

Image interpolation based on nonlocal self-similarity

Qiang Guo^{a,b,c,*}, Caiming Zhang^{a,b,c}, Qian Liu^c, Yunfeng Zhang^{a,b}, Xiaohong Shen^{a,b}

^a School of Computer Science and Technology, Shandong University of Finance and Economics, East Erhuan Road, Jinan, 250014, China

^b Shandong Provincial Key Laboratory of Digital Media Technology, East Erhuan Road, Jinan, 250014, China

^c School of Computer Science and Technology, Shandong University, Shunhua Road, Jinan, 250101, China

*Corresponding author, e-mail: qguo2010@126.com

Received 30 Jan 2013

Accepted 27 Dec 2013

ABSTRACT: This paper presents an efficient method for the resolution enhancement of low resolution (LR) images. Inspired by the great success of nonlocal filtering, the proposed method interpolates a missing pixel by a nonlocal estimation which takes into account the self-similarity of images and the geometric duality between LR images and high resolution (HR) images. This method firstly exploits the SAI method to generate an initial HR image of the given LR image. Then the initial HR image is further improved by a nonlocal estimation that uses the given LR image as a guiding image. Pixel correction is applied to reduce the distortion of the HR estimate. Experimental results demonstrate that the proposed method provides competitive performance.

KEYWORDS: downsampling, image filtering, nonlocal estimation

INTRODUCTION

Image interpolation, which reconstructs a high resolution (HR) image from its low resolution (LR) version, is a classic and yet unsolved problem in image processing. It is a vital technique which has numerous applications in consumer electronics such as photographic printing, image zooming, and resolution upconversion for high definition displays. In addition, image interpolation is commonly used in medical imaging, satellite remote sensing, video surveillance, computer vision, and many other important fields.

In this paper, we consider the following observation model for LR image

$$\mathbf{y} = D\mathbf{x}$$

where the observed LR image \mathbf{y} is a downsampled version of the HR image \mathbf{x} , and D represents a non-invertible downsampling operator. In a general interpolation scheme, an interpolated pixel can be expressed by a weighted average of all the pixels in a given spatial window, which is written as

$$\mathbf{x}(i) = \sum_{j \in Z} w(d_{i,j})\mathbf{y}(j)$$

where i and j are pixel indexes, Z is a spatial window, $d_{i,j}$ is a distance measure between pixels i and j , and $w(x)$, called the interpolation kernel, is a weight to be determined from the input data.

Over the past several decades, various interpolation kernel functions, such as nearest neighbour kernel, bilinear kernel, bicubic kernel and cubic spline interpolation¹⁻³, have been suggested and are well founded on approximation theory. The simplest example of such kernel functions from a computational standpoint is the nearest neighbour interpolation weight which is of the form

$$w_n(x) = \begin{cases} 1, & 0 < |x| < 0.5, \\ 0, & 0.5 \leq |x|. \end{cases}$$

The most popular kernel function used in practice is the bicubic interpolation weight² which is defined as

$$w_c(x) = \begin{cases} (a+2)|x|^3 - (a+3)|x|^2 + 1, & 0 \leq |x| \leq 1, \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a, & 1 < |x| \leq 2, \\ 0, & 2 < |x| \end{cases}$$

where a is usually set to -0.5 or -0.75 . The main advantage of these kernel functions is their relatively low computational complexity which allows their use in real-time processing systems. However, all of them inherently assume smoothness constraints on the image and fail to adapt to varying pixel structures due to the spatial invariance of the interpolation kernel. Consequently, the resulting images using these methods sometimes suffer from visual artefacts such as blurring, ringing, and jaggies⁴.

