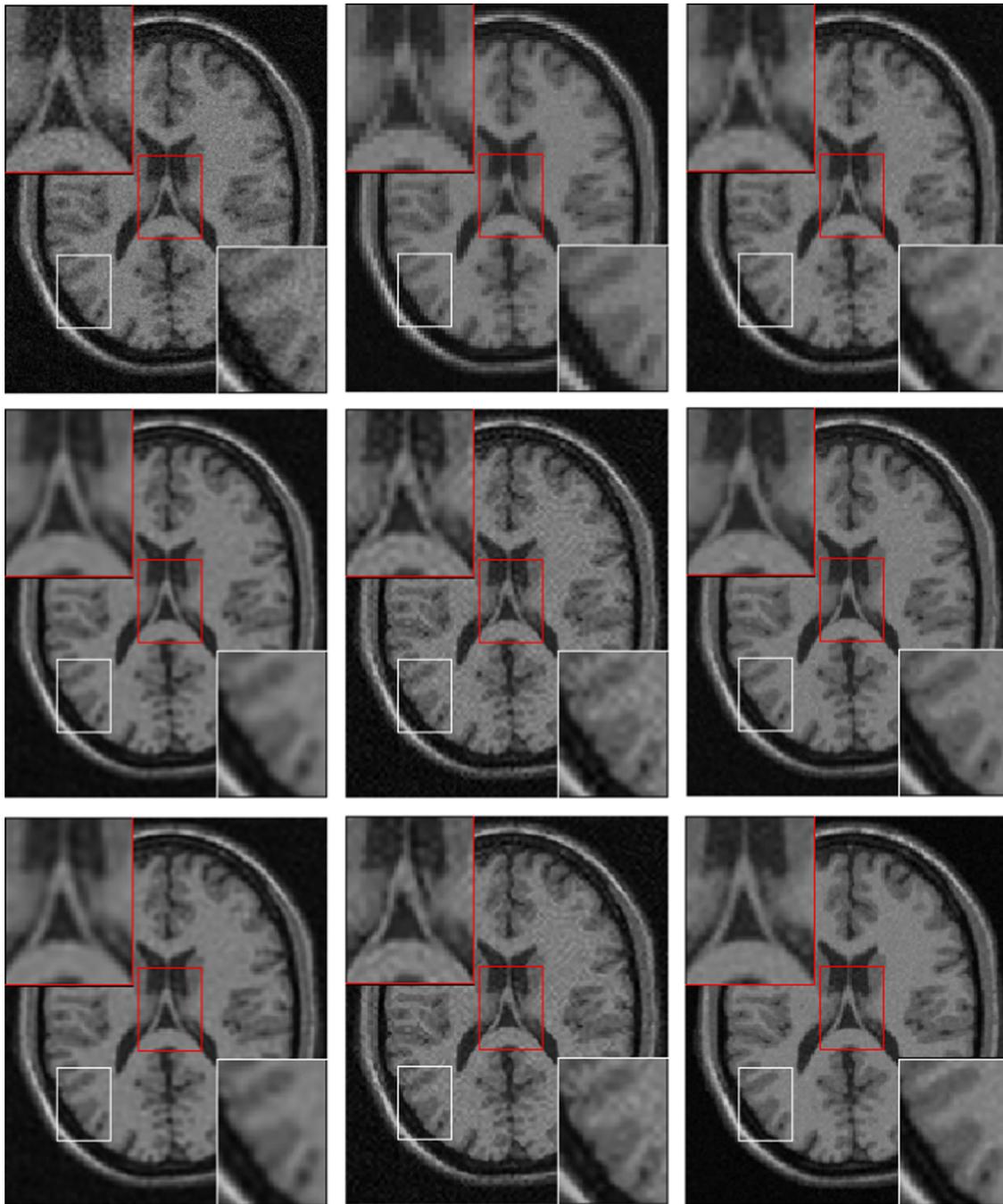
Fig. 8. Comparison of SR results on noisy data (magnified by a factor of 2). Top row: original, SBI and TV. Middle row: NLUP, SRA and SRNL. Bottom row: ASAR, SRNI and SRNINL. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
The average of PSNR and SSIM on data with 9% noise (magnified by a factor of 2).

