JOINT SUB-BAND BASED NEIGHBOR EMBEDDING FOR IMAGE SUPER-RESOLUTION

Sijie Song[1], *Yanghao Li*[1], *Jiaying Liu*[1*], *Zongming Guo*[1,2]

[1]Institute of Computer Science and Technology, Peking University, Beijing, China
[2]Cooperative Medianet Innovation Center, Shanghai, China

ABSTRACT

In this paper, we propose a novel neighbor embedding method based on joint sub-bands for image super-resolution. Rather than directly reconstructing the total spatial variations of the input image, we restore each frequency component separately. The input LR image is decomposed into sub-bands defined by steerable filters to capture structural details on different directional frequency components. Then the neighbor embedding principle is employed to reconstruct each band, respectively. Moreover, taken the diverse characteristics of each band into account, we adopt adaptive similarity criterions for searching nearest neighbors. Finally, we recombine the generated HR sub-bands by applying the inverting sub-band decomposition to get the final super-resolved result. Experimental results demonstrate the effectiveness of our method both in objective and subjective qualities comparing with other state-of-the-art methods.

Index Terms— Image Super-Resolution (SR), Neighbor Embedding, Sub-Bands, Steerable Filters

1. INTRODUCTION

Image Super-Resolution (SR) reconstruction refers to generating the high-resolution (HR) output with inspiring visual qualities from a degraded low-resolution (LR) input image. The technique has been widely studied and used in many areas, ranging from medical image processing, image compression to satellite imaging. Nevertheless, due to the loss of information during the degradation, SR reconstruction which is an ill-posed problem, is still a challenging work.

SR reconstruction algorithms can be roughly classified into three categories: interpolation-based, reconstruction-based and learning-based. Interpolation-based methods estimate missing pixels according to the known pixels with linear or non-linear interpolation algorithms, such as new edge-directed interpolation (NEDI) [1] and soft-decision adaptive interpolation (SAI) [2]. The reconstruction-based methods [3, 4] adopt a maximum posterior probability (MAP) framework. Various regularization terms are imposed as the prior knowledge to describe the properties of natural images. In these methods, the prior information in prediction functions are designed by human and hard to model the diversified patterns in natural images.

Learning-based methods rely on large external datasets, attempting to capture the relationship between LR patches and their HR counterparts. Different models have been proposed to deal with the relationship. Some methods are under Markov Random Field (MRF) framework. Each LR patch has several HR candidates, which can be regarded as MRF framework and solved through graph cuts or belief propagation. But these methods with high computational complexity are time-consuming. To overcome the problem, Yang *et al.* [5] proposed sparse representation-based SR, referring that a patch can be estimated as a linear combination of a few pre-specified atom patches with few of linear coefficients being nonzeros. Moreover, Chang *et al.* [6] introduced neighbor embedding, assuming that the LR and HR patches form manifolds with similar local geometry. This algorithm is easy to understand and operate, which draws much attention.

However, structural features of the image may reflect on different directions and frequency components, especially when there are rich textures. Recovering the whole spatial variations directly as the SR methods described above always leads to the textural details smoothed out. Thus, we propose a novel neighbor embedding method based on joint sub-bands for image super-resolution. To describe features of the texture on each frequency band, we first decompose the image into several directional sub-bands filtered by a bank of orientation selective band-pass filters. In the meantime, a high-pass image and a low-pass residue are yielded. We adopt the neighbor embedding method to reconstruct each sub-band independently, as well as the high-pass and low-pass image with adaptive similarity criterions. The final reconstructed image is generated by combining all bands together through inverting sub-band decomposition.

The rest of this paper is organized as follows. The overview of neighbor embedding algorithm is described in Sec.2. In Sec.3, we have demonstrated the joint sub-band based neighbor embedding method. Experimental results are given in Sec.4. And concluding remarks are given in Sec.5.

* Corresponding author

This work was supported by National Natural Science Foundation of China under contract No.61472011, National High-tech Technology R&D Program (863 Program) of China under Grant 2014AA015205 and Beijing Natural Science Foundation under contract No.4142021.

