STRUCTURE PRESERVING SINGLE IMAGE SUPER-RESOLUTION

Fan Yang, Don Xie, Huizhu Jia, Rui Chen, Guoqing Xiang and Wen Gao*

National Engineering Laboratory for Video Technology, Peking University, Beijing 100871, China

ABSTRACT

In this paper, we present a novel structure preserving method for single image super-resolution to well construct edge structures and small detail structures. In our approach, the sharp edges are recovered via a novel edge preserving interpolation technique based on a well estimated gradient field and the edge preserving method, which incorporate the local and non-local structure information. The gradient of interpolated high-resolution(HR) image is then regarded as an edge preserving constraint to reconstruct the detail structures. Experimental results demonstrate that the new approach can reconstruct faithfully the HR images with sharp edges and texture structures, and annoying artifacts (blurring, jaggies, ringing, etc.) are greatly suppressed. It outperforms the state-of-the-art approaches, based on subjective and objective evaluations.

Index Terms— Super-resolution, structure preserving, gradient field.

1. INTRODUCTION

Image enlargement, reconstructing a sharp HR image from its low resolution(LR) counterpart, has a wide range of applications such as medical imaging, remote sensing, consumer electronics, etc. The human visual systems are highly sensitive to edge and texture structures, therefore reconstruction edges and texture structures while suppressing visual artifacts is a critical requirement for image upsampling.

Conventional linear interpolators, such as cubic convolution interpolation[1], have a low complexity but produce annoying artifacts. To preserve sharp edges, many edge-directed interpolation methods have been proposed[2][3]. Li[4] proposed to estimate the covariance of the high-resolution image from the the low-resolution image, and then use the estimated covariance to adapt the interpolation. Another representative edge-guided interpolators is proposed by Zhang and Wu[5], in which a 2-D piecewise autoregressive model is used to estimate missing pixels in groups. Hung and Siu[6][7] presented a robust soft-decision interpolation algorithm using Weighted least-squares estimation for both parameter and data estimation steps. Jing[8] proposed to use edge prior for image interpolation. Wu[9] proposed a high order context to image interpolation. However, these methods interpolate the missing pixels mainly in local region, less or without considering non-local information.

The iterative back-projection[10] technique is a representative reconstruction based method. It can minimize the reconstruction errors efficiently by an iterative process. However, it can produce undesired artifacts along the edges, because the reconstruction errors are back projected into the reconstructed image isotropically. To suppress artifacts, Dai[11] used bilateral filtering to guide the back-projection process, Wang[12] employed the edge sharpness preserving constraint to improve the HR image quality.

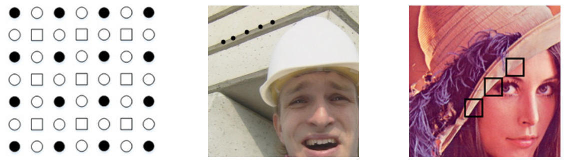
In this paper, we propose a two-step sharpness preserving upsampling method. In the first step, sharp edges in the recovered HR result are ensured by a novel gradient guided interpolation technology which enlarges image based on well estimated gradient field and non-local information. In the second step, to recover small-scale textures, we use a sharpness preserving reconstruction algorithm. Our approach performs significantly better than both the interpolation and the reconstruction based methods in preserving image sharpness while suppressing artifacts(Codes are available at http://www.escience.cn/people/FanYang/index.html.).

2. EDGE PRESERVING INTERPOLATION

This section presents a novel image interpolation method to generate an initial HR image, for the following reconstruction process. Without the loss of generality, we assume that the HR image is first smoothed by Gaussian function, and then downsampled to yield the LR image by a factor of two as illustrated by Fig.1(a). For a local region around an edge, the image intensity variation along the edge is locally similar and piecewise stationary as illustrated by Fig.1(b). Generally, natural images often contain repeated patterns and structures, as shown in Fig.1(c). A natural idea is to employ such local and non-local redundancies to improve the image interpolation quality. Therefore, our interpolation method first estimates the gradient of the HR image by fusing gradients in non-local region, and then the missing pixels are estimated as the weighted average of chosen LR pixels in local region based on gradient guided. In the next we develop a non-local post-processing technique to improve the image interpolation quality.