	SBI	TV	NLUP	SRA	SRNL	ASAR	SRNI	SRNINL
PSNR	30.205	30.694	30.841	30.895	30.881	30.847	30.944	**32.281**
SSIM	0.6883	0.7124	**0.7529**	0.6831	0.7438	0.7503	0.6862	0.7455

used in the previous experiments and magnify a LR image by two consecutive steps with a magnification of 2 for each step. For TV, NLUP, ASAR, and SRNI, we simply magnify the LR image by a factor of 4 since the codes provided by the authors are capable of executing a magnification of 4 directly. Fig. 10 presents one of the reconstructed slices by all the methods and the corresponding numeric results are reported in Table 6. From Fig. 10, we can still see that the proposed SRNINL algorithm produces the best result among all the tested algorithms. However, Table 6 shows that the PSNR of SRNINL is lower than that of SRNL, though it generates visually more appealing image than SRNL does. This is because for the proposed SRNINL algorithm, an extra post-filtering operation (if we add Algorithm 1, the sparse derivative prior based post-filtering operation into SRNL, it turns out to be SRNINL) may degrade
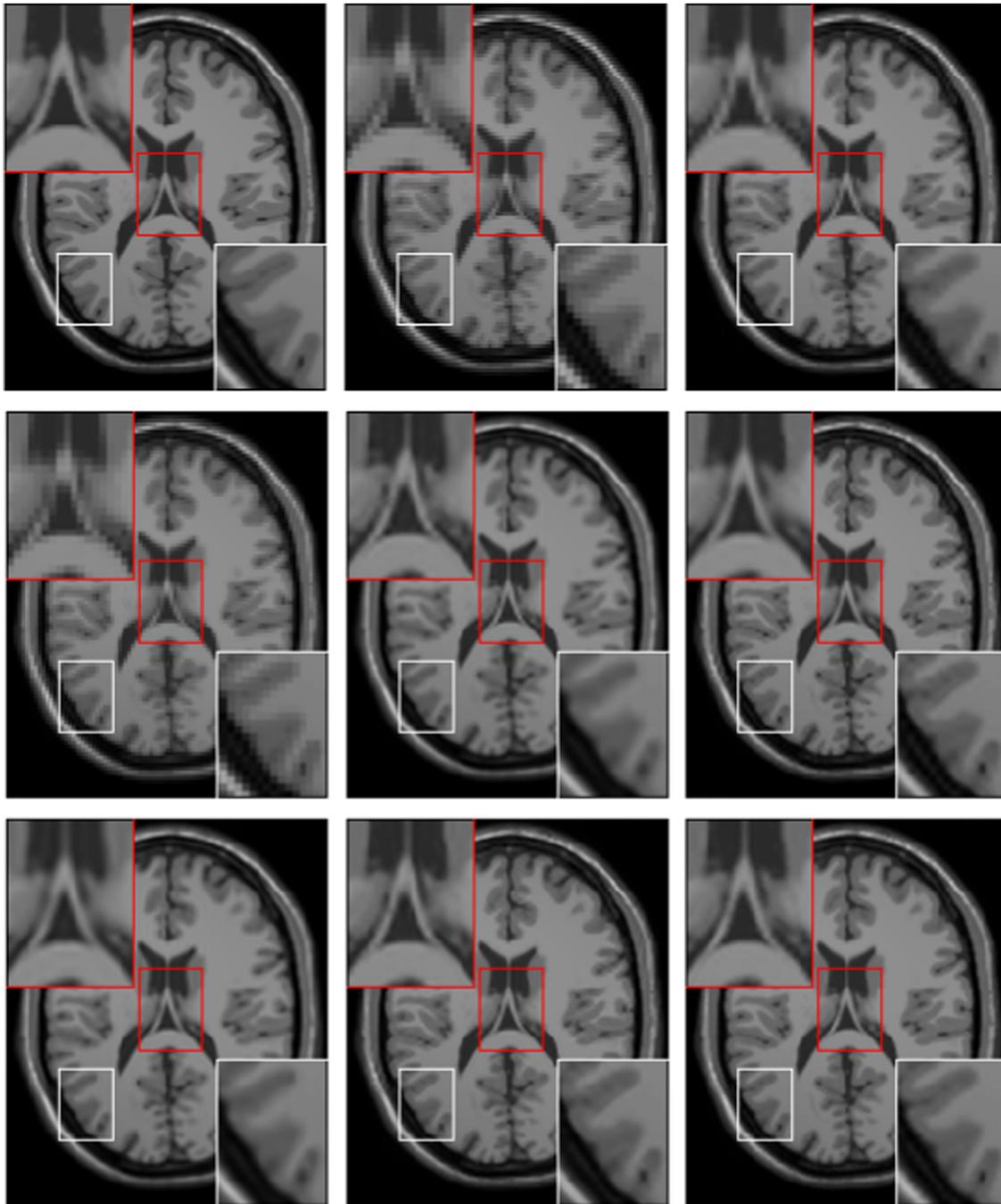
Fig. 9. Comparison of SR results on intensity inhomogeneous data (magnified by a factor of 2). Top row: original, SBI and TV. Middle row: NLUP, SRA and SRNL. Bottom row: ASAR, SRNI and SRNINL.

Table 5
The average of PSNR and SSIM on data with 40% intensity non-uniformity (magnified by a factor of 2).