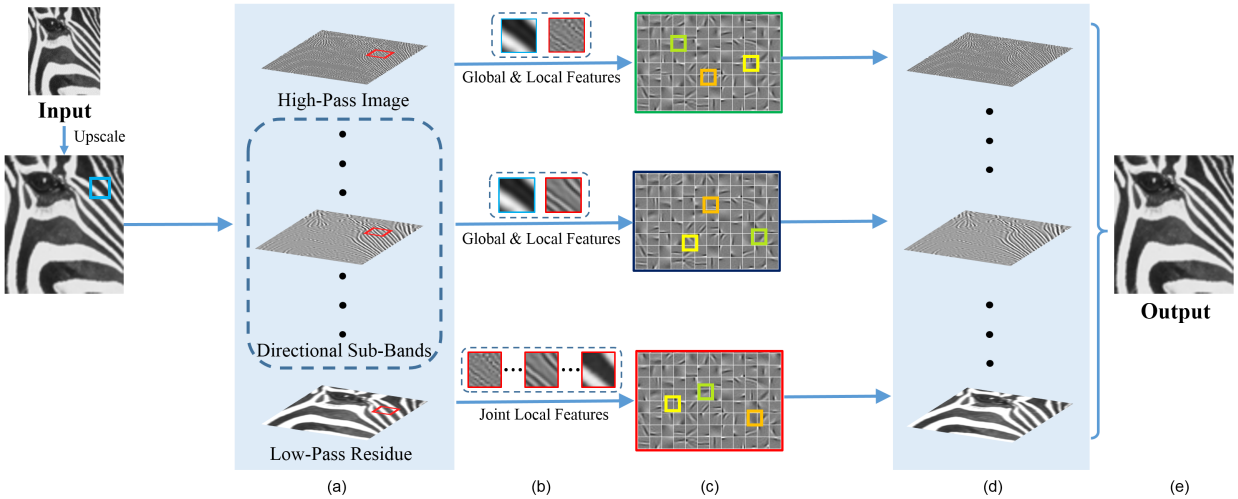


Fig. 1. The framework of joint sub-band based neighbor embedding for image super-resolution. (a) Different frequency components obtained from steerable pyramid transform. (b) References for similar patches considering global and local features. (c) Nearest neighbors from external library. (d) Reconstruction based on neighborhood regression for each frequency component. (e) Super-resolved image through inverse steerable pyramid transform.

2. OVERVIEW OF NEIGHBOR EMBEDDING

Neighbor embedding (NE) methods have shown good performance on image super-resolution reconstruction. The traditional neighbor embedding proposed by [6], locally linearly embedding (LLE), is based on the assumption that LR and HR patches share the similar local geometry and neighborhood relationship. This approach usually reconstructs the images using coupled dictionaries. An input LR image is typically separated into overlapping patches. For each LR patch, its local geometry is characterized by how a feature vector can be linearly represented by its similar patches in the feature space. And the goal is to reconstruct its HR counterpart as a weighted average of neighbors from the HR dictionary, using the same coefficients estimated in the LR space. Then, the target HR image is restored by integrating the HR patches according to their positions and averaged wherever they overlap.

3. JOINT SUB-BAND BASED NEIGHBOR EMBEDDING

In this section, we explain our proposed sub-band based neighbor embedding method in details. The framework of this method can be viewed in Fig.1.

3.1. Image Decomposition with Steerable Filters

The self-inverting and multi-orientation steerable pyramid transform [7, 8] at one scale is first employed to extract different frequency components from the input LR image X_t. By computing the response of a set of steerable filters, we can obtain the direction selective sub-bands $\{X_t^i\}_{i=1}^N$ with N orientations, as well as the high-pass image X_t^0 and the residual low-pass information X_t^{N+1}, which can be formulated as:

$$X_t^i = \begin{cases} \mathcal{F}^{-1}\left(\mathcal{F}\left(X_t\right) f\left(\theta^i\right)\right) & i = 1, ..., N \\ \mathcal{F}^{-1}\left(\mathcal{F}\left(X_t\right) g_i\right) & i = 0 \; or \; N+1 \end{cases}, \quad (1)$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ represent FFT and inverse FFT, $f\left(\theta^i\right)$ denotes the directional bandpass filter oriented at θ^i, and g_i is the high-pass or low-pass filter which can be calculated according to the frequency of bandpass filters. An example of the decomposition result can be viewed in Fig.2.

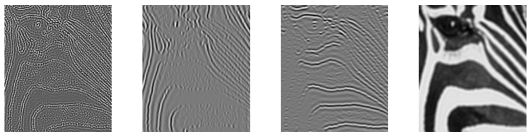


Fig. 2. Representations of steerable pyramid transform with one scale and two orientations. From left to right: high-pass image, two directional sub-bands (oriented at $0°$, $90°$), and low-pass residue.