*The corresponding author, Don Xie is with Peking University, also with Cooperative Medianet Innovation Center and Beida Information Research.

978-1-4673-9961-6/16/$31.00 ©2016 IEEE



(a) (b) (c)

Fig. 1. (a) Spatial relationship among LR samples (black dot) and the missing pixel (empty circles and empty squares). (b) Pixels (black dot) along the edge. We assume that these pixels have similar values. (c) Image repetitive structures.

2.1. Non-local gradient fusion

In order to obtain the gradient of HR image, first, we apply conventional bicubic interpolation to the LR image so as to obtain the initial HR image. Second we use Sobel kernels to convolve with the initial HR image to get the approximated gradient of HR image. Let $y(i_0, j_0)$ be an initially interpolated HR pixel at location (i_0, j_0), the empty dot in Fig.1(a), and $G(i_0, j_0)$ be its gradient. Denote $y(i, j)$ as an LR pixel at location (i, j), the black dot in Fig.1(a), and $G(i, j)$ as the gradient of it. In general, repeated structures in natural images have similar gradient features. Let S be the search window, the gradient of $y(i_0, j_0)$ is computed as the weighted average of all the gradients of LR pixels in S

$$G(i_0, j_0) = \sum_{y(i,j)\in S} w(i, j)G(i, j) \tag{1}$$

where the weight $w(i, j)$ depends on the similarity between pixels $y(i_0, j_0)$ and $y(i, j)$. Denote $N(i_0, j_0)$ as the squared neighborhood of $y(i_0, j_0)$, $N(i, j)$ as the squared neighborhood of $y(i, j)$. The similarity between two pixels $y(i_0, j_0)$ and $y(i, j)$ is measured by the similarity of the gray level intensity and distribution. The distance of gray level intensity between $y(i_0, j_0)$ and $y(i, j)$ is defined by

$$d(i, j) = \left\|N(i_0, j_0) - N(i, j)\right\|_2^2 \tag{2}$$

where $\|\cdot\|$ is the L_2 norm operator. For the gray level distribution, we use perceptual hashing[13] for block hashing so as to obtain a concise representation of gray level distribution. Denote binary image block $H(i_0, j_0)$ and $H(i, j)$ as the hash value of $N(i_0, j_0)$ and $N(i, j)$. For a block $N(i, j)$, first, the mean of the pixel values is computed. Next, for each pixel in $N(i, j)$, a 1 is assigned to the corresponding coordinate in $H(i, j)$ if the pixel value is greater than the mean; otherwise a 0 is assigned. The distance of gray level distribution between $y(i_0, j_0)$ and $y(i, j)$ is defined by

$$h(i, j) = \left\|H(i_0, j_0) - H(i, j)\right\|_1^1 \tag{3}$$

Similar to the non-local means denoising[14], we set $w(i, j)$ as the exponential function of distance $d(i, j)$ and $h(i, j)$

$$w(i, j) = \frac{1}{Z(i_0, j_0)} e^{-\frac{d(i,j)}{\sigma_1}} e^{-\frac{h(i,j)}{\sigma_2}} \tag{4}$$

where $Z(i_0, j_0)$ is the normalization constant

$$Z(i_0, j_0) = \sum_{y(i,j)\in S} e^{-\frac{d(i,j)}{\sigma_1}} e^{-\frac{h(i,j)}{\sigma_2}} \tag{5}$$

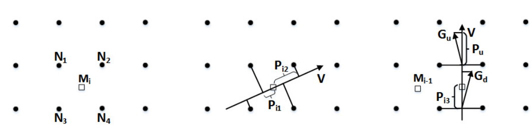
the parameter σ_1 and σ_2 controls the decay of the exponential function.