	SBI	TV	NLUP	SRA	SRNL	ASAR	SRNI	SRNINL
PSNR	34.1714	38.2919	34.4714	40.2691	40.2734	40.1395	40.2795	**42.1689**
SSIM	0.9108	0.9711	0.9118	0.9771	0.9812	0.9757	0.9809	**0.9930**

pixel-by-pixel based quantitative measures (e.g., PSNR) under large magnification factors. However, the results in this paper demonstrate that pixel-by-pixel based quantitative measures (e.g., PSNR) may work well in noisy cases, SSIM produces more reliable statistics for noiseless images. The corresponding CPU time of each method is also listed in Table 3 for comparison. Except NLUP, all the other methods need more time to reconstruct the final HR image since more unknown pixels are needed to be calculated when magnification factor increases from two to four. The reason why NLUP spends less time for large magnification factors is that it reaches convergence very quickly when there is no sufficient information among neighboring pixels (the larger magnification factor, the less corelation among neighboring pixels). On the other hand, the proposed SRNINL is still faster than ASAR.
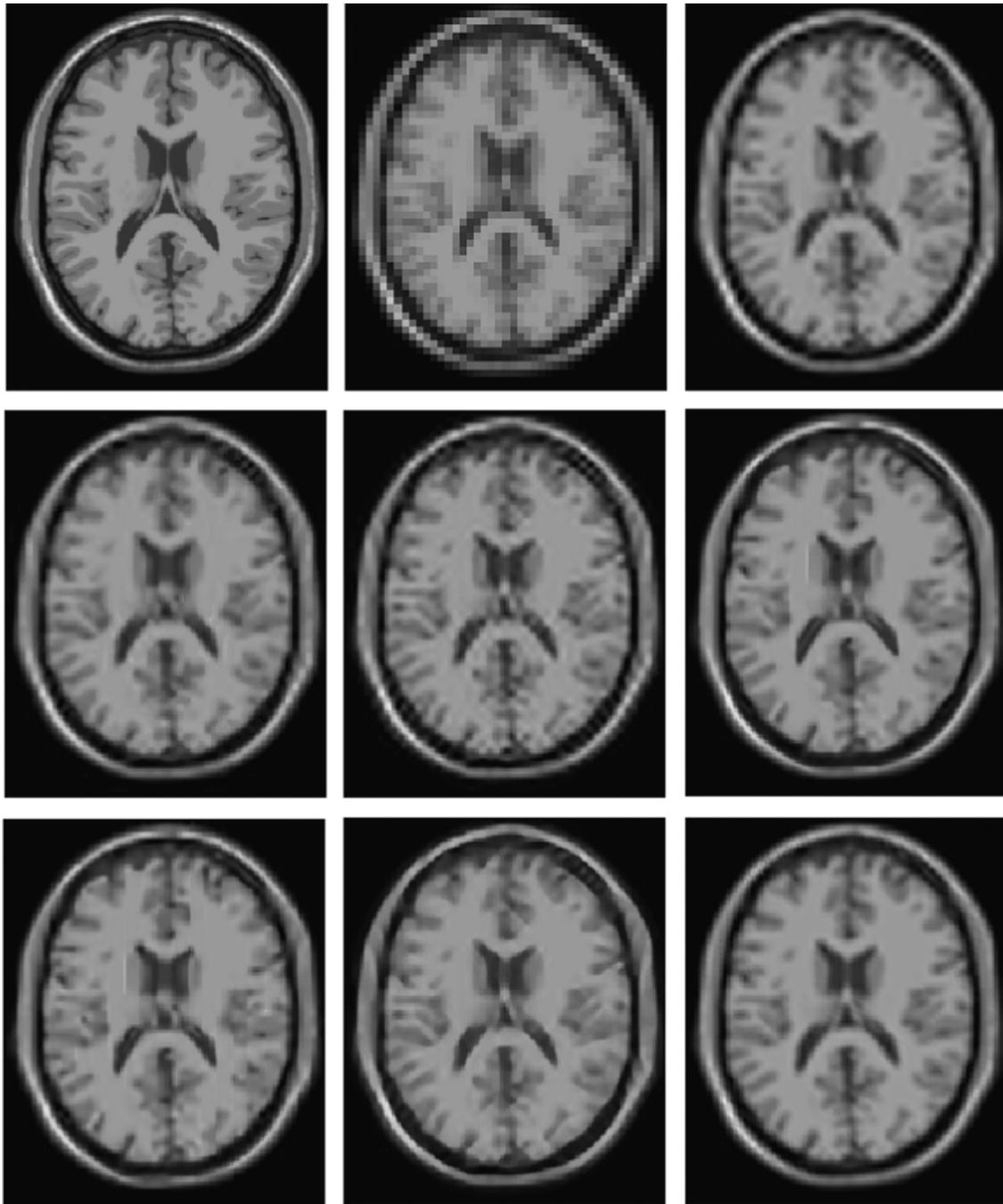
Fig. 10. Comparison of SR results (magnified by a factor of 4). Top row: original, SBI and TV. Middle row: NLUP, SRA and SRNL. Bottom row: ASAR, SRNI and SRNINL.