Inspired by [9], the motivations we decompose the input image in frequency domain are twofold. (i) Structural patterns like edges are usually more prominent in one directional sub-band. Decomposing the image conveys such a property explicitly so that we are able to recover more details on this sub-band and get sharper edges in the final result. (ii) Richer textures can be synthesised because each frequency component is recovered independently. The combination of them can provide textures not existing in the training set.

3.2. Neighbor Regression for Each Frequency Component

Each frequency band of the input LR image is reconstructed independently to generate the corresponding HR bands

$\{Y_t^i\}_{i=0}^{N+1}$. We can perform the decomposition on the training images with steerable pyramid transform described in Sec.3.1, and yield coupled dictionaries on each frequency component. To formulate this problem, let $\mathcal{X}^i = \{x_s^{ij}\}_{j=1}^{\mathcal{N}}$ and $\mathcal{Y}^i = \{y_s^{ij}\}_{j=1}^{\mathcal{N}}$ be the LR and HR patch dictionaries of the band with subscript i, respectively. $\mathcal{N}$ is the dictionary size. After separating the band X_t^i of input LR image X_t into small patches, for each LR patch x_t^i, we can obtain its K nearest neighbors N_l^i in the training set $\mathcal{X}^i$ through K-nearest neighbor (K-NN). Then K appropriate weights are estimated to represent LR patch x_t^i by solving the least squares problem and apply these weights to the HR domain to yield its HR counterpart. To obtain better estimations more efficiently, we employ Ridge Regression to regularize the problem by l_2-norm coefficients, which can be formulated by:

$$\min_{\alpha_t^i} \|x_t^i - N_l^i \alpha_t^i\|_2^2 + \lambda \|\alpha_t^i\|_2^2, \tag{2}$$

where λ is the regularization term coefficient. Then α_t^i can be solved as:

$$\alpha_t^i = \left({N_l^i}^T N_l^i + \lambda I \right)^{-1} {N_l^i}^T x_t^i. \tag{3}$$

The corresponding HR patch y_t^i is given by applying the same reconstruction weights to corresponding neighbor HR patches N_h^i in the HR domain as follows:

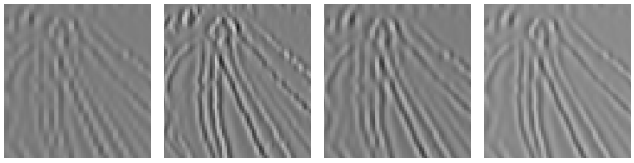
$$y_t^i = N_h^i \alpha_t^i. \tag{4}$$

The HR band Y_t^i is reconstructed by integrating reconstructed HR patches. The overlap portions of patches are averaged among different patches. However, the qualities of reconstructed patches rely heavily on their nearest neighbors. It is important to formulate good similarity criterions when performing retrieval algorithms, which we will discuss in the following subsection.

3.3. Similarity Metrics with Global and Local Features

After decomposition, the features of complex textures reflect on each frequency band, some of which may become sharper on one directional band while simpler on other bands. As shown in Fig.3, the pattern on the directional band of the bicubic interpolated image is cracked. However, it is continuous on the corresponding band of its ground truth. Obviously, we can not reconstruct continuous stripe of this sub-band if the broken pattern is the only reference for similar patches during the reconstruction. Thus, we consider joint features for similarity metrics.

For a patch $x_t^i (i = 0, ..., N)$ from the high-pass image or sub-bands, it is not accurate enough to find its similar patches in the training set only in accordance with its local features as shown in Fig.3(c). Thus, for the high-pass image and sub-bands, we also introduce the global structural information from the bicubic interpolated image. Taken both the global and local features into consideration, we develop the distance function for the patch $x_t^i (i = 0, ..., N)$ in K-NN searching as follows:



(a) (b) (c) (d)

Fig. 3. (a) Sub-band of upscaling image. (b) Corresponding sub-band of ground truth. (c) Results with only local features referred for similarity. (d) Results with global and local features referred for similarity.

$$dis\left(x_t^i, x_s^{ij}\right) = \|\nabla x_t^i - \nabla x_s^{ij}\|_2^2 + \eta \|\nabla x_t - \nabla x_s^j\|_2^2, \tag{5}$$

where ∇ denotes the gradient operator, and x_t, x_s^j are spatial co-location patches from the upscaling images for x_t^i, x_s^{ij}, respectively. The first term represents the distance of local features between patch x_t^i and x_s^{ij}, while that of the global features is measured by the second term. In addition, η allows us to balance the contribution of these two terms.