2.2. Gradient guided interpolation

As previously described, let the LR image I_l be a down sampled version of the HR image I_h by a factor of two. There are two different kinds of missing pixels as shown in Fig.1(a). Interpolation is done in two steps. The first step interpolates the HR pixels marked by empty squares, and the second step interpolates the remaining unknown pixels marked by empty circles. Fig.2(a) illustrates the spatial configuration of known pixels and interpolated pixel involved in the first step. For a interpolated pixel M_i of the HR image, its four neighbors are known LR pixels, denoted by $N_j, j = 1, 2, 3, 4$. The value of M_i is estimated as the weight average of N_j

$$M_i = \sum_{j=1}^{4} \omega(i, j)N_j \tag{6}$$

where the weights $\omega(i, j)$ are estimated via the gradient information. Let $\beta_k = \{\theta | \frac{k\pi}{4} - \frac{\pi}{8} \le \theta < \frac{k\pi}{4} + \frac{\pi}{8}\}, k = 0, 1, ..., 7$, denote the eight octants of the 2D plane, where $\theta \in [-\frac{\pi}{8}, \frac{15\pi}{8})$ is the angle of a gradient in the positive direction of the x-axis. we group the eight octants into two sets Ω_1 and Ω_2, where $\Omega_1 = \{\beta_1, \beta_3, \beta_5, \beta_7\}$, $\Omega_2 = \{\beta_0, \beta_2, \beta_4, \beta_6\}$. According to the image prior which assumes that the image intensity variation along the edge is locally similar, in order to preserve edge sharpness, we propose to estimate the weights in two ways. Let θ_M denote the angle of gradient of missing pixel M_i in the positive direction of the x-axis.



(a) (b) (c)

Fig. 2. (a)Spatial configuration in the first step. (b) θ_M belongs to set Ω_1. V represent the gradient direction of unknown pixel M_i. (c) θ_M belongs to set Ω_2.

1) If θ_M belongs to set Ω_1, as shown in Fig.2(b), similar to the gradient guided image interpolation[8], we project the LR pixel N_j to the gradient direction of interpolated pixel M_i, and set $P_{(ij)}$ as the projection length of pixel N_j. The longer the projection length $P_{(ij)}$, the smaller weights LR pixel N_j will carry when estimating the missing pixel M_j. We define the weights as

$$\omega(i, j) = \frac{1}{C(i)} e^{-\frac{P_{(i,j)}}{\sigma_3}} \tag{7}$$

where $C(i) = \sum_{j=1}^{4} e^{-\frac{P_{(i,j)}}{\sigma_3}}$ is the normalizing constant, the parameter σ_3 controls the decay of the exponential function.

2) If θ_M belongs to the set Ω_2, the projection length of four neighbors are almost the same, as shown in Fig.2(c). The weights are nearly identical when calculated via Eqn.(7). To preserve the horizontal and vertical edges, that have maximum gradient in local region, we let the interpolated pixel close to the side with smaller gradient. Therefore, we redefine the weights as

$$\omega(i,j) = \frac{1}{C(i)'}\delta_j e^{-\frac{P_{(i,j)}}{\sigma_3}} \tag{8}$$

where δ_j is the scaling coefficient with the initial value 1. Assume that θ_M is the nearest to the vertical(Fig.2(c)), denote the statistical gradient on the upside as $G_u = (G(i,1) + G(i,2))/2$, and the corresponding downside is $G_d = (G(i,3) + G(i,4))/2$. Then projecting the gradients G_u and G_d to the gradient direction of M_i, and setting P_u, P_d as the projection length. We update the δ_j as

$$\begin{cases} \delta_1 = \delta_2 = 0 & P_u - P_d > T \\ \delta_3 = \delta_4 = 0 & P_d - P_u > T \\ \delta_1 = \delta_2 = 0 & |P_u - P_d| \le T \& V_d \le V_u \\ \delta_3 = \delta_4 = 0 & |P_u - P_d| \le T \& V_d > V_u \end{cases} \tag{9}$$

where T is the threshold, $V_u = (N_1 + N_2) - 2M_{i-1}$, $V_d = (N_3 + N_4) - 2M_{i-1}$, M_{i-1} is the previous interpolated pixel.

