BILATERAL BACK-PROJECTION FOR SINGLE IMAGE SUPER RESOLUTION

†*Shengyang Dai,* ‡*Mei Han,* †*Ying Wu,* ‡*Yihong Gong*

†EECS Department, Northwestern University, Evanston, IL 60208
‡NEC Laboratories America, Inc., Cupertino, CA 95014

ABSTRACT

In this paper, a novel algorithm for single image super resolution is proposed. Back-projection [1] can minimize the reconstruction error with an efficient iterative procedure. Although it can produce visually appealing result, this method suffers from the chessboard effect and ringing effect, especially along strong edges. The underlining reason is that there is no edge guidance in the error correction process. Bilateral filtering can achieve edge-preserving image smoothing by adding the extra information from the feature domain. The basic idea is to do the smoothing on the pixels which are nearby both in space domain and in feature domain. The proposed bilateral back-projection algorithm strives to integrate the bilateral filtering into the back-projection method. In our approach, the back-projection process can be guided by the edge information to avoid across-edge smoothing, thus the chessboard effect and ringing effect along image edges are removed. Promising results can be obtained by the proposed bilateral back-projection method efficiently.

1. INTRODUCTION

The objective of image super resolution (SR) [2] is to obtain high resolution (HR) images from low resolution (LR) inputs. It is widely applicable in video communication, object recognition, HDTV, image compression, *et al*. In this paper, we mainly focus on the super resolution task given one single low resolution input image.

The generation process of low resolution images can be modeled as a combination of smoothing and down-sampling operations of the natural scenes by low quality sensors. Super resolution is an inverse problem of this generation process, so it is under-determined due to the information loss. One criteria of solving this inverse problem is minimizing the reconstruction error. In other words, the result which can produce the same low resolution image as the input one is preferred.

Various methods are proposed in the literature to regularize the inverse problem. Two of the most extensively explored image modeling priors are the image smoothness prior and the edge smoothness prior. Simple filtering/interpolation algorithms (*e.g.*, bilinear/bicubic interpolation) can produce smooth high resolution images, which are usually blurry, thus have limited image quality. To preserve edge sharpness, edge directed interpolation [3, 4] is proposed to prevent cross-edge interpolation. However, locating high precision edge position itself is a non-trivial task. Level-set [5] and multiple-scale tensor voting [6] methods are explored to get smooth edges. Edge preserving smoothness prior on large neighborhood is proposed in [7]. In [8], a soft edge smoothness term which can measure the average length of discrete level lines is incorporated into an objective function to produce smooth soft edges. It is applied on alpha channel to achieve a uniform treatment of edges with different strength. Instead of image prior modeling, many researchers use image exemplar directly. An image can be modeled as a Markov Random Fields as in [9], This idea is extended onto domain-specific video SR in [10].

Generally speaking, exemplar-based algorithms can produce high contrast details in HR images, while powerful database indexing/searching method and efficient MRF optimization algorithm are usually necessary. On the other hand, methods based on image modeling priors are more efficient. One critical issue is how to handle image edges in a satisfactory way. Simple interpolation strategies tend to produce blurry results, while edge preserving methods may remove image details in regions without strong edges.

Back-projection [1] method can minimize the reconstruction error efficiently by an iterative algorithm. In each step, the current reconstruction error is back-projected to adjust the image intensity. Although this method can improve the image quality greatly, it suffers from some unsatisfying artifacts, such as the ringing effect and the chessboard effect. The underlining reason is the usage of *isotropic* back-projection kernel. In fact, due to the under-determined nature of the SR task, there exist a lot of minimizers for the reconstruction error. It is very likely that the isotropic back-projection kernel leads to unsatisfactory results, since the edge information is totally ignored throughout the update procedure.

In this paper, the bilateral back-projection method is proposed to solve the problems associated with the original one, when it is applied to single image SR. We first show that, for any given positive integer scaling factor, the original back-projection algorithm can minimize the reconstruction error efficiently under certain conditions. Then, the idea of bilateral filtering is employed to guide the back-projection process. The image edge information is integrated to avoid across-

edge projection, thus the ringing effect and chessboard effect can be removed. The proposed bilateral back-projection algorithm is introduced in Sec. 2. The experiment results are shown in Sec. 3 and Sec. 4 concludes the paper.

