Edge-Preserving Single Image Super-Resolution*

Qiang Zhou[1], Shifeng Chen[1], Jianzhuang Liu[1,2] and Xiaoou Tang[1,2]
[1]Shenzhen Key Laboratory for Computer Vision and Pattern Recognition
Shenzhen Institutes of Advanced Technology,Chinese Academy of Sciences, China
[2]Department of Information Engineering, The Chinese University of Hong Kong, China
{qiang.zhou,shifeng.chen}@siat.ac.cn, {jzliu,xtang}@ie.cuhk.edu.hk

ABSTRACT

This paper proposes a novel approach to single image super-resolution. First, an image up-sampling scheme is proposed which takes the advantages of both bilateral filtering and mean shift image segmentation. Then we use a shock filter to enhance strong edges in the initial up-sampling result and obtain an intermediate high-resolution image. Finally, we enforce a reconstruction constraint on the high-resolution image so that fine details can be inferred by back projection. Since strong edges in the intermediate result are enhanced, ringing artifacts can be suppressed in the back projection step. We compare our algorithm with several state-of-the-art image super-resolution algorithms. Qualitative and quantitative experimental results demonstrate that our approach performs the best.

Categories and Subject Descriptors

I.4.3 [**Image Processing and Computer Vision**]: Enhancement

General Terms

Algorithms, Experimentation

Keywords

Image Super-Resolution, Edge-Preserving

1. INTRODUCTION

The goal of single image super-resolution is to recover a high-resolution image from a low-resolution image. Image super-resolution is important in many applications of multimedia such as playing a video on a higher-resolution screen. Although much work has been available, super-resolution has not be solved very well yet. As pointed out in [2], the generation process of a low-resolution image can be modeled as smoothing and down-sampling a high-resolution image. Single image super-resolution is a challenging problem because there is inevitable information loss in down-sampling and the number of unknowns in the recovery exceeds the number of observed data.

Previous approaches to super-resolution can be categorized into three groups: interpolation based, reconstruction based, and learning based. The comparison of different approaches is summaried in Table 1. Interpolation based methods [1, 9] do not introduce high-frequency information. Therefore, their results often appear over-smooth and have jagged artifacts along the edges.

The reconstruction based approaches [4, 5, 14] are formulated under the framework of enforcing a reconstruction constraint and imposing prior knowledge on high-resolution images. The reconstruction constraint requires that the high-resolution image via smoothing and down-sampling should be as close as possible to the low-resolution image. In [13], Shan et al. utilize the heavy-tailed distribution prior of natural image gradient for image/video super-resolution. Different from above mehtods, statistical edge features between the low-resolution and high-resolution image are used in [5, 14]. In these methods, the imposed prior is not true for arbitrary images and many artifacts (e.g., ringing) may appear in the high-resolution image.

The category of learning based approaches [6, 15, 19, 18, 11, 21, 10],also known as image hallucination techniques, is originally proposed in [6] for learning high frequency details that do not appear in the low-resolution image from a number of training examples. The framework can be formulated by a Markov Random Field (MRF) and solved by belief propagation. The learning based methods are often time-consuming. To speed up their algorithm, Sun et al. [15] propose to learn only the primal sketch priors (e.g., edges, ridges, and corners) for image hallucination. In [21], sparse representation is incorporated into the learning step. Recently, Sun et al. [16] propose a context-constrained hallucination method which pays much attention to textural regions. Because of insufficient training examples, high frequency artifacts often appear in the results of a learning based approach.

Almost all image super-resolution methods [6, 15, 4, 13, 14, 21] need an initial high-resolution image for latter steps via up-sampling the input low-resolution image. Previous approaches often use interpolation methods (e.g., bicubic interpolation) to obtain the initial high resolution result. Such a simple method severely degrades the final result. In this paper, we propose a novel up-sampling method to ob-

*Area chair: Lei Chen

MM'11, November 28–December 1, 2011, Scotsdale, Arizona, USA.

Table 1: Comparison of different super-resolution approaches.