To solve these problems, many interpolation algorithms have been developed which incorporate more a priori knowledge into image models to improve the visual quality of images. In Ref. 5, to preserve edge structures of images, Li and Orchard proposed a new edge-directed interpolation (NEDI) method that takes advantage of the geometric duality between the LR covariance and the HR covariance. This algorithm interpolates missing pixels by the estimated covariance. Zhang and Wu proposed to interpolate a missing pixel in preset multiple directions, and then compute the final estimate by weighted averaging the directional interpolation results, in which the weights are derived by linear minimum mean square-error estimation⁶. In Ref. 7, Zhang and Wu proposed a soft-decision estimation algorithm for adaptive interpolation (SAI). This interpolation method exploits a 2D autoregressive model to model the edge structures and reconstructs the missing pixels via soft-decision estimation. Rather than estimating one weight at a time in isolation, SAI estimates a group of weights jointly by computing the ordinary least squares estimate

$$\mathbf{w}_i^L = \arg \min_{\mathbf{w}_i} \left\| \mathbf{y}_i - \mathbf{y}_{i,n} \mathbf{w}_i \right\|_2^2 \quad (1)$$

where \mathbf{y}_i is the LR pixel vector in block i , $\mathbf{y}_{i,n}$ is the neighbour vector of \mathbf{y}_i , and \mathbf{w}_i is the interpolation weight vector. For each block, we can compute \mathbf{w}_i in the closed-form solution of (1) by setting the derivative to 0, which is given by

$$\mathbf{w}_i^L = \left(\mathbf{y}_{i,n}^T \mathbf{y}_{i,n} \right)^{-1} \mathbf{y}_{i,n}^T \mathbf{y}_i.$$

This method can generally achieve high visual quality images. However, it assumes the spatial correlation between the HR pixels is same as the one between the LR pixels. Unfortunately, this assumption does not always hold true for texture-rich images. When the assumption is violated, some overfitting artefacts are produced at edges and textures. The reason is that the ordinary least square estimation is not robust to outliers. That is, small numbers of outliers can dramatically affect the accuracy of interpolated pixels.

To solve the overfitting problem, some robust interpolation techniques have been proposed in the literature. Liu et al proposed a robust interpolator based on the moving least squares estimation and regularized local linear regression⁸. In Ref. 9, Hung and Siu exploit the weighted least square estimation to handle outliers, in which the interpolation weight vector is computed from

$$\mathbf{w}_i^{WL} = \arg \min_{\mathbf{w}_i} \left\| \mathbf{s}_i^{1/2} \left(\mathbf{y}_i - \mathbf{y}_{i,n} \mathbf{w}_i \right) \right\|_2^2$$

where \mathbf{s}_i is a diagonal matrix for weighting, which is defined by

$$\mathbf{s}_i = \exp \left(- \frac{2 \|\mathbf{y}_i - \mathbf{y}_{i,n}\|_2}{\sigma_1} - \frac{2 \|\mathbf{z}_i - \mathbf{z}_{i,n}\|_2}{\sigma_2} \right)$$

where \mathbf{z}_i and $\mathbf{z}_{i,n}$, respectively, represent the pixel coordinates of \mathbf{y}_i and $\mathbf{y}_{i,n}$, and the parameters σ_1 and σ_2 adjust the intensity similarity and the spatial similarity, respectively. Although these methods create better image quality than SAI in both subjective and objective criteria, they still produce visual artefacts in some cases and have large memory and CPU time requirements.

Recently, sparsity priors have attracted more attention and found many successful applications in the image processing field. In Ref. 10, a sparsity-based interpolation method was introduced to enhance image resolution by estimating preserved high frequency information from the LR image in the wavelet domain. Following this scheme, Mueller et al proposed an iterative image interpolation method based on the wavelet and contourlet transforms, which improves the visual quality of resulting images by enforcing a sparsity constraint on the transform coefficients¹¹. Demirel and Anbarjafari proposed to interpolate the LR image based on hybrid wavelet transforms. This method uses a discrete wavelet transform to decompose the given LR image into different subbands. The high frequency subbands are interpolated by the bicubic interpolation algorithm. Then the stationary wavelet transform is used to correct the estimated high frequency coefficients. Corrected subbands are combined by the inverse discrete wavelet transform to achieve the HR image¹². These methods based on sparsity priors assume that the LR image is a low-pass output of the wavelet transform applied to the HR image. However, this assumption is not tenable because the LR image also contains high frequency information. In practice, therefore, sparsity-based interpolation methods are inefficient as the resulting images tend to be oversmoothed.

