

EXPLOITING MULTI-SCALE SPATIAL STRUCTURES FOR SPARSITY BASED SINGLE IMAGE SUPER-RESOLUTION

Yongqin Zhang, Jiaying Liu, Wei Bai, Zongming Guo*

Institute of Computer Science and Technology, Peking University, Beijing 100871, China

ABSTRACT

To improve the performance of sparsity-based single image super-resolution (SR), we propose a joint SR framework of structure prior based sparse representation (SPSR). The proposed SPSR algorithm exploits the multi-scale spatial structural self-similarities, the gradient prior and nonlocally centralized sparse representation to formulate a constrained optimization problem for high-resolution image recovery. The high-resolution image is firstly initialized by exploiting cross-scale patch redundancy in an image pyramid from single input low-resolution image. Then the sparse modeling of the image SR problem is proposed to refine it further, where the gradient histogram preservation is incorporated as a regularization term. Finally, an iterative solution is provided to solve the problem of model parameter estimation and sparse representation. Experimental results on image super-resolution validate the generality, effectiveness and robustness of the proposed SPSR algorithm.

Index Terms— Image super-resolution, sparse representation, self-similarities, sparse coding, dictionary learning.

1. INTRODUCTION

Single image super-resolution is still a challenging problem in the field of image restoration [5, 8]. It refers to the recovery of a suitable high-resolution (HR) image from an observed low-resolution (LR) image. The learning-based methods that exploit the information from training images have become the focus of the study in recent years [4, 6, 9]. Specifically, the sparse representation is used to train the dictionary and to estimate the target HR image. Recently, sparse coding with adaptive dictionary learning in LR and HR spaces has been received great attention in image super-resolution since it is an efficient technique for signal modeling. Yang *et al.* [9] proposed the coupled dictionaries trained from the LR and HR image patch pairs for single image super-resolution. Although the assumed similarities of geometry and sparse representation exist between the LR and HR spaces [9], they ignore

the case that there are many possible HR counterparts related to the LR input. Considering the sparse coding noise, Dong *et al.* [4] proposed a nonlocally centralized sparse representation (NCSR) model with very encouraging performance, which exploits patch redundancy within and across scales to train adaptive sub-dictionaries [6]. However, for the images with complex structures, these methods [4, 6, 9] cannot recover the small edges and the fine structural details. In addition, an obvious drawback of these methods is that their SR performance depends heavily on the noise level of input LR images. Therefore, there is still much space to further improve the performance of single image SR by exploiting prior knowledge of natural images.

It is observed that the image structures have high correlation between an image and its resized one with small-scale changes. Inspired by the redundancy of the multi-scale spatial image structures, we propose a novel joint framework of structure prior based sparse representation for single image super-resolution. The initial HR image is constructed through multistep magnification by exploiting the structural self-similarities in the scale and space domains. The gradient histogram of the observed LR image is incorporated as a gradient regularization term of sparse modeling. The priors of image gradient and nonlocally centralized sparsity are exploited for dictionary training and HR image reconstruction. The main contributions of our work can be summarized as follows: The prior knowledge of multi-scale spatial image structures is investigated and combined with other priors about the gradients and the sparsities to form the joint framework of single image super-resolution.

The remainder of this paper is organized as follows. Section 2 provides the detailed descriptions of the proposed SPSR algorithm. Experimental results are given in Section 3. The conclusions are drawn in Section 4.

2. SINGLE IMAGE SUPER-RESOLUTION

2.1. SR Model

For an observed LR image \mathbf{y} , the problem of image super-resolution is generally modeled as

$$\mathbf{y} = \mathbf{D}\mathbf{H}\mathbf{x} + \mathbf{v}, \quad (1)$$

* Corresponding author.

This work was supported by National Natural Science Foundation of China under contract No.61201442, Beijing Natural Science Foundation under contract No.4142021 and China Postdoctoral Science Foundation funded project under contract No.2013M530481.

where the degradation matrices \mathbf{D} and \mathbf{H} are separately the down-sampling and blurring operators, \mathbf{x} is the original image, and v is the noise term. Given a LR image, our goal is to produce a HR image such that its underlying high frequency details are recovered while preserving the structures from the original HR image. The flowchart of our proposed SPSR approach is shown in Figure 1, where the initialized HR image I_P is the enlarged output from a given LR image y by the multistep magnification technique, Φ is the learned dictionaries by k-means clustering and principal component analysis (PCA), w_S is the related weights of the similar patch groups by block matching, x_M is the updated HR image by the nonlocal means, x_G is the updated HR image by the gradient regularization with gradient histogram preservation, x_B is the updated HR image by the data fidelity constraint, and x_H is the final reconstructed HR image.