Categorization	Description	Disadvantages
Interpolation based	using different interpolation techniques	over-smooth, jagged artifacts
Reconstruction based	using reconstruction constraint and image prior	ringing artifacts, imposing additional prior
Learning based	learning high frequency details from training set	high frequency artifacts, relying training set

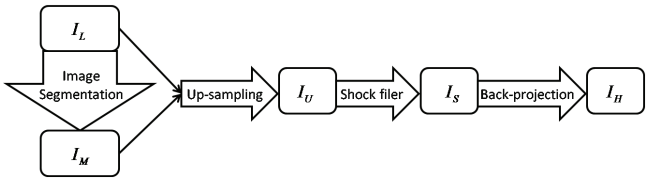


Figure 1: Framework of our approach.

tain the initial high-resolution result. Our scheme incorporates a soft-edge and a hard-edge constraints in up-sampling. The soft-edge constraint is enforced via bilateral filtering [17] which smoothes the image while preserving edges. The hard-edge constraint is enforced after applying the mean shift image segmentation algorithm [3].

After obtaining the initial result via up-sampling, we use the complex shock filter [7] to enhance strong edges in the high-resolution image, instead of imposing prior knowledge on the high-resolution image which is used in most previous methods [4, 13, 5, 14]. These methods with a prior have two main disadvantages. First, the prior is not suitable for all images. With it imposed, many artifacts may be introduced to the high-resolution image or the result appears over-smooth. Second, solving for a high-resolution image with a prior is computationally expensive.

Finally, we enforce a reconstruction constraint on the high-resolution image and the final result is solved by back projection [8]. The results obtained by the original back projection [8] often suffer from ringing artifacts. Due to the combination of up-sampling scheme and the complex shock filter, our algorithm obtains results without noticeable ringing artifacts.

Compared with previous work, our algorithm has two advantages. The first is that it works without any prior and thus can handle arbitrary images. The second is that it preserves edges excellently.

2. FORMULATION OF OUR APPROACH

Let I_L be the input low-resolution image, I_U be the initial high-resolution result obtained by our up-sampling method, I_S be the intermediate result after the shock filter, and I_H be the final high-resolution image. The framework of our algorithm is shown in Fig. 1.

2.1 Up-Sampling

In previous image up-sampling algorithms, the filter is constructed considering only spatial information with useful intensity information discarded. Motivated by bilateral filtering in [17], we propose a novel up-sampling scheme considering both spatial and intensity information. Bilateral filtering can be defined as

$$B(p) = \frac{1}{W(p)} \sum_{q \in \Omega} I(q) G_{\sigma_s}(\|p-q\|) G_{\sigma_r}(\|I(p)-I(q)\|), \quad (1)$$

where I is the input image, and Ω is the set of all pixels of

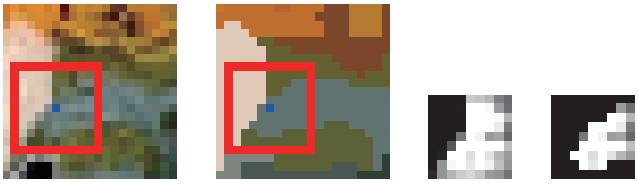


Figure 2: Illustration of the soft-edge and hard-edge constraints in our up-sampling scheme. (a) One patch of a low-resolution image (Fig. 4 (a)). The blue pixel is the current center and the red rectangle denotes its filtering neighborhood region. (b) The result of mean shift segmentation. (c) The bilateral filtering weight $BF(p_l, q)$ for the blue pixel. (d) The final up-sampling weight $BF(p_l, q)M(p_l, q)$ for the blue pixel. A larger intensity in (c) and (d) denotes a larger weight value. These values are scaled for visualization.

the image, p and q denote pixel locations over the image, and $W(p)$ is the normalization constant at p, and G_{σ_s} and G_{σ_r} are two Gaussian functions with standard deviations σ_s and σ_r, respectively.

The intuition of bilateral filtering is to smooth the image using pixels which are close both in the spatial domain and the intensity domain. We call this as soft-edge constraint. In our algorithm, to interpolate a target pixel, we choose its nearby pixel with similar colors. In order to preserve edges better, we also propose a hard-edge constraint. We first perform segmentation on the input low-resolution image by the mean shift algorithm [3]. Then for each pixel, we use only pixels in the same segment for up-sampling.