Table 6
The average of PSNR and SSIM on clean data (magnified by a factor of 4).

	SBI	TV	NLUP	SRA	SRNL	ASAR	SRNI	SRNINL
PSNR	31.785	33.170	33.319	33.211	**34.110**	33.459	33.250	34.079
SSIM	0.7182	0.8614	0.8652	0.8581	0.8861	0.8640	0.8565	**0.9096**

6.4. Tests on ADNI dataset

The data used in this section were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [45]. The Principal Investigator of ADNI is Michael W. Weiner, MD, VA Medical Center and University of California. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations. The primary goal of ADNI has been to test whether serial MR imaging, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

We download the data and repeat the same procedure as we did in the previous experiments. Fig. 11 presents one of the

reconstructed slices by all the methods and the corresponding numeric results are reported in Table 7. From Fig. 11, we can see that for the down-right area (marked by the red box), SRA and ASAR produce the most blurry results. By incorporating the nonlocal similarity information properly, SRNL and NLUP produce more clear tissue structures. Though SRNI produces better result than SRNL does, by jointly using sparse representation prior, nonlocal similarity and sparse derivative prior, the proposed SRNINL recovers more fine details. Note that in Table 7 the quantitative measures of NLUP are higher than that of SRNI, which is in contradiction with the visual perception shown in Fig. 11.

6.5. Tests on Cardiac MRI dataset

The Cardiac MRI dataset [46] contains 4D MR images acquired from 33 subjects. Each subject's sequence consists of 20 frames and 8–15 slices along the long axis, for a total of 7980 images. Since the original data contains time axis, we simply choose the volume data by setting t to zero and repeat the same procedure as we did in the previous experiments. Fig. 12 presents one of the reconstructed slices by all the methods and the corresponding numeric results are

reported in Table 8. From Fig. 12, we can see that for the tissues marked by the small white and red boxes, SRA and ASAR still produce the most blurry results. On the other hand, SRNL and NLUP produce more clear tissue structures since nonlocal similarity information is properly incorporated into the reconstruction process. The result yielded by SRNI is slightly better than that of SRNL. Nevertheless, our proposed SRNINL still produces the best tissue structures among all the tested methods. Again, as we found in Table 7, though SRNI generates visually more appealing image than NLUP does, in Table 8, the quantitative measures of NLUP are higher than that of SRNI.

7. Discussion

In this paper, we propose a new algorithm for reconstructing a HR image from a single LR MR image by jointly using sparse representation prior, nonlocal similarity, and sparse derivative prior. The proposed method has been demonstrated to achieve much better results than many state-of-the-art algorithms in terms of both quantitative measures and visual perception. Our main contribution is threefold: (1) the use of multi-scale first- and
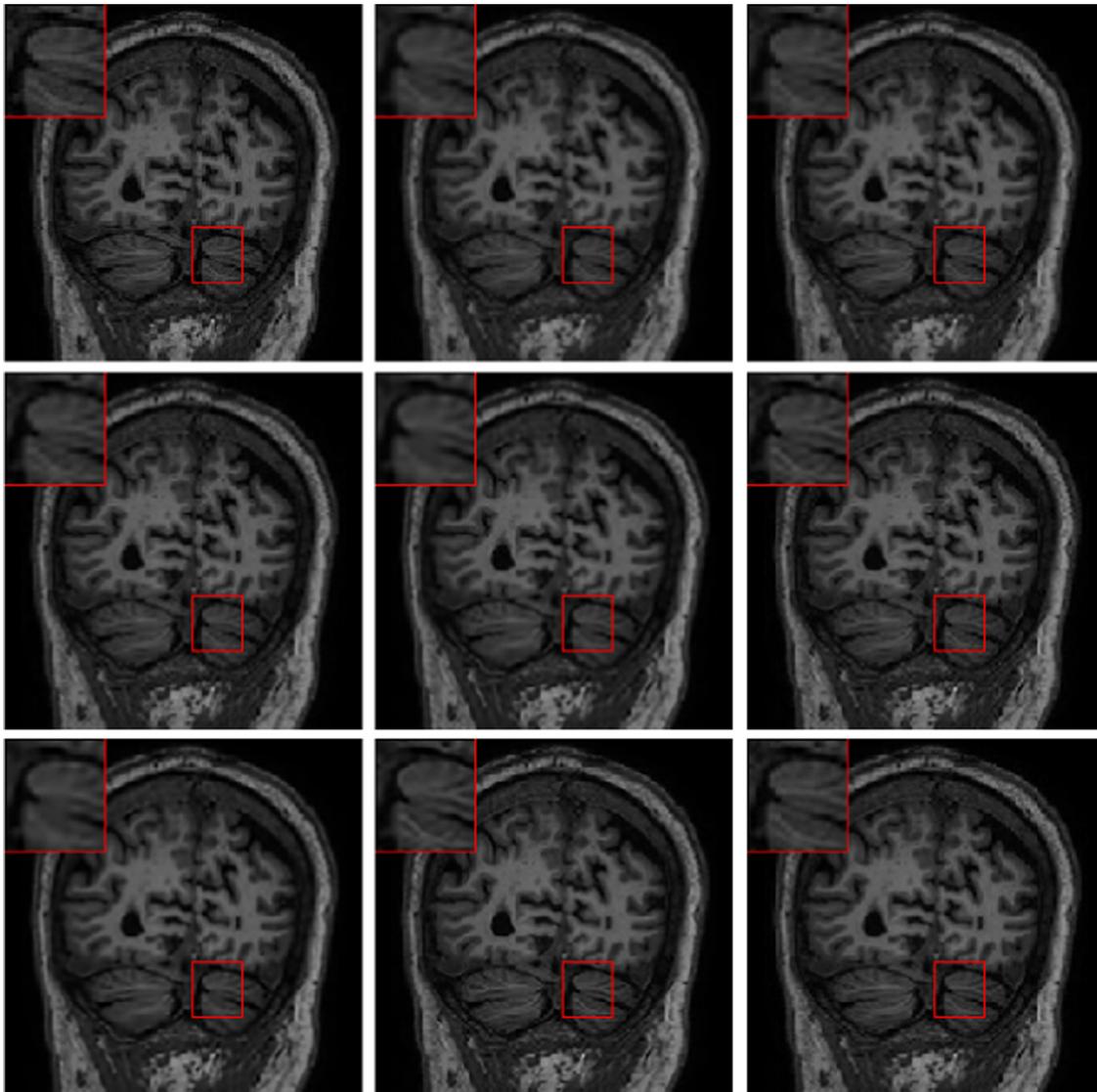
Fig. 11. Comparison of SR results on ADNI dataset (magnified by a factor of 2). Top row: original, SBI and TV. Middle row: NLUP, SRA and SRNL. Bottom row: ASAR, SRNI and SRNINL. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 7
The average of PSNR and SSIM on ADNI dataset (magnified by a factor of 2).