As previously described, first type pixels are interpolated first. In the second step of the interpolation, the interpolated first type pixels act as the known pixels, as shown in Fig.3. The spatial configuration of known pixels and the interpolated pixel is rotated by $45°$ to interpolate the second type pixels.

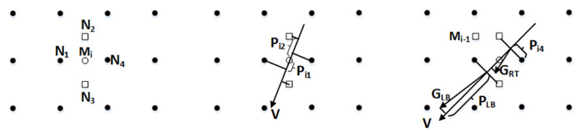


Fig. 3. Spatial configuration in the second step.

2.3. Non-local data fusion

The previous estimation of HR pixels only uses LR pixels in local region, although the edge can be well preserved, it may produce many "jaggies" as shown in Fig.4(a). The interpolation error can be viewed as the noise. To improve the image quality, we propose to fuse the interpolated HR pixels with the similar LR pixels to reduce the noise. Similar to Eqn.(1), for a missing pixel $y(i_0, j_0)$, we update the pixel value as

$$y(i_0, j_0) = \sum_{y(i,j)\in S} w(i,j)y(i,j) \tag{10}$$

where $y(i,j)$ is the LR pixel, $w(i,j)$ is calculated by Eqn.(4). Here, block matching is performed in HR image obtained by gradient guided interpolation. The visual quality of the interpolated image is improved greatly as shown in Fig.4(b).

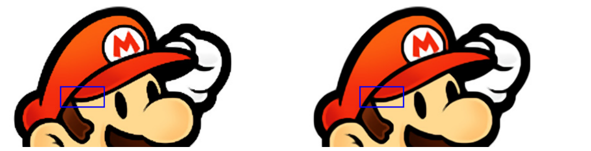


Fig. 4. Image upsampling with the factor of 2. From left to right: the result without non-local data fusion, the result with non-local data fusion.

3. SMALL DETAIL STRUCTURE PRESERVING RECONSTRUCTION

The gradient guided interpolation can reconstruct sharpness edges, but little consideration is given to the reconstruction of detail structures. Therefore, in this step, we will focus on the reconstruction of small details on the premise that assures the sharpness of edge structures. Given the LR image I_l and interpolated HR image I_h^0, we estimate the HR image by minimizing the following energy function that enforce the constraints in both image domain and gradient domain

$$I_h^* = arg\min_{I_h} \left\| [I_h \otimes G]_{\downarrow(n)} - I_l \right\|_2^2 + \lambda \left\| \sum_{\varphi\in\Psi} \varphi I_h - \varphi I_h^0 \right\|_2^2 \tag{11}$$

where $\otimes$ is the convolution operator. G represents the Gaussian kernel and $\downarrow(n)$ means downsampling image with the factor n. φ and Ψ are gradient extraction operator and corresponding operator set respectively. Let I_h^0 represent the HR image obtained by gradient guided interpolation in the first setp.We use λ to control the relative weights of the data fidelity term and the gradient regularization term. Larger λ places larger importance on the gradient domain constraint, which helps to produce sharp edges with little artifacts.