The principal challenge of image interpolation is how to preserve edge structures when increasing the resolution of an image. Inspired by the great success of nonlocal mean filtering for image denoising^{13,14}, we propose an iterative nonlocal interpolation method to overcome the above difficulty. The proposed method interpolates a missing pixel by a nonlocal estimation, which uses the self-similarity of natural images and the geometric duality between the LR image and the HR one.

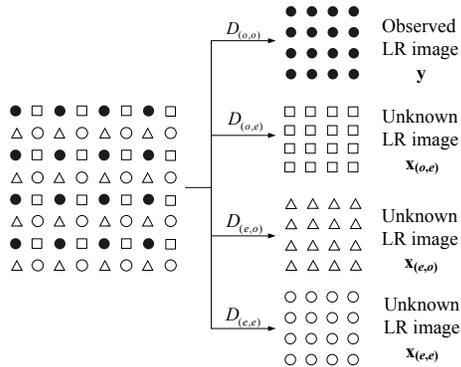


Fig. 1 Graphical illustration of an observed LR image from an HR image by downsampling with a decimating factor of two. The solid dots are the observed LR image pixels, the circles, squares and triangles, respectively, represent the missing HR pixels in different positions.

PROPOSED METHOD

Without the loss of generality, we assume that the observed LR image y is generated by decimating the ground-truth HR image x by a factor of two. Specifically, the downsampling is applied row-wise then column-wise in the ground-truth HR image. Let $D_{o,o}$ denote the operator decimating the pixels in odd rows and odd columns. Similarly, $D_{o,e}$, $D_{e,o}$ and $D_{e,e}$ represent odd-row and even-column decimation, even-row and odd-column decimation, and even-row and even-column decimation, respectively. By applying these four decimation operators on an HR image successively, four LR images are produced, as illustrated in Fig. 1. Let the observed LR image y be generated by $D_{o,o}$ and other three unknown LR images be denoted by $x_{o,e}$, $x_{e,o}$ and $x_{e,e}$, respectively. For simplicity of notation, we use x_u , ($u = 1, 2, 3$) to represent three unknown LR images. Our goal is to estimate x_u ($u = 1, 2, 3$) from y .

Nonlocal self-similar interpolation (NSSI)

In natural images, there are often many similar patterns and features. This redundancy in natural images, especially in textured patches, has been used in image denoising. Although the aforementioned LR images are different from each other, they are very similar with respect to pixel intensities and the geometric structure. Fig. 2 shows four LR images extracted from the Lena image by the aforementioned decimation operators, and illustrates the photometric and structure similarities among these LR images.

In essence, the goal of image interpolation is to estimate the other three unknown LR images by

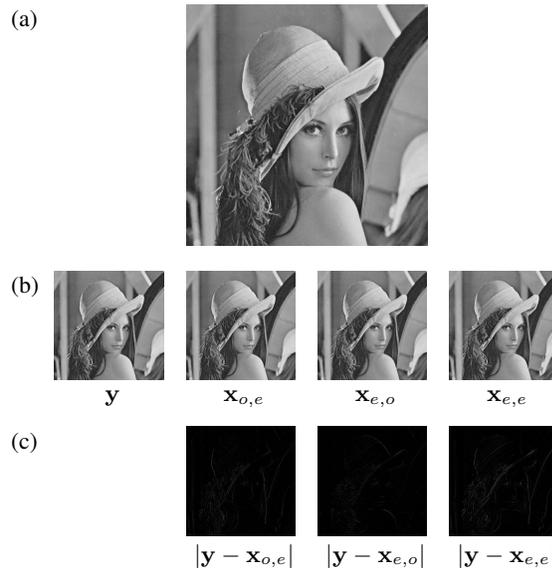


Fig. 2 Illustration of the similarity among LR images generated by different decimation operators: (a) original HR image, (b) four LR images, (c) the absolute errors between the observed LR image y and the unknown ones x_u , ($u = 1, 2, 3$).