	SBI	TV	NLUP	SRA	SRNL	ASAR	SRNI	SRNINL
PSNR	35.6354	35.7876	36.4898	35.6288	35.7580	35.6512	35.8142	**37.7246**
SSIM	0.8929	0.8983	0.9199	0.8860	0.9050	0.8871	0.9056	**0.9281**
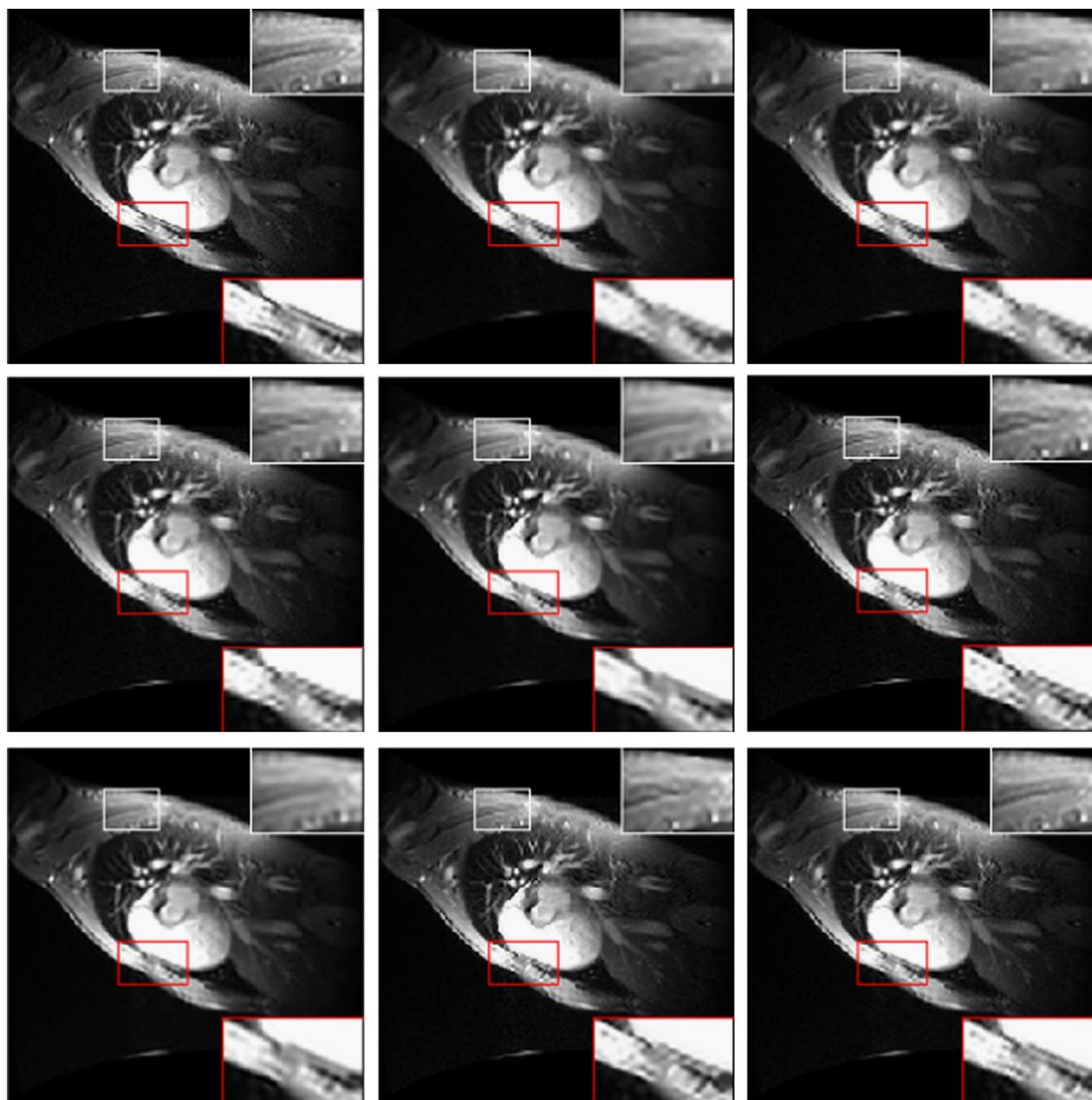
Fig. 12. Comparison of SR results on Cardiac MRI dataset (magnified by a factor of 2). Top row: original, SBI and TV. Middle row: NLUP, SRA and SRNL. Bottom row: ASAR, SRNI and SRNINL. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 8
The average of PSNR and SSIM on Cardiac MRI dataset (magnified by a factor of 2).

