Multiscale Semilocal Interpolation With Antialiasing

Kai Guo, Xiaokang Yang, *Senior Member, IEEE*, Hongyuan Zha, Weiyao Lin, and Songyu Yu

Abstract—**Aliasing is a common artifact in low-resolution (LR) images generated by a downsampling process. Recovering the original high-resolution image from its LR counterpart while at the same time removing the aliasing artifacts is a challenging image interpolation problem. Since a natural image normally contains redundant similar patches, the values of missing pixels can be available at texture-relevant LR pixels. Based on this, we propose an iterative multiscale semilocal interpolation method that can effectively address the aliasing problem. The proposed method estimates each missing pixel from a set of texture-relevant semilocal LR pixels with the texture similarity iteratively measured from a sequence of patches of varying sizes. Specifically, in each iteration, top texture-relevant LR pixels are used to construct a data fidelity term in a maximum *a posteriori* estimation, and a bilateral total variation is used as the regularization term. Experimental results compared with existing interpolation methods demonstrate that our method can not only substantially alleviate the aliasing problem but also produce better results across a wide range of scenes both in terms of quantitative evaluation and subjective visual quality.**

Index Terms—**Antialiasing, image interpolation, iterative multiscale, semilocal.**

I. INTRODUCTION

IMAGE INTERPOLATION addresses the problem of obtaining a high-resolution (HR) image from its low-resolution (LR) counterpart. It is fundamental to many real-world applications, such as biometrics, surveillance, and security, where the resolution of the captured images tend to be very low and the apparent aliasing effects often appear due to the limited number of charge-coupled-device pixels used in commercial digital cameras. Based on the Nyquist–Shannon sampling theorem [1], [2], if the sampling frequency is smaller than twice the maximum frequency of HR image, all the image information above half the sampling frequency are removed. Then, the

K. Guo was with the Institute of Image Communication and Information Processing, Shanghai Key Laboratory of Digital Media Processing and Transmission, Shanghai Jiao Tong University, Shanghai 200240, China. He is now with DMC Research and Development Center, Samsung Electronics, Suwon 443-742, Korea (e-mail: visionkai@gmail.com).

X. Yang, W. Lin, and S. Yu are with the Institute of Image Communication and Information Processing, Shanghai Key Laboratory of Digital Media Processing and Transmission, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xkyang@sjtu.edu.cn; wylin@sjtu.edu.cn; syyu@sjtu.edu.cn).

H. Zha is with the College of Computing, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: zha@cc.gatech.edu).
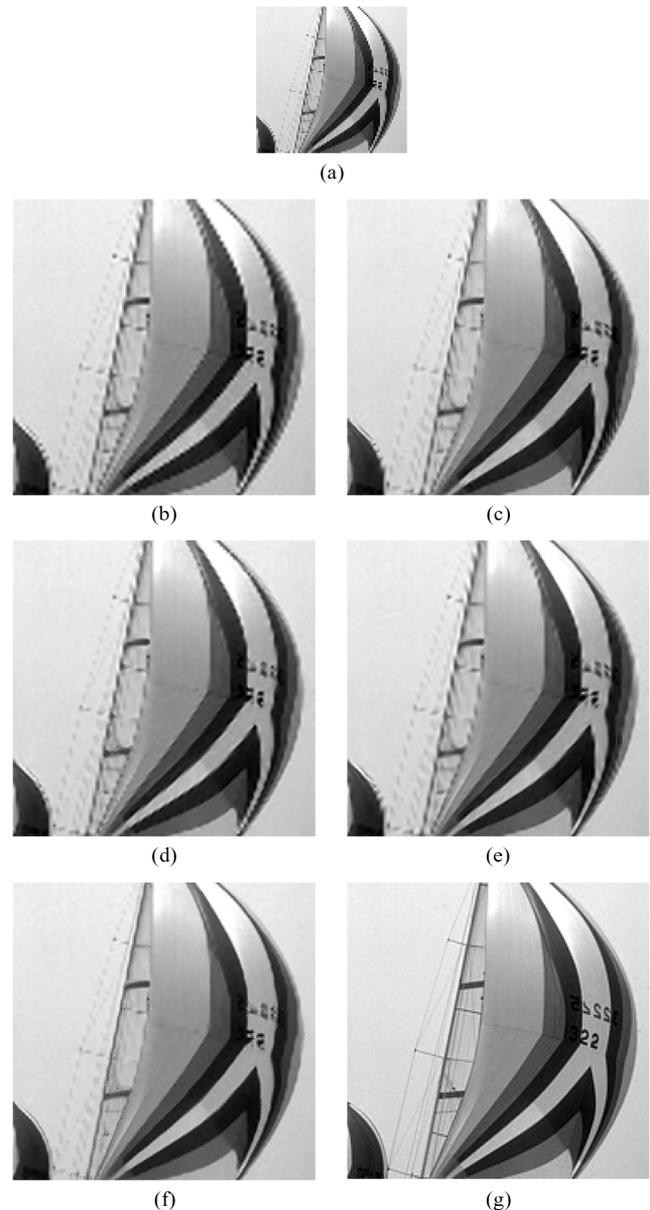
Fig. 1. Comparison of different methods on the image "Sail" [5]. (a) Observed LR image. (b) Bicubic interpolation. (c) NEDI method [6]. (d) Method in [7]. (e) SAI method [8]. (f) Our method. (g) Ground-truth HR image.