using the observed one. In this paper, therefore, we develop an efficient interpolation method which takes into account the self-similarity of natural images to improve the quality of interpolated images. Since all pixels in x_u are missing, an initial interpolation is needed to start this process. We use the SAI method to generate the initial result. For simplification of notation, we denote the initial result of SAI by \hat{x}_u^0 .

Unlike conventional interpolation methods using the local weights, the proposed method uses nonlocal weights to estimate the missing pixel, in which the missing pixel is estimated as the weighted average of its neighbourhoods

$$\hat{x}_u(i) = \frac{\sum_{j \in S_i} w(i, j) y(j)}{M(i)} \tag{2}$$

where S_i is the search region around i , $M(i)$ is a normalizing term, $M(i) = \sum_{j \in S_i} w(i, j)$, and $w(i, j)$ is the nonlocal weight. The nonlocal weight is defined as

$$w(i, j) = \exp \left(- \frac{\|x_u(N(i)) - y(N(j))\|^2}{2h^2} \right)$$

where $N(i)$ denotes the square patch centred at i , $\|\cdot\|$ represents the Euclidean distance between $x_u(N(i))$ and $y(N(i))$, and h is a parameter that controls the decay of the exponential expression. The nonlocal

weight compares the neighbourhoods around pixels i and j . If the pixel i is similar to the pixel j , the nonlocal weight $w(i, j)$ in the average will be large. Conversely, if the two pixels differ substantially, $w(i, j)$ will be very small. Since the nonlocal weights make use of the redundant information among similar pixels, using (2) gives better results than the methods with local weights.

Pixel correction

The HR image obtained by the aforementioned non-local estimation is not perfect as it contains some structural distortions. Fig. 3 displays the structural distortions for the test image Mandrill. These structural distortions can be considered as structural noise. The BM3D filter, which exploits both self-similarities and sparsity in natural images, is an efficient tool for noise reduction¹⁵. From Fig. 3c we can see that BM3D is also an effective way to reduce the structural noise caused by the nonlocal estimation. BM3D contains two transform-domain filtering stages, namely, hard thresholding filtering and Wiener filtering. The first stage yields a basic estimate of the noisy image by applying hard thresholding filtering in the 3D transform domain to remove most of the noise, and the second stage uses the basic estimate as the reference image to further refine the output of the first stage by Wiener filtering.

Although this two-stage scheme can result in BM3D achieving the best performance, experiments show that the Wiener filtering in the second stage blurs edges and causes visible artefacts. Hence for the purpose of image interpolation, we only perform the hard thresholding filtering stage of the BM3D algorithm. The procedures (for detailed descriptions, see Ref. 15) are as follows:

- (i) *Grouping by block classification.* Use the block matching method to classify image blocks and form a 3D block group.
- (ii) *Filtering 3D transform coefficients by hard thresholding.* Apply a hybrid wavelet transform to the formed 3D block groups, perform hard thresholding on the transform coefficients to reduce the noise, and produce estimates of all blocks by applying the inverse 3D transform to the filtered coefficients.
- (iii) *Aggregation by weighted averaging.* Generate the final estimate of the noise-free image by weighted averaging over all filtered blocks.

This filtering process can be formulated by

$$\hat{x}_{u,f} = A \left(W^{-1} \left(T \left(W \left(G(\hat{x}) \right) \right) \right) \right) \quad (3)$$

where W , G and A represent the hybrid 3D wavelet transform, block grouping, and block aggregation procedures, respectively, and T is the hard thresholding operator which is defined by

$$T(x) = \begin{cases} x, & |x| < \lambda, \\ 0, & \text{otherwise,} \end{cases}$$

where λ is the given threshold.