	SBI	TV	NLUP	SRA	SRNL	ASAR	SRNI	SRNINL
PSNR	37.4095	37.5550	38.4542	37.6537	37.8962	37.7649	37.9704	**39.6499**
SSIM	0.9303	0.9350	0.9528	0.9224	0.9400	0.9230	0.9408	**0.9703**

second-order derivative analysis to estimate the missing high-frequency information, (2) the joint use of sparse representation and nonlocal similarity under a unified $L1$-norm minimization framework, and (3) the use of sparse derivative prior based post-processing in MR image SR reconstruction.

The use of multi-scale edge analysis to estimate high-frequency information was first proposed by Rueda et al. [21] based on two considerations: (1) compared with low-frequency information, high-frequency information has more influence on the reconstruction of sharp edges, and (2) image coherence and regularity are also

preserved via multi-scale analysis to some extent, which makes patch overlapping unnecessary. However, only multi-scale first-order derivative analysis was applied in [21]. Since most edges in images are often gradual transitions from one intensity to another, using first-order derivative one would usually get a curve with rising gradient magnitude, and then a falling gradient magnitude. Extracting the ideal edge is thus a matter of finding this curve with optimal gradient. For patch-based SR image reconstruction, however, such curve may pass through many neighboring patches, the exact location of the edges would probably be lost for these neighboring patches are separately reconstructed. Second-order derivate, on the other hand, could be used to extract optimal edge location by finding where it is zero. We must point out that though only the simplest second-order derivative analysis is used in this paper, using other more reliable ones, e.g., Laplacian operator, Marr–Hildreth operator, is also possible.

Nonlocal similarity refers to the fact that the patches with similar patterns can be spatially far from each other, we can collect them in the whole image and process them simultaneously for various purposes, e.g., image denoising [29,35], deblurring [16,34], and SR [12,30]. In this paper, by incorporating the nonlocal similarity regularization into the process of MR image SR reconstruction, we got better results than only using either nonlocal similarity or sparse representation prior. Moreover, thanks for the averaging operation of similar patches across the image, the proposed method is more robust to noise than TV, SRA and SRNI. Finally, the proposed SRNINL algorithm handles nonlocal similarity information in pixel domain, how about coding nonlocal similarity via sparse representation and including it in the same unified SR image reconstruction framework, this is one of our future research directions.

Though natural image prior has been proven to be effective in various natural image processing tasks, in this paper, we validate that it is also applicable to MR image processing. In fact, we can see that a MR image consists mainly of zero gradient regions interspersed with occasional strong gradient transitions. By comparing the results yielded by SRNL and SRNINL, it is easy to find that a sharp edge is preferred over a blurry one in the final output when applying the natural image prior based post-processing operation. However, a more direct and elegant approach for jointly using sparse representation prior, nonlocal similarity, and sparse derivative prior is to incorporate a sparse derivative prior based regularization term into (19) as follows

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\{ \|\mathbf{y}^* - \boldsymbol{\Gamma}\boldsymbol{\Psi}_l\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 + \zeta\left(|\mathbf{D}*\boldsymbol{\Psi}_l\boldsymbol{\alpha}|/\sigma\right)^{\alpha} \right\} \quad (29)$$

where $\mathbf{D}$ is the desired localized, oriented, and bandpass filter, and ζ is a parameter controlling the contribution of sparse derivative prior. Unfortunately, for sparse derivative prior, the exponent parameter α must satisfy $0 < \alpha < 1$, which makes solving (29) via directly applying standard $L1$-norm optimization techniques impossible. One of our future research directions will focus on finding an efficient algorithm to solve (29).