The low-pass residue is more smooth due to its low frequency, which results in the difficulty extracting gradient features. But it is coherent and contains enough structural information. So for a patch x_t^{N+1} from the low-pass image, we need joint local features from its corresponding patches of other bands. The similarity metric on patch x_t^{N+1} can be given by:

$$dis\left(x_t^{N+1}, x_s^{(N+1)j}\right) = \sum_{i=0}^{N+1} \|\nabla x_t^i - \nabla x_s^{ij}\|_2^2. \tag{6}$$

In the end, to generate the super-resolved result Y_t, the HR bands $\{Y_t^i\}_{i=0}^{N+1}$ are combined through inverting the steerable pyramid decomposition, which can be solved by:

$$\begin{aligned} Y_t = &\mathcal{F}^{-1}\left(\sum_{i=1}^{N} \mathcal{F}\left(Y_t^i\right) f\left(\theta^i\right)\right) + \\ &\mathcal{F}^{-1}\left(\mathcal{F}\left(Y_t^0\right) g_0 + \mathcal{F}\left(Y_t^{N+1}\right) g_{N+1}\right). \end{aligned} \tag{7}$$

Besides, the nonlocal redundancy [10] is also employed to the generated HR image Y_t to enhance the final result. For each patch y_t of Y_t, we seek its similar patches y_t^l and constrain the prediction error to be minimum, which can be formulated as:

$$Y_t = \arg\min \sum_{y_t \in Y_t} \|y_t - \sum_{l=0}^{L-1} w^l y_t^l\|_2^2, \tag{8}$$

where the nonlocal weight w^l is defined in [10], depending on the distance between y_t^l and y_t. Moreover, the classical back projection constraint is also performed to confirm that the blurred and down-sampled version of Y_t matches the given LR image X_t.

4. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed method, we conduct experiments of $2\times$ on several test sets (Set5, Set14 and B100) used in the previous literature [11]. For a fair comparison, we also adopt the training set consisting of 91 images

Table 1. Average PSNR(dB) results by $2\times$ on three test sets

Test Set	Bicubic	ScSR[5]	ANR [11]	BPJDL [12]	SRCNN [13]	NE	Proposed	Gain vs. SRCNN
Set5	33.68	36.00	35.84	36.20	36.34	35.84	**36.59**	0.25
Set14	30.23	31.93	31.80	32.02	32.17	31.79	**32.33**	0.16
B100	29.56	30.92	30.82	31.00	31.14	30.76	**31.22**	0.08

from ScSR [5]. The LR input images are generated from the original HR images by bicubic downsampling with the scaling factor. In our experiment, we decompose the image into the high-pass image ($\mathcal{B}^0$), four directional sub-bands oriented at $0°$, $45°$, $90°$ and $135°$ ($\mathcal{B}^1$–$\mathcal{B}^4$), as well as the low-pass residue ($\mathcal{B}^5$). For all of the bands, the regularization parameter λ in Eq.(2) is set to be 0.15. For the reconstruction in terms of the high-pass image and sub-bands, the patch size is 5×5, and the parameter η in Eq.(5) is set to be 1. For the reconstruction of low-pass image, the patch size is 9×9.

To confirm the effectiveness of our method, we first compare the PSNR results of each generated band with the corresponding band extracted from the HR image yielded by our NE method without decomposition in Table 2. It is clear that through image decomposition, we can achieve better frequency components in all bands. And the final result has an improvement of 1.52 dB.

We also compare the proposed algorithm with different methods, including Bicubic, ScSR [5], ANR [11], BPJDL [12] and SRCNN [13]. Besides, to prove the influence of image decomposition, our NE algorithm without decomposition is also compared with the proposed method. Except NE, all the results are obtained from the original authors' codes.

The objective results are shown in Table 1. For color images, we only calculate PSNR for the illuminance channel. Our proposed method outperforms other state-of-the-art methods in all the test sets. The detailed PSNR results for each single image are released in our website[1].