Since the blocks are overlapping in Step i of BM3D, each block can belong to several groups. After filtering 3D transform coefficients by hard thresholding on block groups, more than one block estimate can be located at the same coordinate. This redundancy can prevent the estimates of three unknown LR images from becoming the same copy of the observed LR image. And the aggregation by weighted averaging over the overlap block estimates can ensure the compatibility of them.

Iterative NSSI

The filtered estimate of the HR image can be used as a guiding image to iteratively update the nonlocal weight. Hence the iterative nonlocal estimate of the HR image is given by modifying (2), which is expressed as

$$\hat{x}_u^k(i) = \frac{\sum_{j \in S_i} w^k(i, j) \mathbf{y}(j)}{M(i)}, \quad (u = 1, 2, 3) \quad (4)$$

where k is the iteration number ($k \geq 1$) and

$$w^k(i, j) = \exp \left(- \frac{\|\mathbf{x}_{u,f}^{k-1}(N(i)) - \mathbf{y}(N(j))\|^2}{2h^2} \right).$$

The aforementioned iterative process is summarized in the following algorithm.

Algorithm 1

- Step 1: For a given LR image \mathbf{y} , use the SAI interpolation method to generate an initial estimate of the HR image \hat{x}_u^0 , and let $k = 1$.
- Step 2: Use \mathbf{y} as a guiding image to calculate the nonlocal estimate \hat{x}_u^k of the HR image by (4).
- Step 3: Correct \hat{x}_u^k to improve the quality of image and produce $\hat{x}_{u,f}^k$ by (3).
- Step 4: Update $\hat{x}_u^k = \hat{x}_{u,f}^k$ and $k = k + 1$.
- Step 5: Repeat Steps 2–4 until the resulting HR image satisfies the stopping criterion.

At the beginning of an iteration of the proposed algorithm, we apply the SAI to generate an initial estimate of the HR image. The given LR image \mathbf{y} is used as a guiding image to estimate the three other

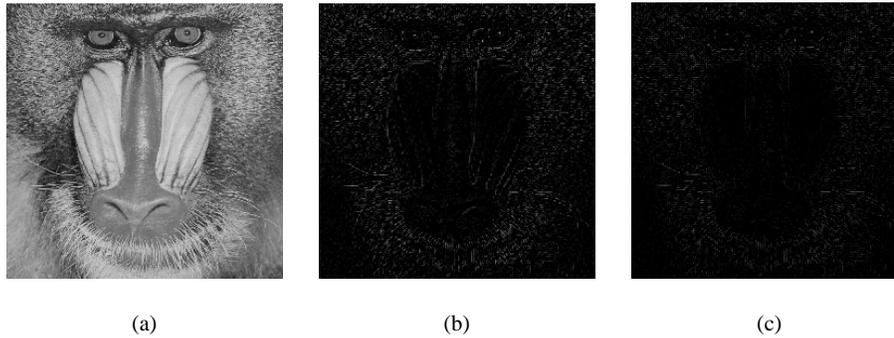


Fig. 3 Structural distortion: (a) original image, (b) distortion of image interpolated by the nonlocal estimation, (c) distortion of image interpolated after applying BM3D.

LR images \mathbf{x}_u , ($u = 1, 2, 3$). The iterative NSSI terminates when the resulting HR image satisfies the following stopping criterion

$$\frac{\|\hat{\mathbf{x}}_u^k - \hat{\mathbf{x}}_u^{k-1}\|}{\|\hat{\mathbf{x}}_u^k\|} \leq \tau$$

where τ is a preset tolerance. We use the relative error as the stopping criterion because it is a simple and common way to decide when to terminate the iteration. This stopping criterion means that the iteration is terminated when the relative error between the previous estimate and current estimate is less than a preset tolerance value.