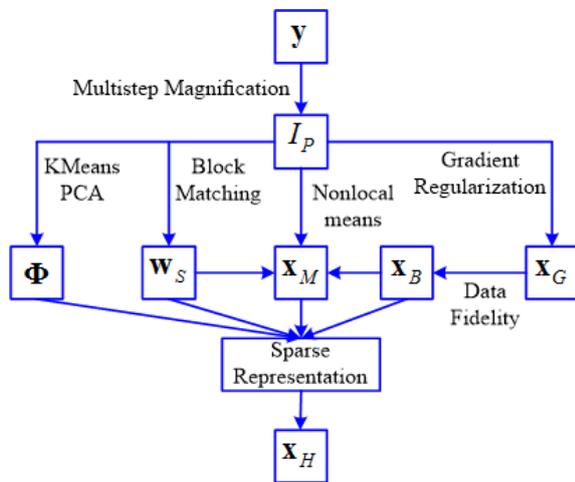


Fig. 1: The flowchart of our SPSR algorithm.

The initialization of the target HR image is a key procedure for single image super-resolution reconstruction. Motivated by the observations of the patch similarity widely existing within scale and across scales of natural images [6], different from the bicubic interpolation method, we design the multistep magnification scheme for the HR image initialization. The scale factor of each step is small enough so that more similar patches from an image pyramid are found for HR image reconstruction. In the image pyramid, assume that the relationship between HR image I_H and LR image I_L is given by

$$I_L = (I_H * G) \downarrow_s, \quad (2)$$

where $*$ is a convolution operator, G is an isotropic Gaussian kernel, and \downarrow_s is a downsampling operator with the scaling factor of s .

For an input LR image I_0 ($I_0 = y$), the downsampled LR images I_n ($n = -1, \dots, -N$) are generated to build the patch pair database B . Similarly, the input LR image I_0 is assumed to be a downsampled result from an unknown HR

image I_p ($p = 1, \dots, P$). For a total scaling factor z , the number of magnification steps is $P = \text{ceil}(\log(z) / \log(z_s))$ where z_s is the magnification factor of each step. Therefore, we can exploit patch similarity across scales to reconstruct the HR image I_p . For a given patch Ω_s of the input image I_0 , the approximate nearest neighbor (ANN) method is used to search most similar patches in the database B from the LR images I_n ($n = -1, \dots, -N$). Assume that Q_i is the similar patch of Ω_s , Γ_i and Π_i denote its corresponding regions of larger size in the given image I_0 and the HR image I_1 , respectively. Note that Q_i and Ω_s are very similar, but they are not completely the same. This relationship between them is also applied to Γ_i and Π_i . The overlapped HR patches Π_i and their weights $\exp(-\|\Omega_s - Q_i\|^2 / \sigma^2)$ are gathered together to reconstruct the HR image I'_1 , where σ is a control parameter for similarity measure. Since there may be some uncovered regions in the initial reconstructed HR image, the back projection algorithm is used to improve image resolution. Then the refined reconstructed HR image I_1 is added to the patch pair database B . These steps mentioned above are implemented iteratively until the final HR image I'_P is obtained. After P magnification steps, the HR image I'_P is downsampled to get the $\times z$ HR image I_P by the bicubic interpolation. Next the HR image I_P is used as the initialized value of image super-resolution reconstruction.

Mathematically, our proposed sparse coding model of single image super-resolution is given as follows:

$$\alpha_y = \arg \min_{\alpha, F} \left\{ \|y - \mathbf{D}\mathbf{H}\Phi \circ \alpha\|_2^2 + \lambda \sum_i \|\alpha_i - \beta_i\|_1 \right. \\ \left. + \mu \|F(\nabla \mathbf{x}) - \nabla \mathbf{x}\|^2 \right\},$$

$$\text{s.t. } \mathbf{h}_F = \mathbf{h}_r, \quad (3)$$

where λ and μ 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 $|F(\nabla \mathbf{x})|$. Note that F is an odd function that is monotonically non-descending in the domain $(0, +\infty)$. On the right side of (3), the first term is the data fidelity of the solution, the second term is the sparse nonlocal regularization [4] and the third term is the gradient regularization [12]. Considering the natural images often contain repetitive patterns, 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 [6]. For the current estimate $\hat{\mathbf{x}}$, the similar patches of $\hat{\mathbf{x}}_i$ are denoted by $\hat{\mathbf{x}}_i^c$, whose coding coefficients are α_i^c . Then β_i can be computed as the weighted average of the sparse codes of the associated nonlocal similar patches:

$$\beta_i = \sum_c \mathbf{w}_i^c \alpha_i^c, \quad (4)$$

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

$$\mathbf{w}_i^c = \frac{1}{W} \exp\left(-\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_i^c\|_2^2 / \tau\right), \quad (5)$$

