WAVELET-BASED SINGLE IMAGE SUPER-RESOLUTION WITH AN OVERALL ENHANCEMENT PROCEDURE

Zongqing Lu[†] *Quan Zou*[†] *Fei Zhou*[†] *Qingmin Liao*[†]

[†] Graduate School at Shenzhen, Department of Electronics Engineering
Tsinghua University, Shenzhen 518055, China

ABSTRACT

In this paper, we address the problem of generating a super-resolution image based on a dictionary of low- and high-resolution exemplars from a single input image in wavelet domain with a overall enhancement procedure. Most methods extract different kinds of features in low-resolution image and high-resolution images to establish the mapping relation. But in this paper, we implement wavelet-transform to extract the same kind of feature to make the mapping more reasonable. Meanwhile we implement local Lipschitz regularity constraint and structure-keeping constraint to preserve the local singularity and edge in our method. Compared with current state-of-art methods on standard images, our method obtains both visual and PSNR improvement.

***Index Terms*—** Image super-resolution, wavelet domain, Lipschitz regularization, structure-keeping

1. INTRODUCTION

Single image super-resolution(SR) aims at generating a high-resolution(HR) image from a low-resolution(LR) image. The core part of the SR methods is to maintain the high frequency infomation of the edge area of the image to make the reconstructed image sharper visually and better in performance. Currently SR methods can be roughly subdivided into three categories. Interpolation-based methods use the linear combination of nearby known pixels to obtain the unknown pixels like bicubic interpolation, or non-linear interpolation like NEDI(new edge directed interpolation) [1]. Reconstruction-based methods map the LR images to HR images using the known prior. Neighbor embedding method in [2] that implements the prior that the manifolds of LR and HR images are locally in similar geometries and LR/HR images can be linearly combined by the LR/HR neighbors. Learning-based methods use machine learning techniques that try to learn the mapping function or some relation between the LR images and HR images. Edge statistics are learned in [3] from natural images as gradient profile prior. A deep convolutional neural network is implemented in [4] to learn an end-to-end mapping between the LR and HR images. Sparse representation based methods [5] learned coupled LR and HR dictionaries to represent the mapping function based on sparse signal representation prior. K-SVD/OMP is implemented in the sparse representation process which obtained lower computational complexity and improved quality in [6]. The sparse representation methods and neighborhood embedding methods are combined in [7, 8]. Timofte et al[7] find the neighbors in the sparse representation dictionaries to represent LR/HR images and use Ridge Regression [9] to reformulate the problem as a least squares regression. Timofte et al [8] use the sparse representation dictionaries to find neighbors in the training data instead to represent the LR/HR images.

Currently, most SR methods [5, 6, 7, 8] used the first- and second- order gradients of patches as the features for LR images and subtracted the bicubically interpolated LR image from the HR image to create the features for HR images. We can see that the features for LR/HR images extracted from different ways so that we can't guarantee the structures of these features in high-dimensional manifold matching well. To solve this problem, we implement the wavelet transform to extract the high-frequency components both in LR and HR image.

Many single image SR methods [5, 6, 10, 11] implement back projection fidelity term to improve the initial results that are obtained from their raw methods. Wang et al[12] and Dong et al [13] implement nonlocal self-similarity which is proved indeed existing in natural images [14] to regularize the optimization problem in SR. Besides the back projection fidelity term and nonlocal self-similarity constraint term, in this paper we implement local Lipschitz regularity constraint and structure-keeping constraint to preserve the local singularity and edge in our method. We combine these four terms to an overall enhancement procedure that significantly improves the result compared with other SR methods in edge-full images.

In the following sections, we will first present the model of our proposed metthod in Section 2. Then we explain details of our proposed method in Section 3, and describe our experiments in Section 4 where we compare the performance of our method to other state-of-art methods. Finally in Section 5 we conclude the paper.

This work was supported by Shenzhen STP(JCYJ20150331151358150)

2. MODEL OF WAVELET-BASED SINGLE IMAGE SR METHODS

Our proposed method builds on theories from learning-based super-resolution methods. So in this section we briefly present the model of our method. And the training and testing phases is shown in 1.