LR image is corrupted by aliasing [3] (such as the mast and jib occurred in the image "Sail" in [see Fig. 1(a)].

Commonly used linear interpolation methods such as the bicubic method [4] perform interpolation based on the homogeneous assumption and are therefore not capable of adapting to various image structures, often producing blurred edges. This motivates many scene-adaptive image interpolation methods

that try to explore dominant image structures and edge information of the LR image. Here, we mention a few examples: Jensen and Anastassiou [9] directly estimated the edge orientation to regularize the interpolated image pixels. Li and Orchard [6] assumed the geometric duality between the LR and HR images and used it to guide the interpolation of missing pixels from a weighted combination of neighboring pixels of the LR image. Zhang and Wu [7] proposed to interpolate the missing pixels along multiple directions first and then fuse them by minimizing mean square error. Zhang and Wu [8] estimated the missing pixels using a 2-D piecewise autoregressive model, which is adaptively learned from the local neighborhood of the LR image. Generally, those edge-directed methods can work well when the edge information of LR image is correctly estimated. However, for observed LR image with aliasing artifacts, these methods cannot recover ground-truth content well [such as the mast and jib occurred in the image "Sail" in Fig. 1(b)–(e)], and may often interpolate overfitting artifacts, particularly at fine textures. Furthermore, other researchers tried to perform image superresolution using training database of LR and HR patch pairs [10]–[12]. However, since these methods are based on an ideal assumption that there are almost no aliasing artifacts in LR images, they still cannot efficiently handle aliasing problems.

Natural images often contain various redundant similar patches, even within a single image; thus, intuitively, patches of corrupted observed images can be restored from a set of similar candidates. This observation has been widely adopted in some applications such as image restoration and denoising [13]–[16]. For example, the nonlocal mean method adaptively estimates image pixel values using the weighted average of the pixel values whose centered regions are similar to the region centered by the pixel being estimated [17]. Inspired by nonlocal means, improvements also consider iteratively increasing the neighborhood size [18] and using varying neighborhood sizes [19] for denoising. The success of those methods very much depend on the observation that there exist redundant similar patches within a single image.

The focus of this paper is to develop image interpolation methods that can effectively address the aliasing problem. Since a natural image normally contains redundant similar patches, the values of missing pixels can be available at texture-relevant LR pixels. Based on this basic idea, we propose to recover each missing pixel with antialiasing from a set of texture-relevant LR pixels within its neighborhood, whose pixel-centered patches are similar to the corresponding patch centered by the missing pixel. Compared with the nonlocal idea, finding the similar patches within a neighborhood substantially reduces computational complexity [17], and this type of methods are generally known as semilocal methods [20].

Specifically, we propose an iterative multiscale interpolation method to estimate each missing pixel from a set of texture-relevant semilocal LR pixels, where the texture similarity is measured from large to small patch sizes iteratively. The similarity measurement with large patch sizes can alleviate the influence of aliasing artifacts of the LR image, whereas the measurement with small patch sizes can avoid overfitting effects for fine and dense textures. To further enhance performance, the outputs of all previous iterations are grouped together as inputs for the
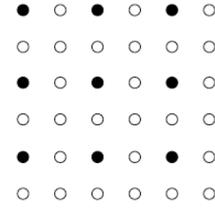

Fig. 2. Generation of an LR image from an HR image by downsampling. (Solid dots) LR image pixels x and (circles) missing HR pixels y.

next smaller scale interpolation: This can not only inherit the advantages of using large-scale patches but also filter out inaccurate results of previous outputs and gradually recover the finer details. In each iteration, top texture-relevant semilocal LR pixels are selected to construct a data fidelity term in a maximum *a posteriori* (MAP) estimation, and a bilateral total variation (TV) [21] is used as the regularization term. Experimental results compared with existing interpolation methods verify that our method can alleviate aliasing artifacts substantially, and at the same time, it also outperforms other methods both in terms of quantitative evaluation and subjective visual quality across a wide range of images.