EXPERIMENTAL RESULTS

In this section, extensive experiments were performed to demonstrate the superiority of the proposed method. We compare the proposed method with existing interpolation methods, namely, bicubic interpolation (BCI)², new edge-directed interpolation (NEDI)⁵, iterative curve based interpolation (ICBI)¹⁶, and soft-decision adaptive interpolation (SAI)⁷. The proposed method is implemented in MATLAB 7.0 and performed on an Intel-i7, 2.93 GHz computer system. The implementations of ICBI, SAI and NEDI are available at www.andreagiachetti.it/icbi, www.ece.mcmaster.ca/~7Exwu, and www.csee.wvu.edu/~7Exinl, respectively.

For the experiments, we use a test images set including eight widely used grey-level images. As illustrated in Fig. 1, LR images are generated by downsampling corresponding test images by a factor of two in both row and column dimensions. The HR images are reconstructed by the methods mentioned above. The performance of these methods is evaluated by visual quality and peak signal-to-noise ratio (PSNR), respectively. For the original HR image \mathbf{x} and the interpolated image $\hat{\mathbf{x}}$, the PSNR of them is

Table 1 PSNR (dB) of test images using different methods.

Images	BCI	NEDI	SAI	ICBI	Our
Monarch	25.27	27.97	29.57	29.12	29.66
Mandrill	19.77	20.71	20.60	19.75	20.76
Lena	30.16	33.74	34.77	34.08	34.84
Barbara	21.41	21.50	22.10	21.33	22.22
Cameraman	23.74	25.47	26.02	25.33	26.14
Man	24.43	25.94	26.45	25.76	26.53
Goldhill	28.67	30.40	30.80	30.16	30.85
House	29.31	31.60	32.91	32.02	33.06

defined as

$$\text{PSNR} = 20 \log_{10} \left(\frac{255}{\text{MSE}} \right)$$

where

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N [\mathbf{x}(i) - \hat{\mathbf{x}}(i)]^2.$$

For quantitative comparisons, the PSNR results generated by all methods for test images are listed in Table 1. In all test images, the proposed method outperforms BCI, NEDI, ICBI and SAI in terms of PSNR. Fig. 4 and Fig. 5 compare the visual results of these interpolation methods on the Monarch and House test images, respectively. It can be observed that all methods exhibit different visual characteristics, particularly in texture areas. The BCI method is sensitive to noise and results in lower fidelity images than the other methods, as justified in Table 1. The NEDI and ICBI methods, respectively, produce prominent directional artefacts and jaggies near sharp edges. The SAI method also exhibits directional artefacts in the smooth region, while it performs better than NEDI, ICBI, and BCI. The proposed method