2. BILATERAL BACK-PROJECTION

2.1. Problem modeling

In theory, the generation process of LR image can be modeled by a combination of the blur effect (due to the atmosphere, the object/camera motion, and the sensor) and the down-sampling operation. By simplifying the blur effect with a single filter g for the entire image, the generation process can be formulated as follows

$$\mathbf{I}^l = (\mathbf{I}^h * g) \downarrow_s, \tag{1}$$

where $\mathbf{I}^h$ and $\mathbf{I}^l$ are the HR and LR images respectively, $*$ is the convolution operator, and $\downarrow_s$ is the down-sampling operator with scaling factor s.

The reconstruction error of an HR image $\mathbf{I}$ is defined as the difference between the LR input image and the synthesized LR image by $\mathbf{I}$ as follows

$$\mathbf{e}_r(\mathbf{I}) = \mathbf{I}^l - (\mathbf{I} * g) \downarrow_s \tag{2}$$

2.2. Back-projection

Back-projection [1] is an efficient algorithm to get the HR image by minimizing the norm of the reconstruction error defined by Eqn. 2. It is originally designed for the case with multiple LR inputs. Given only one LR input image, the updating procedure can be summarized as doing the following two steps iteratively:

- Compute the LR error $\mathbf{e}_r(\mathbf{I}_t^h)$ by Eqn. 2.
- Update the HR image by back-projecting the error as follows

$$\mathbf{I}_{t+1}^h = \mathbf{I}_t^h + \mathbf{e}_r(\mathbf{I}_t^h) \uparrow_s * p, \tag{3}$$

where $\mathbf{I}_t^h$ is the HR image at the t-th iteration, $\uparrow$ is the up-sampling operator, p is a constant back-projection kernel.

It is proved in [1] that for $s = 1$ (the problem of SR is equivalent to the problem of deblurring in this case) and multiple LR input images, the back-projection algorithm can converge to the desired deblurred image, which satisfies Eqn. 1 for all LR inputs under their corresponding geometry transform, with an exponential rate under certain conditions.

We first extend the above assertion to the case with arbitrary positive integer scaling factor and single LR input image, and show that the back-projection iteration converges to the desired image, which satisfies Eqn. 1, for arbitrary positive integer s under certain conditions.

Theorem 1 *By updating the HR image with the back-projection iteration, $\mathbf{I}_t^h$ will converge to a desired image $\mathbf{I}^c$, which satisfies Eqn. 1, with an exponential rate for all $s \geq 1$, given $||\delta - g * p \downarrow_s ||_1 < 1$.*

The proof of Theorem. 1 is presented in Appendix. A. It means that by applying the back-projection method iteratively, the reconstruction error can be minimized efficiently for any positive integer scaling factor s, with one LR input image, when $||\delta - g * p \downarrow_s ||_1 < 1$. Similar to the discussion in [1], back-projection filter corresponds to smaller value of $||\delta - g * p \downarrow_s ||_1$ will have faster converge speed, while it may produce numerically instable result.

2.3. Bilateral filtering

Bilateral filtering [11] is a non-linear filtering technique which can combine image information from both of the space domain and the feature domain in the filtering process. It can be represented by the following equation

$$\mathbf{h}(x) = \frac{1}{k(x)} \sum_y \mathbf{I}(y) c(x, y) s(\mathbf{I}(x), \mathbf{I}(y)), \tag{4}$$

where $\mathbf{I}$ and $\mathbf{h}$ are the input and output images respectively, x and y are pixel positions over the image grid, $c(x, y)$ and $s(\mathbf{I}(x), \mathbf{I}(y))$ measure the spatial and photometric affinity between pixel x and pixel y respectively, and

$$k(x) = \sum_y c(x, y) s(\mathbf{I}(x), \mathbf{I}(y)) \tag{5}$$

is the normalization factor at pixel x. The functions $c(\cdot)$ and $s(\cdot)$ are usually chosen as follows