Let I_M be the result of mean shift image segmentation. By combining both the soft-edge constraint and the hard-edge constraint, our up-sampling is formulated as

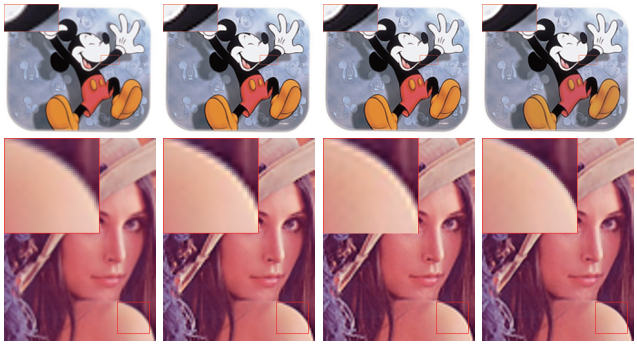
$$I_U(p_h) = \frac{1}{W(p_h)+\beta}\left(\sum_{q \in \Omega} BF(p_l, q) M(p_l, q) I_L(q) + \beta I_L(p_l)\right), \quad (2)$$

$$BF(p_l, q) = G_{\sigma_s}(\|p_l - q\|) G_{\sigma_r}(\|I_L(p_l) - I_L(q)\|), \quad (3)$$

$$M(p_l, q) = \begin{cases} 1 & \text{if } I_M(p_l) = I_M(q) \\ 0 & \text{if } I_M(p_l) \neq I_M(q) \end{cases}, \quad (4)$$

where β is a balance parameter, p_h is pixel position in the high-resolution image, and p_l is the position in the low-resolution image corresponding to p_h. If the coordinates of p_l are not integers, $I_L(p_l)$ and $I_M(p_l)$ are obtained by nearest interpolation on the low-resolution image. The term $\beta I_L(p_l)$ in (2) is used to emphasize the importance of the filter center p_l.

Fig. 2 gives an example to illustrate the soft-edge constraint and hard-edge constraint. From this example, we can see that the color of each pixel in the high-resolution



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

Figure 3: Comparison of different algorithms at 3× magnification on the "mickey" and "Lena" images. (a) Bicubic interpolation. (b) Sharpen bicubic (bicubic interpolation + unsharp mask in Adobe Photoshop). (c) Back projection. (d) Our approach. (Better viewed on the screen with more than 400% zooming-in.)

image is determined by both the soft-edge and hard-edge constraints.

2.2 Shock Filter

Based on the up-sampling result I_U, a shock filter is employed to enhance strong edges in the high-resolution image. The shock filter is a blind deconvolution technique and originally developed in [12] for image enhancement. The classical 2D shock filter formulation is:

$$I_t = -\text{sign}(I_{\eta\eta})\|\nabla I\|, \tag{5}$$

where $\nabla I = (I_x, I_y)_T$ and $I_{\eta\eta} = I_{xx}I_x^2 + 2I_{xy}I_xI_y + I_{yy}I_y^2$ denote the gradient and the second derivative, respectively.

The drawback of the original shock filter is that it is very sensitive to noise. Consequently, it enhances both true edges and noise. Later improved shock filter algorithms have been proposed. In this paper, we choose the complex shock filter in [7] to obtain the enhanced image. General 2D complex shock filter is defined as

$$I_t = -\frac{2}{\pi}\arctan(a\text{Im}(\frac{I}{\theta}))\|\nabla I\| + \lambda I_{\eta\eta} + \tilde{\lambda} I_{\xi\xi}, \tag{6}$$

where $I_{\xi\xi} = I_{xx}I_y^2 - 2I_{xy}I_xI_y + I_{yy}I_x^2$, $\lambda = re^{i\theta}$ is a complex scalar, $\tilde{\lambda}$ is a real scalar, the parameter a controls the sharpness of the slope near zero, θ is the phase angle of the complex part, and $\text{Im}(\frac{I}{\theta})$ denotes the complex part of $\frac{I}{\theta}$. More details can be found from [7].

Then we update I_S by the following operation:

$$I_S^{t+1} = I_S^t + I_t * d_t \tag{7}$$

where d_t is a iteration step. In our experiments $d_t = 0.1$ and the initial input in the iteration procedure is I_U.