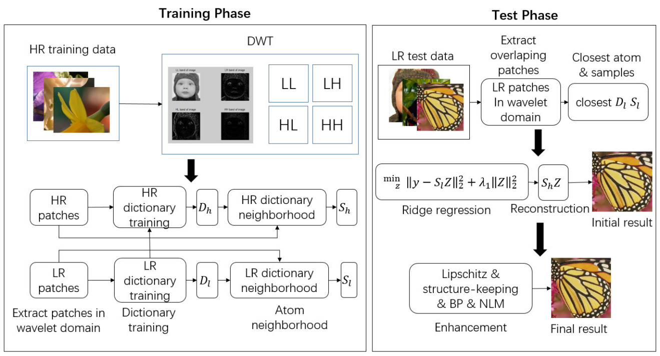

Fig. 1. Training Phase of our method

Firstly our method collects patches from the 91 training images[5] in 4 wavelet domains. Then it uses a sparsity constraint to jointly train the LR/HR dictionaries to represent the LR/HR patches[6] in each wavelet domians.

Then we use the local neighborhood samples S_l of each LR dictionary atom in each wavelet domain to represent patches with Ridge Regression. Formulated below:

$$\min_{\alpha} \|S_l \alpha - p_l\|_2^2 + \lambda \|\alpha\|_2 \tag{1}$$

where S_l contains the K training samples that lie closest to the dictionary atom to which the input patch p_l is matched, and K is a constant we need to set. The distance measure used for neighbor search in out method is the Euclidean distance.

After we get the reconstruction coefficient α, we use corresponding HR neighbor samples S_h and then reconstruct the initial HR image patch in wavelet domain. Then we averagely add these overlapping HR patches and implement inverse wavelet transform to get a initial HR image.

Then we implement back-projection fidelity term with Lipschitz regularity constraint, structure-keeping constraint and nonlocal self-similarity constraint to enhance the result.

3. MORE DETAILS OF PROPOSED METHOD

In this section, we present the details of the training phase and the ehancement phase.

3.1. Training

Currently many methods [5, 6, 7, 8] used the first- and second-order gradients of patches as the feature for LR images. But we notice that all these features are limited. These features can not represent the whole high frequencies details. Meanwhile wavelet transform is a perfect way to extract the whole local high frequencies details and our experimental results will illustrate it. We implement discrete wavelet transform to the training LR/HR images and we can get four wavelet domains(LL,LH,HL,HH) LR/HR images, then collect overlapping patches from them. Sparse dictionaries are learned independently in each wavelet domain for LR/HR images. Specifically we use K-SVD for the LR dictionaries in each wavelet domain and pseudo-inverse for the HR dictionaries in each wavelet domain, just like [6].

3.2. Enhancement

3.2.1. Lipschitz regularization

The local maxima of wavelet transform modulus capture the sharp variation pixels of an image and their evolution across scales characterizes the local Lipschitz regularity of the image. For example, left part of Fig.2 shows a two-dimensional image and its wavelet transform at several scales. And right part of Fig.2 shows the propagatation of extrema points across the scales in the $10^t h$ column of the image.

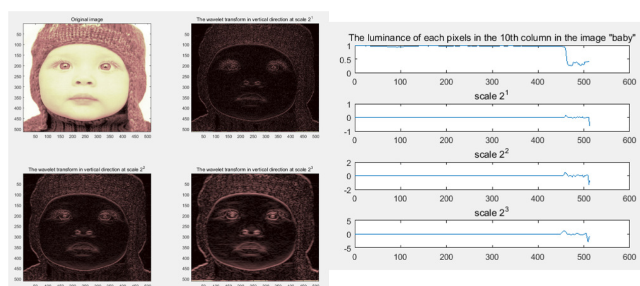

Fig. 2. Left: Pseudocolor image of $Baby$ and its LH components of 2-D wavelet transform in three scales. Right: Propagation of extrema points across the scales for 2-D waveform in the $10^t h$ column of the image $Baby$

The singularities in the signal induce peaks in the wavelet transform propagate across scales, and the values of the peaks corresponding to the same singularity change across the scales according to an exponential function. In particular, a function f is uniformly Lipschitz α (defined in [15]) over an interval (a, b) if and only if there exists a constant $K > 0$ such that for all $x \in (a, b)$, the wavelet transform of $f(x)$ satisfies

$$|W_s f(x)| <= K s^{\alpha} \tag{2}$$

The wavelet transform of f at scale s and position x, denoted by $W_s f(x)$. If $f(x)$ is differentiable but not continuously differentiable at x_0, then it is Lipschitz $\alpha = 1$ at x_0 and the corresponding wavelet transform maxima behave as $O(s)$

around x_0. If f is discontinuous but bounded in the neighborhood of x_0, then $\alpha = 0$ at x_0, and the corresponding maxima remain constant across the scales. And for Dirac function, $\alpha = -1$ at x_0 and the corresponding wavelet transform maxima behave as $O(1/s)$ around x_0.