$$c(x, y) = \exp(\frac{-||x - y||_2^2}{2\sigma_c^2}), \tag{6}$$

$$s(u, v) = \exp(\frac{-||u - v||_2^2}{2\sigma_s^2}). \tag{7}$$

The underlining idea of the bilateral filtering is to do the smoothing according to pixels not only close in the space domain, but close in the feature domain as well, thus the edge sharpness is preserved by avoiding the cross edge smoothing. Bilateral filtering is closely related to other edge preserving techniques such as nonlinear diffusion and adaptive smoothing [12].

2.4. Bilateral back-projection

The back-projection procedure in Sec. 2.2 can be explained as a process to correct the reconstruction error iteratively. In each step, the LR error $\mathbf{e}_r(\mathbf{I}^h)$ is back-projected to HR image by a kernel p. However, as we mentioned above, since a lot of information is lost in the generation process, the problem of super resolution is severely under-determined. There might be multiple solutions to minimize the reconstruction error, even for multiple LR input images [13, 14]. Many minimizers are not satisfying although with zero reconstruction error. Since minimizing the reconstruction error is the only objective for back-projection algorithm, the iteration step may converge to some unsatisfactory result. The most commonly observed artifacts are the ringing effect and the chessboard effect [6]. The underlining reason for these artifacts is the isotropic nature of the back-projection kernel p. The error correction step propagates the error without considering the

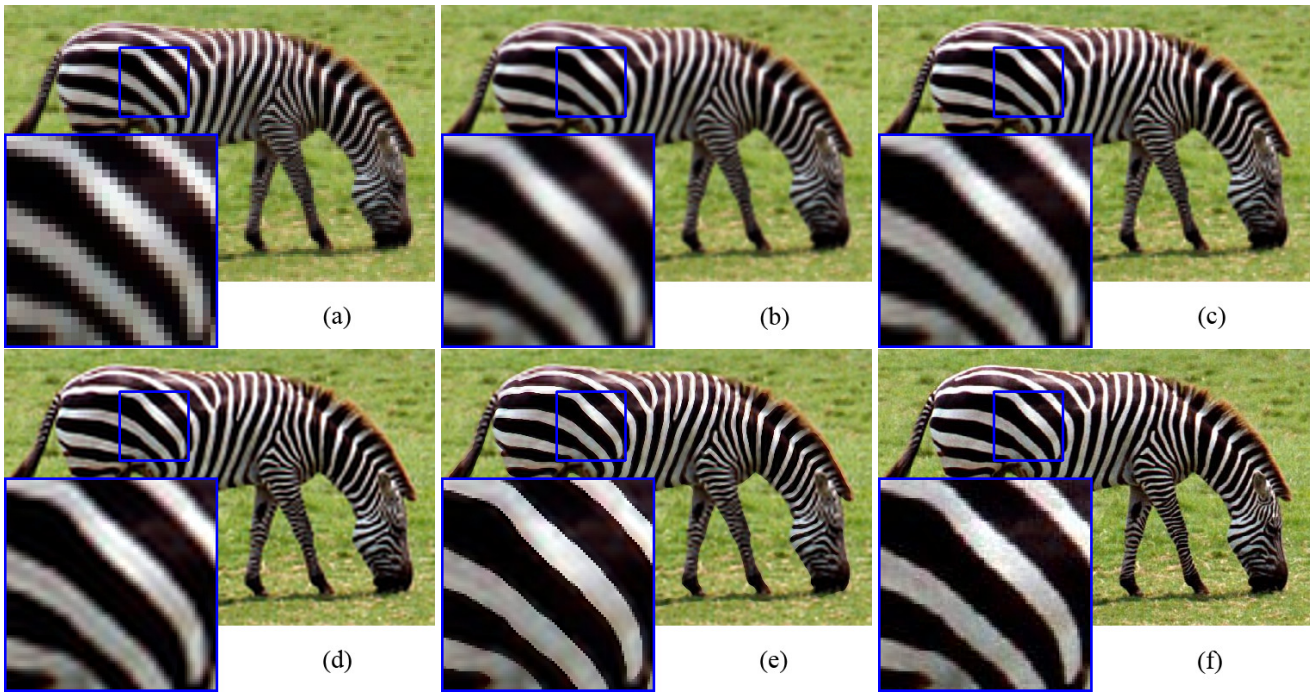

Fig. 1. (a) LR input image (b) bicubic interpolation (c) sharpened bicubic (d) back-projection (e) bilateral BP (f) ground truth