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

2.2. Iterative Solution

In our algorithm, we proposed an alternating minimization method to solve the image SR problem in (3) so that the constrained optimization is carried out with some variables fixed in cyclical fashion. First, the noise level σ_v is also estimated from the selected patches without high frequency from the LR image \mathbf{y} by the PCA technique [7, 10, 11]. Next, the initial value I_P of the HR image is acquired from the input LR image \mathbf{y} by the multistep magnification method that exploits structural self-similarities in both scale and space domains. Then for the current estimation I_P 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 PCA sub-dictionary is trained. Subsequently, for each patch, 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 SR problem in (3) is reduced to the sub-problem of gradient histogram preservation:

$$\min_F \|F(\nabla \mathbf{x}) - \nabla \mathbf{x}\|_2^2, \text{ s.t. } \mathbf{h}_F = \mathbf{h}_r. \quad (6)$$

Thus according to the estimated noise deviation σ_v , we can compute the reference histogram of image gradients and update the transform function F by solving the reduced sub-problem in (6). After that, for the fixed Φ and F , the image SR problem in (3) is reduced to the sub-problem in the following form:

$$\arg \min_{\alpha} \left\{ \begin{array}{l} \|\mathbf{y} - \mathbf{DH}\Phi \circ \alpha\|_2^2 + \lambda \sum_i \|\alpha_i - \beta_i\|_1 \\ + \mu \|F(\nabla \mathbf{x}) - \nabla \mathbf{x}\|_2^2 \end{array} \right\}. \quad (7)$$

where λ is a constant to weight the l_1 -norm sparsity regularization. To improve the reconstruction of sparse signals, we adopt an adaptively reweighting method [4] that exploits the image nonlocal redundancy to estimate the parameter λ . To solve the convex minimization sub-problem in (7), we first update the HR image \mathbf{x} by the gradient descent method:

$$\hat{\mathbf{x}}^{(t+1/2)} = \hat{\mathbf{x}}^{(t)} + \delta \begin{pmatrix} (\mathbf{DH})^T (\mathbf{y} - \mathbf{DH}\hat{\mathbf{x}}^{(t)}) \\ + \mu \nabla^T (\mathbf{f} - \nabla \hat{\mathbf{x}}^{(t)}) \end{pmatrix}, \quad (8)$$

where δ is a constant. The update process of HR image in (8) can be divided into two stages: the gradient regularization and the fidelity constraint. Assume that \mathbf{R}_i denotes the matrix

extracting the patch \mathbf{x}_i from the image \mathbf{x} at the location i . The sparse coding coefficients α_i are updated as follows:

$$\alpha_i^{(t+1/2)} = \Phi_k^T \mathbf{R}_i \hat{\mathbf{x}}^{(t+1/2)}, \quad (9)$$

where Φ_k , $k = 1, \dots, K$ is the PCA sub-dictionary of which cluster the patch $\hat{\mathbf{x}}_i$ falls into. The nonlocal means β_i of α_i can be estimated by using (4). By employing the iterative shrinkage operator [3] applied to each element of α_i , we can further update the coding coefficients α_i :

$$\alpha_i^{(t+1)} = S_{\lambda/c} \left(\Phi^T \circ (\mathbf{DH})^T \Theta / c + \alpha_i^{(t+1/2)} - \beta_i \right) + \beta_i, \quad (10)$$

where $\Theta = \mathbf{y} - \mathbf{DH}\Phi \circ \alpha_i^{(t+1/2)}$, $S_{\lambda/c}$ 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 follows:

$$\begin{aligned} \hat{\mathbf{x}}^{(t+1)} &= \Phi^{(t+1)} \circ \alpha^{(t+1)} \\ &= \left(\sum_{i=1}^l \mathbf{R}_i^T \mathbf{R}_i \right)^{-1} \sum_{i=1}^l \left(\mathbf{R}_i^T \Phi_k^{(t+1)} \alpha_i^{(t+1)} \right). \end{aligned} \quad (11)$$

The above iterative procedures are executed repeatedly until the convergence is achieved [1]. To give further clarification of the specific implementation of our proposed SPSR algorithm, it is summarized in **Algorithm 1**.

Algorithm 1 Pseudocodes of SPSR-Based Super-Resolution

Input: a LR image \mathbf{y} and a total scaling factor z .

Output: a HR image \mathbf{x}_H .