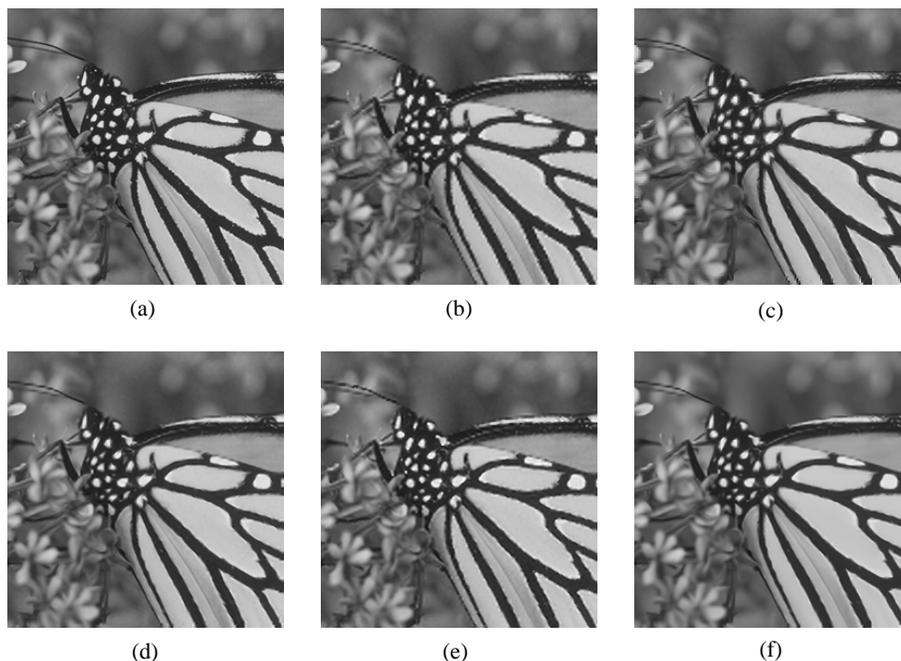


Fig. 4 Performance comparison on Monarch image: (a) original image, (b) BCI, (c) NEDI, (d) SAI, (e) ICBI, (f) the proposed method.

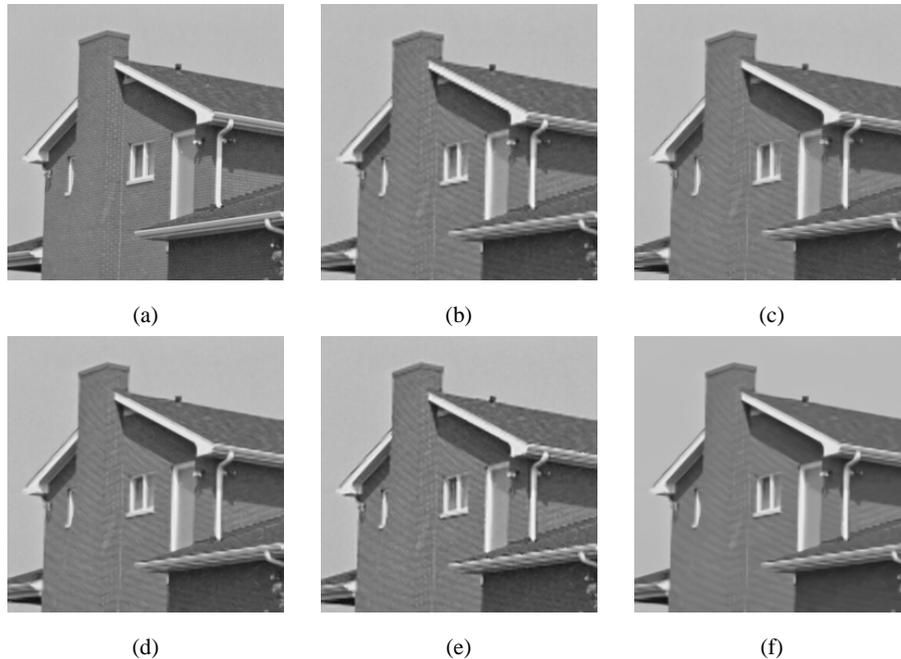


Fig. 5 Performance comparison on House image: (a) original image, (b) BCI, (c) NEDI, (d) SAI, (e) ICBI, (f) the proposed method.

produces the most visually pleasant results among all competing methods, and it exhibits less aliasing artefacts than other methods. Fig. 6 shows the evolution of the PSNR along the iterations, which is produced

by the proposed method on the test image Mandrill. In Fig. 6, experimental convergence is evident for the proposed method.

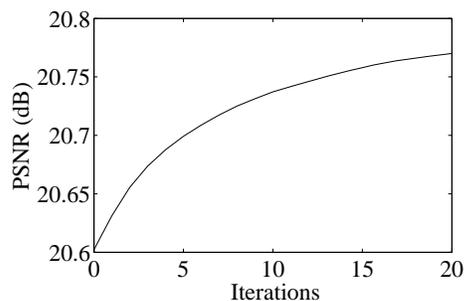


Fig. 6 PSNR Value as a function of the iterations for the proposed method.