For the local extremum in the wavelet domain of signals, we can rewrite (Equation2) in discrete formulation

$$W_{2^j} f[x_m^{(j)}] = K_m (2^j)^{\alpha_m}, j = 1, \ldots, J, \tag{3}$$

where $x_m^{(j)}$ is the location of the local extremum at scale 2^j corresponding to the m_{th} extremum, α_m is the Lipschitz regularity of f at the extremum point, and K_m is a nonzero constant. Then a LR image can be treated as a smoothed HR image. An unknown scaling filter is implemented on a HR image to generate the corresponding LR image. Then we denote LR image at scale 2^1 and the HR image that we want to be restored at scale 2^0. Then we can extrapolate every extremum point in the wavelet domain of HR image by scaling the LR image $J - 1$ times, the parameters α_m and K_m in (Equation3) can be estimated via linear regression on

$$\log_2(W_{2^j} f[x_m^{(j)}]) = \log_2(K_m) + j\alpha_m, j = 1, \ldots, J, \tag{4}$$

and all the extremum points in HR image denoted by E_0 can be estimated by $E_0 = W_{2^0} f[x_m^{(0)}] = K_m$.

For the image that we want to enhance, we implement a 2-D discrete wavelet transform to the LR image. By scaling the LR image several times, we can obtain the extremum points in the wavelet domain of the HR images that we want to reconstruct. We expect

$$\min_X \|E(X) - E_0\|_2^2, \tag{5}$$

to be small, where $E()$ is the operation to obtain the all extremum points in wavelet domain of current HR image. As the $E()$ operation is an unlinear operation, it can not be formulated into a matrix form. Instead, we change the loss function in (Equation5) to

$$\min_X \|X - IWA(E_0 + E^r)\|_2^2, \tag{6}$$

where X stands for the current HR image, and IWA is a inverse wavelet transform, E_0 is the extremum points that we obtained from the corresponding LR image before and E^r is the other part of the wavelet domain of HR image X. More details can be found in [15]. Then we can implement an iterative gradient desent method to solove this regularization term as below.

$$X_{t+1} = X_t - \rho(X_t - IWA(E_0 + E_t^r)), \tag{7}$$

where X_t is the estimate of the reconstruction result after t^{th} iteration, ρ is the step of the gradient descent.

3.2.2. *Structure-Keeping Constraint*

Normally, a LR image preserves the structure of the corresponding HR image quite well. Regular super resolution methods like interpolation usually blur the structure. To enhance the structures in our reconstruction results, we can use a roughly structure regularization to constrain our reconstruction results. Xu et al [16] proposed a relative total variation(RTV) to extract meaningful structures under complicate texture patterns. The relative total variation is

$$RTV = D_x(p_i)/(L_x(p_i) + \epsilon) + D_y(p_i)/(L_y(p_i) + \epsilon), \tag{8}$$

where $D_{x/y}(p_i)$ stands for the windowed total variation in x/y direction in the pixel p_i, $L_{x/y}(p_i)$ stands for the windowed inherent variation in x/y direction in the pixel p_i, and ϵ is a small positive number to avoid division by zero. more details in [16].

In an image, pixels with textures and strutures yield large D. But pixels with only textures are generally smaller than the pixels with textures and structures on the measure of L. It is shown that relative total variation(RTV) is simple and yet effective to make main strutures in an image stand out, which means that it can sharpen the structures area in a blurred image.

To preserve this structure in a reconstructed HR image, we add a structure-keeping term $\min_X \|X - X^s\|$ to our loss function, where X^s is the structure image of the X obtained from [16].