Table 2. Comparison for each band of *butterfly*

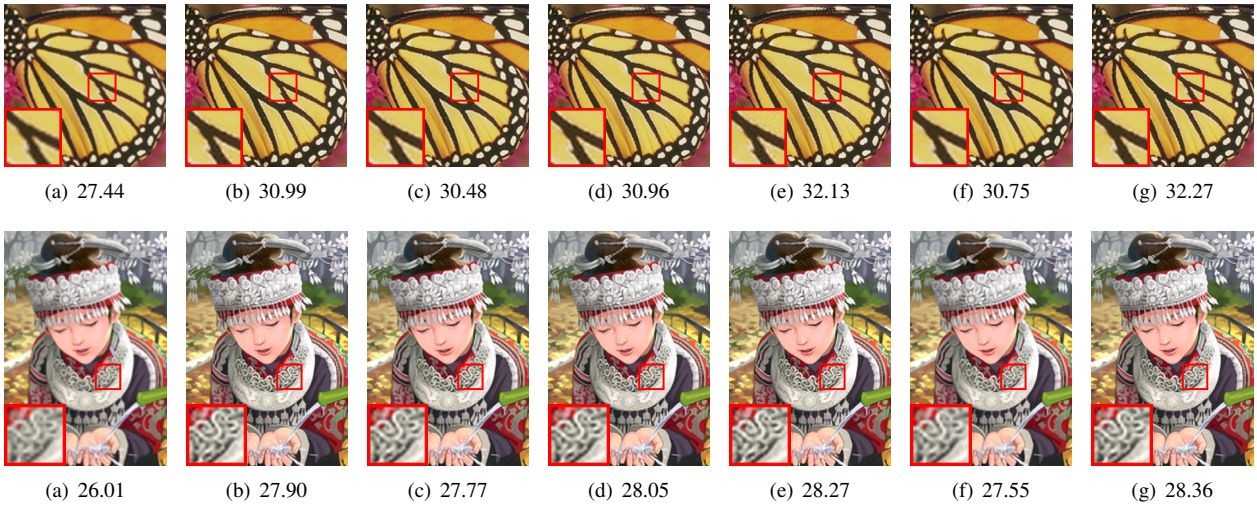
Frequency band	Bicubic	NE	Proposed
$\mathcal{B}^0$	32.74	34.55	**35.18**
$\mathcal{B}^1$ $(\theta = 0°)$	35.64	39.27	**40.63**
$\mathcal{B}^2$ $(\theta = 45°)$	37.06	39.79	**40.97**
$\mathcal{B}^3$ $(\theta = 90°)$	34.96	38.45	**39.42**
$\mathcal{B}^4$ $(\theta = 135°)$	34.47	39.49	**40.90**
$\mathcal{B}^5$	38.92	48.76	**49.28**
Combined result	27.44	30.75	**32.27**

Fig.4 presents some subjective results. Because the features on each band are specifically recovered, we get sharper edges and abundant textures while the results of other methods tend to be more blurred or unnatural. These observations illustrate that our method performs favorably against several state-of-the-art algorithms.

5. CONCLUSION

In this paper, we develop a joint sub-band based neighbor embedding method for image super-resolution. The image is processed by steerable pyramid decomposition to capture details on each frequency band. By reconstructing them independently, we can recover sharper edges and richer textures. Experimental results indicate the proposed method achieves better results in both quantitative and qualitative evaluations.

[1]http://www.icst.pku.edu.cn/course/icb/Projects/JSNE.html



(a) 27.44 (b) 30.99 (c) 30.48 (d) 30.96 (e) 32.13 (f) 30.75 (g) 32.27

(a) 26.01 (b) 27.90 (c) 27.77 (d) 28.05 (e) 28.27 (f) 27.55 (g) 28.36

Fig. 4. Comparison of PSNR(dB) results by $2\times$ on the (*butterfly*, *comic*) images: (a) Bicubic, (b) ScSR [5], (c) ANR [11], (d) BPJDL [12], (e) SRCNN [13], (f) NE, (g) Proposed. The red block with its corresponding magnification on the left-bottom corner of each image shows the reconstruction details.

6. REFERENCES

[1] X. Li and M.T. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1521–1527, 2001.

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

[3] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 372–379, 2000.

[4] Z. Lin and H.-Y. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 83–97, 2004.

[5] 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, 2010.

[6] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. I–I, 2004.

[7] E.P. Simoncelli and W.T. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *Proc. IEEE Int'l Conf. Image Processing (ICIP)*, vol. 3, pp. 444–447, 1995.

[8] W.T. Freeman and E.H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

[9] A. Singh and N. Ahuja, "Super-resolution using sub-band self-similarity," in *Proc. Asian Conference on Computer Vision (ACCV)*, vol. 9004, pp. 552–568, 2014.

[10] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 60–65, 2005.

[11] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, pp. 1920–1927, 2013.

[12] L. He, H. Qi, and R. Zaretzki, "Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 345–352, 2013.

[13] C. Dong, C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2015.