Conflict of interest statement

None declared.

Acknowledgement

References

[1] H. Greenspan, Super-resolution in medical imaging, Comput. J. 52 (1) (2008) 43–63.

[2] X.G. Wang, X. Tang, Hallucinating face by eigentransformation., IEEE Trans. Syst. Man Cybern. Part C 35 (3) (2005) 425–434.

[3] D. Zhang, H.F. Li, M.H. Du, Fast MAP-based multiframe super-resolution image reconstruction, Image Vision Comput. 23 (7) (2005) 671–679.

[4] S.C. Park, M.K. Park, M.G. Kang, Super-resolution image reconstruction: a technical overview, IEEE Signal Process. Mag. (May, 21–36, 2003).

[5] A. Herment, E. Roullot, I. Bloch, O. Jolivet, A.D. Cesare, F. Frouin, J. Bittoun, E. Mousseaux, Local reconstruction of stenosed sections of artery using multiple MRA acquisitions, Magn. Reson. Imaging 49 (4) (2003) 731–742.

[6] R. Shilling, T. Robbie, T. Bailloeul, K. Mewes, R. Mersereau, M. Brummer, A super-resolution framework for 3-D high-resolution and high-contrast imaging using 2-D multislice MRI, IEEE Trans. Med. Imaging 28 (5) (2009) 633–644.

[7] H. Greenspan, G. Oz, N. Kiryati, S. Peled, MRI inter-slice reconstruction using super-resolution, Magn. Reson. Imaging 20 (5) (2002) 437–446.

[8] R. Islam, A.J. Lambert, M.R. Pickering, Super resolution of 3D MRI images using a Gaussian scale mixture model constraint, in: Proc. ICASSP, 2012, pp. 849-852.

[9] S. Baker, T. Kanade, Limits on super- resolution and how to break them, IEEE Trans. Pattern Anal. Mach. Intell. 24 (9) (2002) 1167–1183.

[10] D. Zhang, J.Z. He, M.H. Du, Morphable model space based face super-resolution reconstruction and recognition, Image Vision Comput. 30 (2) (2012) 100–108.

[11] F. Rousseau. Brain hallucination, in: Proc. the European Conference on Computer Vision: Part I, 2008, pp. 497–508.

[12] J.V. Manjón, P. Coupé, A. Buades, V. Fonov, D.L. Collins, M. Robles, Non-local MRI upsampling, Med. Image Anal. 14 (12) (2010) 784–792.

[13] E. Candès, T. Tao, Near optimal signal recovery from random projections: universal encoding strategies? IEEE Trans. Inf. Theory 52 (12) (2006) 5406–5425.

[14] W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally centralized sparse representation for image restoration, IEEE Trans. Image Process 22 (4) (2013) 1620–1630.

[15] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process 19 (11) (2010) 2861–2873.

[16] W. Dong, L. Zhang, G. Shi, X. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, IEEE Trans. Image Process 20 (7) (2011) 1838–1857.

[17] L. Michael, L.D. David, J.M. Santos, M.P. John, Compressed sensing MRI, IEEE Signal Process. Mag. 25 (2) (2008) 72–82.

[18] J.C. Ye, S. Tak, Y. Han, H.W. Park, Projection reconstruction MR imaging using FOCUSS, Magn. Reson. Med. 57 (4) (2007) 764–775.

[19] G. Adluru, C. McGann, P. Speier, E.G. Kholmovski, A. Shaaban, E.V.R. DiBella, Acquisition and reconstruction of undersampled radial data for myocardial perfusion magnetic resonance imaging, J. Magn. Reson. Imaging 29 (2) (2009) 466–473.

[20] S. Ravishankar, Y. Bresler, MR image reconstruction from highly undersampled k-space data by dictionary learning, IEEE Trans. Med. Imaging 30 (5) (2011) 1028–1041.

[21] Andrea Rueda, Norberto Malpica, Eduardo Romero, Single-image super-resolution of brain MR images using overcomplete dictionaries, Med. Image Anal. 17 (1) (2013) 113–132.

[22] D. Zhang, M.H. Du Super-resolution image reconstruction via adaptive sparse representation and joint dictionary training, in: Proc. CISP, 2013, pp. 492-497.

[23] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D 60 (1992) 259–268.

[24] S.D. Babacan, R. Molina, A.K. Katsaggelos. Total variation super resolution using a variational approach, in: Proc. Int. Conf. Image Process, 2008, pp. 641–644.