I. Initialization

- Set the initial parameters λ , μ , δ and c ;
- For the LR image \mathbf{y} , the HR image is initialized as I_P for sparse modeling by the multistep magnification;

II. Outer loop: for each iteration $t = 1$ to T do

- Update $\{\Phi_k\}$ by k-means clustering and PCA;
 - Update $\hat{\mathbf{x}}^{(t)}$ by the gradient regularization;
 - Inner loop: for each iteration $j = 1$ to J do
 - 1) Compute $\hat{\mathbf{x}}^{(t+1/2)}$ by the data fidelity;
 - 2) Compute $\alpha_i^{(t+1/2)} = \Phi_k^T \mathbf{R}_i \hat{\mathbf{x}}^{(t+1/2)}$, where Φ_k is the dictionary assigned to $\hat{\mathbf{x}}_i^{(t+1/2)}$;
 - 3) Compute λ and β_i of $\alpha_i^{(t+1/2)}$;
 - 4) Update $\alpha_i^{(t+1)}$ again by using (10);
 - 5) Reconstruct the estimate $\hat{\mathbf{x}}^{(t+1)}$ using (11);
 - Update the HR image $\mathbf{x}_H = \hat{\mathbf{x}}^{(t+1)}$.
-

3. EXPERIMENTAL RESULTS

The basic parameters of our SPSR algorithm are set as follows: the patch size is 6×6 with the overlap width equal to 4 between the adjacent patches, $K = 64$, $T = 9$, $J = 80$, $\mu = 0.005$ and $\delta = 7$. The proposed SPSR algorithm was compared with the state-of-the-art methods published recently [4] for verifying its validity subjectively and objectively. Just like the other methods [4], the simulated LR image is generated by the blurring and downsampling operator. That is, a HR image is first blurred with a 7×7 Gaussian kernel with standard deviation 1.6, and then downsampled by a scaling factor in both horizontal and vertical directions. In practice, the acquired LR image is often corrupted by noise that makes the super-resolution more difficult. Thus, it is also necessary to verify the robustness of the super-resolution methods to noise. The Gaussian noise with standard deviation 5 was added to the clean LR images to produce the simulated noisy LR images. For a set of test images, the PSNR results of our SPSR algorithm, the bicubic interpolation method and the NCSR method [4] on the simulated noiseless or noisy LR images are shown in Table 1, respectively. As can be seen from Table 1, it is found that the average PSNR gains of our SPSR algorithm over the existing best method, *i.e.*, NCSR [4], are 0.14 and 0.10 dB for the noiseless and noisy cases, respectively. The detailed reconstructed HR results of the different methods for the test images are shown in Figure 2 to 3. The visual comparisons demonstrate that the proposed SPSR algorithm have better recovery results of fine structures and sharp edges than them [4]. As seen from the experimental results, the proposed SPSR algorithm works well for a wide variety of images, and can reach better super-resolution results than the state-of-the-art methods.

4. CONCLUSIONS

In this paper, we addressed the single image super-resolution problem. The multi-scale spatial structural self-similarities are studied and then incorporated in our proposed SPSR framework. The initial estimation of HR image is obtained from an input LR image by the multistep magnification technique. The final HR image is reconstructed by building the sparse representation model with regularization constraints of both the nonlocally centralized sparsity and the gradient histogram preservation. Experimental results demonstrated that our SPSR algorithm is promising and competitive to the state of the arts, and outperforms other leading SR methods both visually and quantitatively.

5. REFERENCES

- [1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality," *Mathematics of Operations Research*, vol.35, no.2, pp.438-457, 2010.

Table 1: PSNR(dB) results on the reconstructed HR images with the total scaling factor 3 for LR images without / with Gaussian noise.

Images	Bicubic	NCSR [4]	SPSR
<i>Bike</i>	20.74/20.68	24.72/23.78	24.78/23.89
<i>Butterfly</i>	20.72/20.69	28.10/26.82	28.34/27.08
<i>Flower</i>	24.73/24.62	29.50/28.08	29.59/28.16
<i>Hat</i>	27.13/26.94	31.26/29.97	31.40/30.04
<i>Leaves</i>	19.80/19.76	27.45/26.25	27.73/26.47
<i>Lena</i>	27.99/27.77	34.28/32.42	34.36/32.49
<i>Parrot</i>	25.36/25.26	30.51/29.51	30.64/29.52
<i>Plants</i>	27.81/27.59	34.04/31.78	34.11/31.80
<i>Average</i>	24.29/24.16	29.98/28.58	30.12/28.68