3.2.3. *Enhancement procedure*

Above all, the whole enhancement procedure is formulated as below

$$\begin{aligned} X^* = &\min_X \|SHX - Y\|_2^2 + a\|(I - W)X\|_2^2 \\ &+ b\|X - IWA(E_0 + E^r)\|_2^2 + c\|X - X^s\|_2^2. \end{aligned} \tag{9}$$

The first term denotes the back-projection fidelity term, Y denotes the corresponding LR image, S denotes a downsampling operator and H denotes a blurring filter. The second term denotes the nonlocal self-similarity regularization, W denotes nonlocal means similar weight matrix defined in [13] and I denotes identity matrix. The third term denotes the Lipschitz regularization, the fourth term denotes the structure-keeping constraint. Meanwhile a, b, c denote the regularization parameters. The solution to (Equation9) can be efficiently computed based on iterative optimization, as in Yang et al [5] and used in the back-projection, as formulated below.

$$\begin{aligned} X_{t+1} = &X_t + \nu H^T S^T (Y - SHX_t) - a(I - W)^T (I - W) X_t \\ &- b(X_t - IWA(E_0 + E_t^r)) + c(X_t^s - X_t) \end{aligned} \tag{10}$$

(a) Bicubic (b) ScSR[5] (c) ANR[7] (d) A+[8] (e) CSC[10] (f) Proposed (g) Ground truth

Fig. 3. Results by $3\times$ on $Butterfly$ image. The red box with its corresponding magnification on the left-bottom of each images shows the details

4. EXPERIMENTAL RESULTS

In this section, we analyze the performance of our method via the reconstruction precision quantity and visual quality compared with other state-of-art methods: SCSR by Yang et al [5], ANR by Timofe et al [7], A+ by Timofe et al [8], CSC by Gu et al [10]. We use Set5 that contains 5 images provided in [17] and Set14 that contains 14 images provided in [6] as test images. In our experiments, the patch size is 9×9. We extract patches from bicubically interpolated LR in wavelet domains to generate LR patches. Meanwhile we extract patches from these HR images in wavelet domains. We use 1024 dictionaries and neighborhood size is 2048. We set ν to 1.8, a to 0.09, b to 1, c to 0.025.

Table. 1 shows PSNR(Peak signal-to-noise ratio) comparation and table.2 shows SSIM(Structural SIMilarity) comparation, meanwhile examples are shown in Fig. 3. We can see that our proposed method outperforms the state-of-arts in these 19 testing images, and its PSNR is in average 0.15dB higher than CSC[10], which is the best among the other methods. Specifically. in Fig. 3, we can see that our proposed method can preserve edges better than the other state-of-art methods in visual quality. And Table.2 also illustrates the effectiveness of our method.

5. CONCLUSION

In this paper, we proposed a new wavelet-based single image super-resolution method. Our contributions are: We extract high-frequency components separately in four wavelet domains for both LR/HR images, which guarantee the features for LR/HR images forming the same structure in the high-dimensional manifold; We constraint the enhancement with local Lipschitz regularity, which is bonus for us to extract the features in wavelet domains; And we also constraint the enhancement with structure image, which can preserve the edge quite well. With above all, our proposed method achieves better results in both reconstruction precision and visual quality.

However our proposed method is not that fast as other state-of-art methods because regularizing with nonlocal self-similarity and structure images significantly increases the complexity of our proposed method. The major of our future investigation is to reduce the computing complexity.

Table 1. PSNR results on image super-resolution with other methods in Set5 and Set14 (scaling factor = 3)

Images	Bicubic	ScSR[5]	ANR[7]	A+[8]	CSC[10]	Proposed
Baby	33.9	34.3	35.1	35.2	**35.3**	**35.3**
Bird	32.6	34.1	34.6	35.5	**35.8**	35.6
Butterfly	24.0	25.6	25.9	27.2	27.1	**28.2**
Head	32.9	33.2	33.6	33.8	**33.8**	**33.8**
Woman	28.6	29.9	30.3	31.2	31.2	**31.5**
Baboon	23.2	23.5	**23.6**	**23.6**	**23.6**	**23.6**
Barbara	26.2	26.4	**26.7**	26.5	**26.7**	26.4
Bridge	24.4	24.8	25.0	25.2	25.2	**25.3**
Coastguard	26.6	27.0	27.1	**27.3**	**27.3**	**27.3**
Comic	23.1	23.9	24.0	24.4	24.4	**24.6**
Face	32.8	33.1	33.6	**33.8**	**33.8**	33.7
Flowers	27.2	28.2	28.5	29.0	29.0	**29.2**
Foreman	31.2	32.0	33.2	34.3	34.2	**34.4**
Lenna	31.7	32.6	33.1	33.5	33.6	**33.7**
Man	27.0	27.8	27.9	28.3	28.3	**28.4**
Monarch	29.4	30.7	31.1	32.1	32.1	**32.9**
Pepper	32.4	33.3	33.8	**34.7**	**34.7**	34.5
Ppt3	23.7	25.0	25.0	26.1	25.9	**26.2**
Zebra	26.6	28.0	28.4	29.0	29.2	**29.3**
Average	28.29	29.13	29.5	30.04	30.06	**30.21**