local edge direction and strength. The cross-edge error propagation may produce ringing effect, and the isotropic kernel results in chessboard effect. To remove such artifacts, more sophisticated updating procedure is necessary.

We propose to use bilateral filtering method to propagate the error according to the edge information. Two important issues need to be addressed here.

First, the information of salient image edges need to be extracted. The method proposed in [8] is applied. It basically has two steps. 1. Each edge segment is decomposed by the alpha matting technique, which describes the neighboring region as a linear combination of two sides of this segment through an alpha channel. A soft edge smoothness prior is applied to super resolve the alpha channel, which is further used to synthesis a HR smooth and sharp edge. 2. To improve the image quality for the regions without salient edge segments, a back-projection based post-processing step is employed. Since we only need salient edge information to guide the back-projection, the resulting image of step 1 is extracted as the edge guidance image, assume it is $\mathbf{I}_g^h$.

Secondly, we recover the image details by the proposed bilateral back-projection algorithm. It can enhance the entire image quality without introducing other artifacts. The initial image $\mathbf{I}_0^h$ is set equal to $\mathbf{I}_g^h$. The original bilateral filtering in Eqn. 4 is modified as follows

$$\mathbf{h}(x) = \sum_y \mathbf{I}(y) c(x,y) s(\mathbf{I}_g^h(x), \mathbf{I}_g^h(y)). \tag{8}$$

Here, $s(\mathbf{I}_g^h(x), \mathbf{I}_g^h(y))$ is used to replace $s(\mathbf{I}(x), \mathbf{I}(y))$ in Eqn. 4, and this filter is applied on the HR error image $\mathbf{e}_r(\mathbf{I}_t^h) \uparrow_s$ in each step. It is different from Eqn. 4 since the feature is not extracted from the input image itself, instead, a fixed edge guidance image $\mathbf{I}_g^h$ is used to convey the desired edge distribution information into the iteration procedure. There are generally two cases. For a homogeneous region in $\mathbf{I}_g^h$, the bilateral back-projection algorithm is the same as the original one, thus the error is back-projected almost isotropically. On the other hand, for a region near a step edge, the error will be only propagated in the part on the same side of the edge as the position corresponding to the LR error on the HR image. We use the same guidance image throughout the entire process instead of updating it dynamically according to $\mathbf{I}_t^h$ for two reasons: (1) The resulting local intensity fluctuation will not influence the updating procedure, thus stable edges can be produced. (2) It also improves the efficiency, since the filter need to be computed only once without updating.

3. EXPERIMENTAL RESULTS

Figure 1 shows the result comparison between different algorithms. Fig. 1(a) is the LR input image. Fig. 1(b) is the result of bicubic interpolation, which is very blurry. Fig. 1(c) shows the result of the sharpened bicubic method, which is obtained by using Photoshop. It has higher contrast than (b), but the chessboard effect is clearly observable. Fig. 1(d) is the result of original back-projection algorithm. The image quality is improved, however, chessboard effect is still observable, and some ringing effect is also introduced, especially in the center of the white regions. Fig. 1(e) shows the result of the proposed bilateral back-projection, clear and sharp edges are obtained without those unsatisfying artifacts. Fig. 1(f) is the ground truth image, which is used to generate the LR input image. One patch is zoomed to better illustrate the difference.

Fig. 2. More experiment results, the first row shows the LR input images, and the second row shows our results.

Although our method can improve the visual effect greatly, the ability of reducing the reconstruction error is similar to the back-projection method. After 15 iteration steps, the RMS error is 12.01 for bilateral back-projection, and 12.99 for back-projection method. More results are shown in Fig. 2.