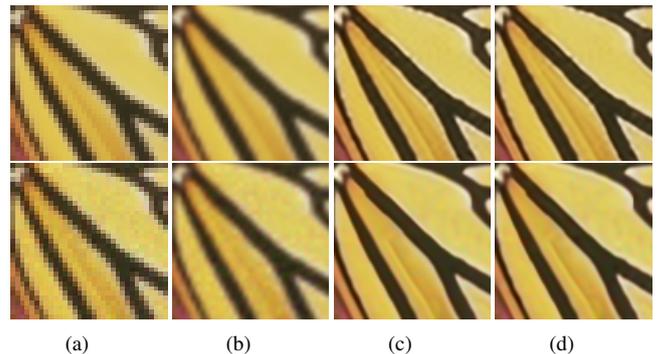


Fig. 2: Visual comparisons of the SR results of the different methods for the *Butterfly* image (the total scaling factor 3, top row: $\sigma_v = 0$, bottom row: $\sigma_v = 5$). From left to right: (a) LR input, (b) bicubic interpolation (top: SSIM=0.7199; bottom: SSIM=0.7039), (c) NCSR [4] (top: SSIM=0.9161; bottom: SSIM=0.8879), and (d) our SPSR algorithm (top: SSIM=0.9202; bottom: SSIM=0.8899).

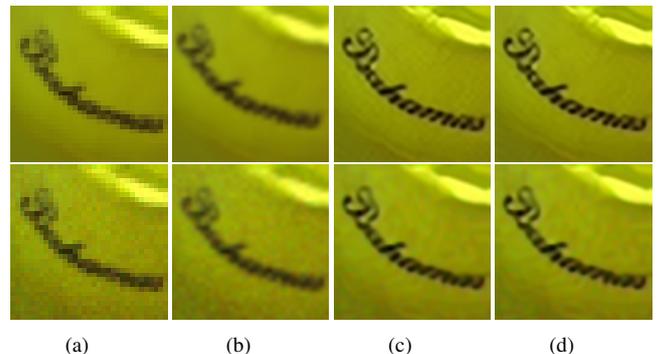


Fig. 3: Visual comparisons of the SR results of the different methods for the *Hat* image (the total scaling factor 3, top row: $\sigma_v = 0$, bottom row: $\sigma_v = 5$). From left to right: (a) LR input, (b) bicubic interpolation (top: SSIM=0.7807; bottom: SSIM=0.7508), (c) NCSR [4] (top: SSIM=0.8702; bottom: SSIM=0.8247), and (d) our SPSR algorithm (top: SSIM=0.8743; bottom: SSIM=0.8234).

- [2] J. F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition-modeling, algorithms and parameter selection," *International Journal of Computer Vision*, vol.67, no.1, pp.111-136, Feb 2006.
- [3] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol.57, no.11, pp.143-1457, Nov 2004.
- [4] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol.22, no.4, pp. 1620-1630, 2013.
- [5] Z. Gan, Z. Cui, C. Chen, and X. Zhu, "Adaptive joint nonlocal means denoising back projection for image super resolution," *IEEE International Conference on Image Processing (ICIP)*, Melbourne, Australia, pp.630-634, September 15-18, 2013.
- [6] D. Glasner, S. Bagon and M. Irani, "Super-resolution from a single image," *IEEE International Conference on Computer Vision (ICCV 2009)*, Kyoto, Japan, pp.349-356, 29 Sept - 2 Oct 2009.
- [7] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Transactions on Image Processing*, vol.22, no.12, pp.5226-5237, 2013.
- [8] Y. Wang, W. Wang, S. Cao, S. Li, L. Xie, and B. Ding, "Self-Similarity Superresolution for Resource-Constrained Image Sensor Node in Wireless Sensor Networks," *Mathematical Problems in Engineering*, vol.2014, Article ID 719408, 10 pages, 2014, <http://dx.doi.org/10.1155/2014/719408>.
- [9] J. Yang, J. Wright, T. S. Huang and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol.19, no.11, pp.2861-2873, Nov 2010.
- [10] Y. Q. Zhang, Y. Ding, J. Liu, and Z. Guo, "Guided image filtering using signal subspace projection," *IET Image Processing*, vol.7, no.3, pp.270-279, Apr 2013.
- [11] Y. Zhang, J. Liu, M. Li, and Z. Guo, "Joint image denoising using adaptive principal component analysis and self-similarity," *Information Sciences*, vol.259, pp.128-141, 2014.
- [12] W. Zuo, L. Zhang, C. Song, and D. Zhang, "Texture Enhanced Image Denoising via Gradient Histogram Preservation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1203-1210, Portland, OR, 23-28 Jun 2013.