Table 2. SSIM results on image super-resolution with other methods in Set5 and Set14 (scaling factor = 3)

Images	Bicubic	ScSR[5]	ANR[7]	A+[8]	CSC[10]	Proposed
Baby	0.9039	0.9046	0.9225	0.9233	**0.9245**	0.9239
Bird	0.9256	0.9398	0.949	0.956	**0.958**	0.9562
Butterfly	0.8215	0.8622	0.872	0.9091	0.9064	**0.9188**
Head	0.8007	0.8036	0.8249	0.8281	**0.8298**	0.8264
Woman	0.8896	0.9044	0.917	0.9288	0.929	**0.9291**
Baboon	0.5439	0.5879	0.5991	0.6064	**0.6092**	0.6059
Barbara	0.7531	0.7633	0.7811	0.7795	**0.7855**	0.7741
Bridge	0.6483	0.6688	0.676	0.684	**0.7139**	0.7112
Coastguard	0.6147	0.6392	0.6575	0.6621	0.6626	**0.6631**
Comic	0.699	0.7571	0.7617	0.7798	0.7805	**0.7909**
Face	0.7984	0.8012	0.8234	0.8271	**0.8283**	0.8257
Flowers	0.8013	0.8301	0.8405	0.8524	**0.8538**	0.8533
Foreman	0.906	0.9133	0.9302	0.94	0.9405	**0.9418**
Lenna	0.8582	0.865	0.8805	0.8851	**0.8864**	0.8848
Man	0.7495	0.776	0.79	0.8	**0.8021**	0.8011
Monarch	0.9198	0.9292	0.9377	0.9471	0.947	**0.9503**
Pepper	0.8698	0.8676	0.8856	0.8921	**0.8923**	0.8907
Ppt3	0.8746	0.906	0.9127	0.9378	0.9305	**0.94**
Zebra	0.7943	0.8298	0.8449	0.8508	**0.8531**	0.8528
Average	0.7985	0.8184	0.8319	0.8416	0.8439	**0.8442**

6. REFERENCES

[1] Xin Li and Michael T Orchard, "New edge-directed interpolation," *Image Processing, IEEE Transactions on*,

vol. 10, no. 10, pp. 1521–1527, 2001.

[2] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. IEEE, 2004, vol. 1, pp. I–I.

[3] Jian Sun, Jian Sun, Zongben Xu, and Heung-Yeung Shum, "Image super-resolution using gradient profile prior," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[4] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[5] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma, "Image super-resolution via sparse representation," *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, 2010.

[6] Roman Zeyde, Michael Elad, and Matan Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, pp. 711–730. Springer, 2010.

[7] Radu Timofte, Vincent Smet, and Luc Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1920–1927.

[8] Radu Timofte, Vincent De Smet, and Luc Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Computer Vision–ACCV 2014*, pp. 111–126. Springer, 2014.

[9] Andreĭ Tikhonov, *Solutions of ill-posed problems*.

[10] Shuhang Gu, Wangmeng Zuo, Qi Xie, Deyu Meng, Xiangchu Feng, and Lei Zhang, "Convolutional sparse coding for image super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1823–1831.

[11] Yu Zhu, Yanning Zhang, Boyan Bonev, and Alan L Yuille, "Modeling deformable gradient compositions for single-image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5417–5425.

[12] Shenlong Wang, Lei Zhang, Yan Liang, and Quan Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2216–2223.

[13] Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *Image Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 1838–1857, 2011.

[14] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 2, pp. 60–65.

[15] S Grace Chang, Zoran Cvetkovic, and Martin Vetterli, "Locally adaptive wavelet-based image interpolation," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1471–1485, 2006.

[16] Li Xu, Qiong Yan, Yang Xia, and Jiaya Jia, "Structure extraction from texture via relative total variation," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, pp. 139, 2012.

[17] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.