2.3 Reconstruction Constraint Refinement

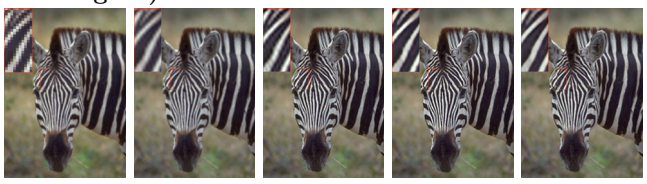
The obtained intermediate result I_S often does not satisfy the reconstruction constraint. Reconstruction error is defined as

$$E(I_L, I_H) = I_L - (I_H \otimes g)\downarrow_s, \tag{8}$$

where $\otimes$ is the convolution operator, g is a spatial filter, and $\downarrow_s$ denotes the down-sampling operator with factor s. The

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

Figure 4: Comparison with a reconstruction based method at 3× magnification on the "Hat" image. (a) Original image zoomed-in by nearest interpolation. (b) Bicubic interpolation. (c) Result by [13]. (d) Our approach. (Better viewed on the screen with zooming-in.)



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

Figure 5: Comparison with another reconstruction based method at 3× magnification on the "Zebra" image. (a) Original image zoomed-in by nearest interpolation. (b) Bicubic Interpolation. (c) Back Projection. (d) Result by [4]. (e) Our approach. (Better viewed on the screen with zooming-in.)

reconstruction constraint requires that the high-resolution image after smoothing and down-sampling should be as close as possible to the input low-resolution image.

As in many image super-resolution methods [15, 14, 4, 13], the reconstruction error can be minimized by the back projection algorithm [8] which is an iterative gradient-based minimization method:

$$I_H^{t+1} = I_H^t + (((I_H^t \otimes g)\downarrow_s - I_L)\uparrow_s) \otimes f \tag{9}$$

where I_H^t is the high-resolution image at iteration t, and g and f are a spatial filter and a Gaussian "back projection" filter, respectively. The intermediate result I_S is used as the initial I_H at the first iteration. In our experiment, the standard variances of g and f are set to $s/2$.

3. EXPERIMENTAL RESULTS

We compare our approach with bicubic interpolation and sharpen bicubic interpolation (using the "unsharp mask" in Adobe Photoshop with the default parameters on the result of bicubic interpolation), and back projection [8] on the "mickey" and "Lena" images in Fig. 3. As shown in the zooming-in region, bicubic interpolation over smoothes the whole image, especially at strong edges. The Sharpen bicubic and back projection methods introduce ringing artifacts on salient edges. The edges of our result are sharp without noticeable artifacts.

Fig. 4 shows the comparison of our approach with a reconstruction based method using the gradient heavy-tailed distribution prior [13] (the code is downloaded from `http://www.cse.cuhk.edu.hk/~leojia/projects/upsampling`) on the "Hat" image, where the results by bicubic interpolation is also given. In Fig. 5, we compare our approach with another reconstruction based method using the soft edge smoothness prior [4] (the results are downloaded from `http://vision.`

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

Figure 6: Comparison with the learning based method at 3× magnification using the "Man" image. (a) Original image zoomed-in by nearest interpolation. (b) Back projection. (c) Result by [4]. (d) Result by [21]. (e) Our approach. (Better viewed on the screen with zooming-in.)

Table 2: RMS values obtained by sveral methods. The best results are in bold.

Images	Bibubic	BP[8]	[13]	[21]	Ours
Man	3.4621	3.3399	4.9643	3.5426	**3.0115**
Zebra	3.9180	3.7476	5.3520	3.9090	**3.6105**
Fire	3.0549	**2.2970**	2.6483	2.6483	2.4373
KNhair	3.6888	3.2124	4.9028	3.6311	**3.1747**

`eecs.northwestern.edu/research/IP/SR`) on the "Zebra" image. The method in [13] suppresses the details in the texture region, such as the region in the rectangle. The result obtained by [4] suffers from ringing artifacts because the imposed prior may not be compatible with the input low-resolution image. On the other hand, our approach preserves the details and sharp edges in the high-resolution results.