To solve the objective function, we use the gradient descent method, and in each iteration, the solution is updated as

$$\begin{aligned} I_h^{t+1} = I_h^t - \tau\Big(& ([[I_h \otimes G]_{\downarrow(n)} - I_l]_{\uparrow(n)}) \otimes G \\ & + \lambda(\sum_{\varphi\in\Psi} \varphi^T \varphi I_h^t - \varphi^T \varphi I_h^0)\Big) \end{aligned} \tag{12}$$

where t is the iteration number and τ is the iteration step, $\uparrow(n)$ means upsampling image with the factor n. For the gradient extraction operators, we use the operators which are similar to Sobel operators. Let $\Psi = \{\varphi_1, \varphi_2\}$, where $\varphi_1 = \begin{bmatrix} -1/4 & 0 & 1/4 \\ -1/2 & 0 & 1/2 \\ -1/4 & 0 & 1/4 \end{bmatrix}$ and $\varphi_2 = \begin{bmatrix} -1/4 & -1/2 & 1/4 \\ 0 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$.

4. EXPERIMENTAL RESULTS

Our method integrates interpolation-based and reconstruction-based methods, therefore, we compare with the three representative algorithms in the two category: the bicubic interpolation method[1], the gradient guided image interpolation[8] and the fast image up-sampling via the displacement field method[12]. We also compare with the sparse representation based algorithm(ScSR)[15]. In all the experimentation we have fixed a search window S of 21×21 pixels and a similarity square neighborhood $N(i,j)$ of 7×7 pixels. The σ_1 is fixed to be the variance of square neighborhood $N(i_0, j_0)$, $\sigma_2 = 0.1$, $\sigma_3 = 0.2$, $\lambda = 0.2$, $\tau = 0.1$. The number of iterations is set to 30. For color images, we apply our method only to the luminance channel and the chromatic channels are interpolated by the bicubic interpolator. Fig.5 lists the eight test images.

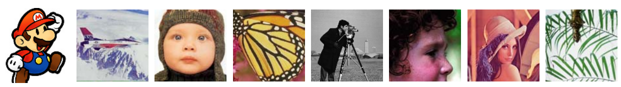


Fig. 5. The eight images in our test.

Table 1. PSNR(dB) results. The best in each row is in bold.

Images	Bicubic[1]	Jing[8]	Wang[12]	ScSR[15]	Ours
Cartoon	29.24	29.52	30.50	30.09	**30.90**
Airplane	31.09	31.63	32.33	31.24	**32.79**
Baby	31.84	32.66	32.90	31.97	**33.81**
Butterfly	28.07	28.27	28.48	28.62	**29.17**
Camera	29.29	29.45	29.63	29.61	**29.94**
Head	29.07	29.62	29.83	29.13	**30.05**
Lena	31.11	32.11	31.95	31.22	**32.91**
Leaf	26.38	26.65	27.74	27.02	**28.08**

For quantitative comparison, in practice, the LR images are obtained by passing the original image through a Gaussian PSF kernel with standard deviation 1 and down-sampling the smoothed image with a factor of 2. The peak signal-to-noise ratio(PSNR) is utilized to measure the upsampling errors. Tables 1 list the PSNR results of the four algorithms for the test images. It can be seen, that our method can improve the P-SNR measure higher than the aforementioned methods.

Visual comparisons are shown in Fig.6, Fig.7. The bicubic interpolator produces very blurred results and clearly visible jaggy. Jing et al.'s method eliminates most of the jaggy, however, the edges are not very sharp and it produces erroneous local structures for regions with high texture details. Wang et al.'s method can produce sharp edges and increase the contrast, but there are some visually unpleasant artifacts along the edge, and the detail structures are not very sharp. The proposed method produces the most visually pleasant results, sharp edges and texture structures with much less artifacts.

5. CONCLUSIONS

In this paper, we proposed a novel algorithm for image enlargement that combines interpolation-based and reconstruction-based methods. In the interpolation process, our method interpolates the missing pixels via the well estimated gradient field and non-local data fusion. In the texture structures reconstruction, we incorporate a gradient constraint to preserve the edge sharpness obtained in interpolation process. The proposed method can improve the result in both subjective visual quality and PSNR measures.