[25] H.A. Aly, E. Dubois, Image up-sampling using total-variation regularization with a new observation model, IEEE Trans. Image Process 14 (10) (2005) 1647–1659.

[26] A. Marquina, S.J. Osher, Image super-resolution by TV-regularization and Bregman iteration, J. Sci. Comput. 37 (2008) 367–382.

[27] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.

[28] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc., Ser. B 58 (1) (1994) 267–288.

[29] L. Zhang, W. Dong, D. Zhang, G. Shi, Two-stage image denoising by principal component analysis with local pixel grouping, Pattern Recognit. 43 (4) (2010) 1531–1549.

[30] M. Protter, M. Elad, H. Takeda, P. Milanfar, Generalizing the nonlocal means to super-resolution reconstruction, IEEE Trans. Image Process 18 (1) (2009) 36–51.

[31] 〈http://sparselab.stanford.edu/〉.

[32] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. imoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process 13 (4) (2004) 600–612.

[33] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes A.C. Evans, Design and construction of a realistic digital brain phantom, IEEE Trans. Med. Imaging 17 (3) (1998) 463–468.

[34] S. Kindermann, S. Osher, P.W. Jones., Deblurring and denoising of images by nonlocal functional, Multiscale Model. Simul. 4 (4) (2005) 1091–1115.

[35] A. Buades, B. Coll, J.M. Morel., A review of image denoising algorithms, with a new one, Multiscale Model. Simul. 4 (2) (2005) 490–530.

[36] I.K. Kwang, K. Younghee, Single-image super-resolution using sparse regression and natural image prior, IEEE Trans. Pattern Anal. Mach. Intell. 32 (6) (2010) 1127–1133.

[37] M.F. Tappen, B.C. Russel, W.T. Freeman. Exploiting the sparse derivative prior for super-resolution and image demosaicing, in: Proc. IEEE Workshop Statistical and Computational Theories of Vision, 2003.

[38] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609.

[39] D.L. Ruderman, W. Bialek, Statistics of natural images: scaling in the woods, Phys. Rev. Lett. 73 (6) (1994) 814–817.

[40] E.P. Simoncelli. Statistical models for images: compression, restoration and synthesis, in: 31st Asilomar Conference on Signals Systems, and Computers. Pacific Grove, CA, 1997, pp. 673–678.

[41] E.P. Simoncelli, Bayesian denoising of visual images in the wavelet domain, Lecture Notes in Statistics, vol. 141, Springer-Verlag, New York (1999) 291–308.

[42] R.W. Buccigrossi, E.P. Simoncelli, Image compression via joint statistical characterization in the wavelet domain, IEEE Trans. Image Process 8 (12) (1999) 1688–1701.

[43] J. Dias. Fast GEM wavelet-based image deconvolution algorithm, in: IEEE International Conference on Image Processing, 2003.

[44] M. Figueiredo, R. Nowak Image restoration using the EM algorithm and wavelet-based complexity regularization, in: International Conference on Image Processing, 2002.

[45] ⟨http://adni.loni.usc.edu/⟩.

[46] A. Alexande, K.T. John, Efficient and generalizable statistical models of shape and appearance for analysis of Cardiac MRI, Med. Image Anal. 12 (3) (2008) 335–357.

[47] J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, J. Zhong, KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2) (2005) 230–244.

[48] M. Irani, S. Peleg, Motion analysis for image enhancement: resolution, occlusion, and transparency, J. Visual Commun. Image Represent. 4 (4) (1993) 324–335.

[49] D. Capel, Image Mosaicing and Super-Resolution, Springer-Verlag, 2004.

Di Zhang received his M.Sc. in Physics in 1996 from Wuhan University, Wuhan, China, and his Ph.D. in Communication and Information Systems in 2005 from South China University of Technology, Guangzhou. From 2005 to 2011, He joined the Department of Computer Science, Shaoguan University, Shaoguan. He is currently an Associate Professor at the School of Information Engineering, Guangdong Medical College, Dongguan. His current research interests include signal processing, image processing and pattern recognition.

Jiazhong He received his Ph.D. degree in communication and information systems in 2007 from South China University of Technology, Guangzhou. He is currently an associate professor at the Department of Physics, Shaoguan University, Shaoguan. His current research interests include pattern recognition and image processing.

Yun Zhao is currently a teacher at the School of Information Engineering, Guangdong Medical College, Dongguan. Her current research interests include computer vision and computer programming.

Minghui Du received his B.Sc. degree in electronics in 1985 from Fudan University, Shanghai, China, and his Ph.D. degree in electronic and communication systems in 1991 from South China University of Technology, Guangzhou, where he is currently a professor with the department of Electronics and communication engineering. Prof. Du is the vice chairman of Guangdong Institute of Graphics and Image, a committee member of the Chinese institute of Graphics and image, a committee member of the Chinese institute of biomedical engineering. His current research interests include digital imaging, medical image processing, signal processing and biomedical engineering.