In Fig. 6, we compare with a learning based method [21] (the matlab code can be obtained from `http://www.ifp.illinois.edu/~jyang29/`) on the "Man" image. We observe that the learning based method introduces many high frequency artifacts from training data. Our result does not have visible artifacts with strong edges close to the input low-resolution image.

Table 2 shows quantitative results by several algorithms, result by other algorithms are not given because their codes are not available. The measure is root mean error (RMS). Table 3 shows results with structural similarity (SSIM) [20] which is one of the most popular image quality measures. From these two tables, we can see that our algorithms performs the best overall.

4. CONCLUSION

In this paper, we have proposed a novel image super-resolution algorithm that can preserve edges well. First, we perform an up-sampling on the low-resolution image by combining both the soft-edge and hard-edge constraints. Then, a complex shock filter robust to noise is used to recover strong edges. Finally, we enforce a reconstruction constraint on the high-resolution result and obtain the final result by back projection. The main advantages of our algorithm are that it does not require image priors and that it preserves edges well. We have tested our algorithms qualitatively and quantitatively. Compared with several state-of-the-art super-resolution algorithms, ours performs the best overall.

5. ACKNOWLEDGMENTS

This work was supported by Introduced Innovative R&D Team of Guangdong Province (Robot and Intelligent Information Technology); Natural Science Foundation of China (60903117, 61070148); and Science, Industry, Trade, and Information Technology Commission of Shenzhen Municipality, China (JC201005270357A, ZYC201006130314A, JC201005270378A, ZYC201006130313A).

Table 3: SSIM values obtained by sveral methods. The best results are in bold.

Images	Bibubic	BP[8]	[13]	[21]	Ours
Man	0.8617	0.8905	0.5767	0.8716	**0.9205**
Zebra	0.8064	0.8885	0.4982	0.8477	**0.8912**
Fire	0.9299	0.9718	0.9667	0.9435	**0.9729**
KNhair	0.8451	**0.9226**	0.6657	0.8818	0.9225

6. REFERENCES

[1] J. Allebac and P. W. Wong. Edge-directed interpolation. In *Proc. ICIP*, 1996.

[2] S. Baker and T. Kanade. Limits on super-resolution and how to break them. In *Proc. CVPR*, 2000.

[3] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. on PAMI*, 24(5):603–619, May 2002.

[4] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong. Soft edge smoothness prior for alpha channel super resolution. In *Proc. CVPR*, 2007.

[5] R. Fattal. Upsampling via imposed edges statistics. In *ACM SIGGRAPH*, 2007.

[6] W. Freeman and E. Pasztor. Learning low-level vision. *IJCV*, 40(1):25–47, 2000.

[7] G. Gilboa, N. A. Sochen, and Y. Y. Zeevi. Regularized shock filters and complex diffusion. *Lecture Notes in Computer Science*, pages 399–413, 2002.

[8] M. Irani and S. Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, 40(4):324–335, 1993.

[9] X. Li and M. Orchard. New edge-directed interpolation.

[10] Z. Lin, J. He, X. Tang, and C.-K. Tang. Limits of learning-based superresolution algorithms. In *Proc. ICCV*, 2007.

[11] W. Liu, D. Lin, and X. Tang. Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation. In *Proc. CVPR*, 2005.

[12] S. Osher and L. I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, 27:919–940, 1990.

[13] Q. Shan, Z. Li, J. Jia, and C.-K. Tang. Fast image/video upsampling. In *ACM SIGGRAPH ASIA*, 2008.

[14] J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *Proc. CVPR*, 2008.

[15] J. Sun, N.-N. Zheng, H. Tao, and H.-Y. Shum. Image hallucination with primal sketch priors. In *Proc. CVPR*, 2003.

[16] J. Sun, J. Zhu, and M. Tappen. Context-constrained hallucination for image super-resolution. In *Proc. CVPR*, 2010.

[17] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. ICCV*, 1998.

[18] Q. Wang, X. Tang, and H. Shum. Patch based blind image super resolution. In *Proc. ICCV*, 2005.

[19] X. Wang and X. Tang. Hallucinating face by eigentransformation. *IEEE Trans. on SMC, Part C*, 35(3):425–434, 2005.

[20] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. on IP*, 13(4):600–612, 2004.

[21] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *Proc. CVPR*, 2008.