CONCLUSION AND DISCUSSION

This paper presents an iterative image interpolation method which uses the SAI interpolation method to generate an initial image of the HR image. The initial HR image is further improved by nonlocal estimation and pixel correction, which use the self-similarity of natural images. Experimental results show that the proposed method is superior to conventional methods.

One possible future direction is to exploit the adaptive parameter h that controls the decay of the exponential weights according to local image statistics. Alternatively, we may use Stein's unbiased risk estimator, which is widely used as an efficient parameter selection method^{17–19}, to derive the adaptive parameter. In addition, we could reduce the computational burden by optimizing the whole process further, particularly the available speed-up methods for computing nonlocal weights.

Acknowledgements: This work was partially supported by Projects of International Cooperation and Exchanges NSFC (61020106001), National Natural Science Foundation of China (61202150, 61173174, 61073080, 61202364), Shandong Province Natural Science Foundation (ZR2011FL028), China Postdoctoral Science Foundation (2013M531600), Projects of Shandong Province Higher Educational Science and Technology Programme (J11LG77), Programme for Scientific Research Innovation Team in Colleges and Universities of Shandong Province, and Doctoral Foundation of Shandong University of Finance and Economics. The authors thank the anonymous reviewers for their valuable comments.

REFERENCES

- Jain R, Kasturi R, Schunck BG (1995) *Machine Vision* McGraw-Hill, New York.
- Keys RG (1981) Cubic convolution interpolation for digital image processing. *IEEE Trans Acoust Speech Signal Process* **29**, 1153–60.
- Hou HS, Andrews HC (1978) Cubic splines for image interpolation and digital filtering. *IEEE Trans Acoust Speech Signal Process* **26**, 508–17.
- Chang SG, Cvetković Z, Vetterli M (2006) Locally adaptive wavelet-based image interpolation. *IEEE Trans Image Process* **15**, 1471–85.
- Li X, Orchard MT (2001) New edge-directed interpolation. *IEEE Trans Image Process* **10**, 1521–7.
- Zhang L, Wu X (2006) An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Trans Image Process* **15**, 2226–38.
- Zhang X, Wu X (2008) Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation. *IEEE Trans Image Process* **17**, 887–96.
- Liu X, Xiong R, Ma S, Gao W (2011) Image interpolation via regularized local linear regression. *IEEE Trans Image Process* **20**, 3455–69.
- Hung KK, Siu WC (2012) Robust soft-decision interpolation using weighted least squares. *IEEE Trans Image Process* **21**, 1061–9.
- Temizel A, Vlachos T (2006) Wavelet domain image resolution enhancement. *IEE Proc Vis Image Signal Process* **153**, 25–30.
- Mueller N, Lu Y, Do MN (2007) Image interpolation using multiscale geometric representation. *Proc SPIE* **6498**, 64980A.
- Demirel H, Anbarjafari G (2011) Image resolution enhancement by using discrete and stationary wavelet decomposition. *IEEE Trans Image Process* **20**, 1458–60.
- Buades A, Coll B, Morel JM (2005) A review of image denoising algorithms, with a new one. *Multiscale Model Simulat* **4**, 490–530.
- Buades A, Coll B, Morel JM (2008) Nonlocal image and movie denoising. *Int J Comput Vis* **76**, 123–39.
- Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans Image Process* **16**, 2080–95.
- Giachetti A, Asuni N (2011) Real-time artifact-free image upscaling. *IEEE Trans Image Process* **20**, 2760–8.
- Van De Ville D, Kocher M (2012) Nonlocal means with dimensionality reduction and SURE-based parameter selection. *IEEE Trans Image Process* **20**, 2683–90.
- Deledalle CA, Duval V, Salmon J (2012) Non-local methods with shape-adaptive patches (NLM-SAP). *J Math Imag Vis* **43**, 103–20.
- Guo Q, Zhang C (2012) A noise reduction approach based on Stein's unbiased risk estimate. *Sci Asia* **38**, 207–11.