6. ACKNOWLEDGE

This work is partially supported by grants from the National Science Foundation of China under contract No.61421062 and No.61520106004 and the Major National Scientific Instrument and Equipment Development Project of China under contract No. 2013YQ030967.

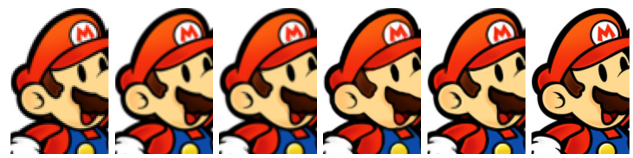
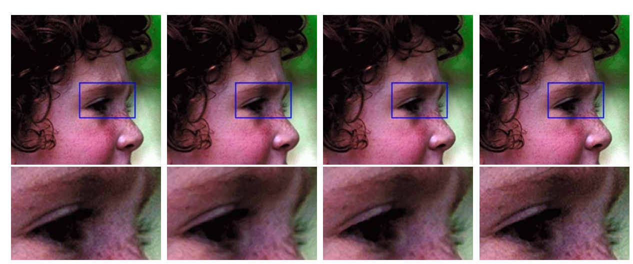


Fig. 6. Super-resolution comparison (2X). From left to right: the input LR image, the result generated by Bicubic[1], the result generated by Jing[8], the result generated by Wang[12], our result, and the ground truth. Please zoom in to see the detail changes.

(a)Camera

(b)Baby



(c)Head

Fig. 7. Super-resolution comparison (4X). From left to right: the result generated by Bicubic, the result generated by Jing, the result generated by Wang, our result.

7. REFERENCES

[1] R. Keys, "Cubic convolution interpolation for digital image processing," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, no. 6, pp. 1153–1160, Dec 1981.

[2] D. Zhang and Xiaolin Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *Image Processing, IEEE Transactions on*, vol. 15, no. 8, pp. 2226–2238, Aug 2006.

[3] A. Giachetti and N. Asuni, "Real-time artifact-free image upscaling," *Image Processing, IEEE Transactions on*, vol. 20, no. 10, pp. 2760–2768, Oct 2011.

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

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

[6] Kwok-Wai Hung and Wan-Chi Siu, "Improved image interpolation using bilateral filter for weighted least square estimation," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sept 2010, pp. 3297–3300.

[7] Kwok-Wai Hung and Wan-Chi Siu, "Robust soft-decision interpolation using weighted least squares," *Image Processing, IEEE Transactions on*, vol. 21, no. 3, pp. 1061–1069, March 2012.

[8] Guangmei Jing, Yi-King Choi, Jiaye Wang, and Wenping Wang, "Gradient guided image interpolation," in *Image Processing (ICIP), 2014 IEEE International Conference on*, Oct 2014, pp. 1822–1826.

[9] Xiaolin Wu and Xiangjun Zhang, "Image interpolation using texture orientation map and kernel fisher discriminant," in *IEEE International Conference on Image Processing 2005*, Sept 2005, vol. 1, pp. I–49–52.

[10] Michal Irani and Shmuel Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *Journal of Visual Communication,Image Representation*, vol. 4, no. 4, pp. 324335, 2002.

[11] Shengyang Dai, Mei Han, Ying Wu, and Yihong Gong, "Bilateral back-projection for single image super resolution," in *Multimedia and Expo, 2007 IEEE International Conference on*, July 2007, pp. 1039–1042.

[12] Lingfeng Wang, Huaiyu Wu, and Chunhong Pan, "Fast image upsampling via the displacement field," *Image Processing, IEEE Transactions on*, vol. 23, no. 12, pp. 5123–5135, Dec 2014.

[13] E. McCarthy, F. Balado, G.C.M. Slvestre, and N.J. Hurley, "A framework for soft hashing and its application to robust image hashing," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, Oct 2004, vol. 1, pp. 397–400 Vol. 1.

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

[15] 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.