4. CONCLUSION

In this paper, we propose the bilateral back-projection method for single image super resolution. The edge information is incorporated with the bilateral filtering method, to remove the artifact caused by the original back-projection algorithm. The error correction on high resolution image is back-projected according to the image edges. Impressive results illustrate the effectiveness of our algorithm.

5. REFERENCES

[1] M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion and transparency," *JVCIP*, 1993.

[2] A.K. Katsaggelos, R. Molina, and J. Mateos, *Super resolution of images and video*, Synthesis Lectures on Image, Video, and Multimedia Processing. Morgan & Claypool, 2007.

[3] J. Allebach and P. W. Wong, "Edge-directed interpolation," in *ICIP*, 1996.

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

[5] B. S. Morse and D. Schwartzwald, "Image magnification using level set reconstruction," in *CVPR*, 2001.

[6] Y. Tai, W. Tong, and C. Tang, "Perceptually-inspired and edge-directed color image super-resolution," in *CVPR*, 2006.

[7] S. Farisiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

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

[9] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, 2002.

[10] D. Kong, M. Han, W. Xu, H. Tao, and Y. Gong, "Video super-resolution with scene-specific priors," in *BMVC*, 2006.

[11] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *ICCV*, 1998.

[12] D. Barash, "A fundamental relationship between bilateral filtering, adaptive smoothing and the non-linear diffusion equation," *IEEE Trans. on PAMI*, vol. 24, no. 6, pp. 844–847, 2002.

[13] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. on PAMI*, vol. 24, no. 9, pp. 1167 – 1183, 2002.

[14] Z. Lin and H. Shum, "Fundamental limits of reconstruction based super-resolution algorithms under local translation," *IEEE Trans. on PAMI*, vol. 26, no. 1, pp. 83 – 97, 2004.

[15] S. Saitoh, V. K. Tuan, and M. Yamamoto, "Convolution inequalities and applications," *Journal of Inequalities in Pure and Applied Mathematics*, vol. 4, no. 3, 2003.

A. PROOF OF THEOREM 1

We first prove that $||\mathbf{e}_r(\mathbf{I}_t^h)||_1$ decreases after each iteration.

$$\begin{aligned}
\mathbf{e}_r(\mathbf{I}_{t+1}^h) &= \mathbf{I}^l - (\mathbf{I}_{t+1}^h * g)\downarrow_s \\
&= \mathbf{I}^l - \left((\mathbf{I}_t^h + (\mathbf{I}^l - \mathbf{I}_t^h * g\downarrow_s)\uparrow_s * p) * g\right)\downarrow_s \\
&= \mathbf{I}^l - \mathbf{I}_t^h * g\downarrow_s - (\mathbf{I}^l - \mathbf{I}_t^h * g\downarrow_s)\uparrow_s * p * g\downarrow_s \\
&= (\mathbf{I}^l - \mathbf{I}_t^h * G\downarrow_s) * (\delta - p * g\downarrow_s) \\
&= \mathbf{e}_r(\mathbf{I}_t^h) * (\delta - p * g\downarrow_s)
\end{aligned}$$

From Young's inequality [15], we have

$$\begin{aligned}
||\mathbf{e}_r(\mathbf{I}_{t+1}^h)||_1 &= ||\mathbf{e}_r(\mathbf{I}_t^h) * (\delta - p * g\downarrow_s)||_1 \\
&\leq ||\mathbf{e}_r(\mathbf{I}_t^h)||_1 \times ||(\delta - p * g\downarrow_s)||_1
\end{aligned}$$

Thus $\mathbf{I}_t^h * g\downarrow_s \rightarrow \mathbf{I}^l$ as $t \rightarrow \infty$ with exponential rate, when $||(\delta - p * g\downarrow_s)||_1 < 1$.

Based on this, it is easy to see that $\mathbf{I}_t^h$ will converge to an image $\mathbf{I}^c$, which satisfies $\mathbf{I}^c * g\downarrow_s = \mathbf{I}^l$.