The rest of this paper is organized as follows: Section II introduces the basic idea for antialiasing. Section III presents the semilocal interpolation method in single scale and single iteration. Section IV describes proposed iterative multiscale semilocal interpolation method. Section V describes the experimental results and comparisons with existing methods. Conclusions are drawn in Section VI.

II. BASIC IDEA FOR ANTIALIASING

There are many downsampling processes that can potentially generate an observed LR image [22]–[24]. In this paper, to focus on the aliasing problem, we assume that the LR image is generated by direct downsampling from the ground-truth HR image.[1]

LR image I_l is assumed to be directly downsampled from ground-truth HR image I_h, as illustrated in Fig. 2. Let $x_i \in I_l$ and $y_j \in I_h$ be the pixels of image I_l and the missing pixels in image I_h that need to be interpolated, respectively. Let Y represent the missing area.

After downsampling, aliasing artifacts appear in the area of high frequencies, where the content and edges are distorted compared with the ground truth, as shown in Fig. 3(a). Since a natural image normally contains redundant similar patches, the values of missing pixels can be available at texture-relevant LR pixels. As shown in Fig. 3(b), the value of missing pixel y_b is available at LR pixel x_a, where the patch centered by x_a is texture relevant, with the patch centered by y_b. Based on this, we can try to interpolate the missing pixels with antialiasing from a set of texture-relevant semilocal LR pixels. That is, we try to recover each missing pixel y_j based on a set of LR pixels x_i, where the textures of the neighborhood patches centered by x_i are similar or relevant to that of the neighborhood patch centered by y_j.

---

[1]As will be demonstrated later in the experimental part, our proposed algorithm, to some extent, can be also applied to some real-world LR images generated with more complicated downsampling process.
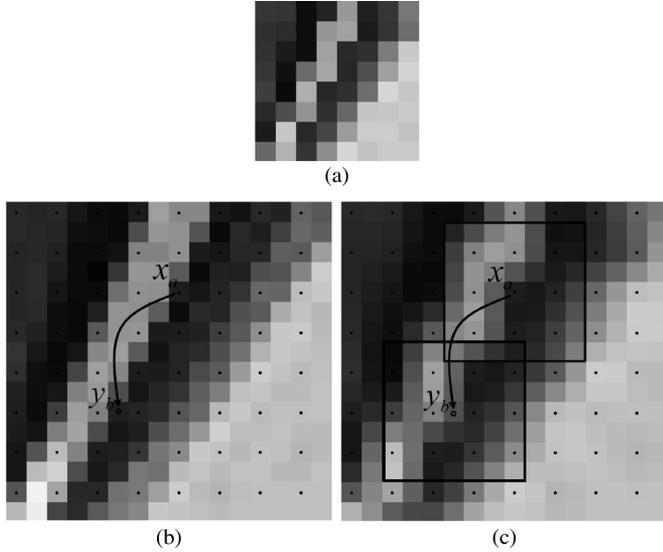
Fig. 3. Illustration of the idea for antialiasing. (Solid dots) Pixels of the observed LR image and (solid square) the pixel-centered patch. (a) Observed LR image with aliasing. (b) Ground-truth HR image. The value of missing pixel y_b is available at LR pixel x_a, where x_a is texture relevant with y_b. (c) Bicubic interpolation with aliasing artifacts. The interpolated value at pixel y_b greatly differs from the ground truth, but we can still find LR pixel x_a as the texture-relevant pixel for y_b through pixel-centered patch matching.

Since all pixels y_j are missing, initial interpolation $\tilde{I}_h$ is needed to start this process. Under the ideal but unrealistic assumption that $\tilde{I}_h$ is the same as ground truth I_h, each missing pixel y_j can be filled by its most similar LR pixel x_i. However, obtaining very accurate initial interpolation from scratch is difficult. Instead, we use bicubic interpolation as initial interpolation $\tilde{I}_h$ because it can recover the missing area of low frequencies well [4]. Unfortunately, the aliasing artifacts occurred in the area of high frequencies still remain. As illustrated in Fig. 3(c), the interpolated value at pixel y_b greatly differs from the ground truth.

To jointly tackle the aliasing problem and the inaccuracy of initial interpolation $\tilde{I}_h$, we propose an iterative multiscale semilocal interpolation method that explores texture-relevant semilocal LR pixels in an iterative multiscale way. This method has two main ideas:

- *Pixel-centered patch matching.* For each missing pixel, neighboring pixels within a patch centered by the current missing pixel can be used as the reference to find texture-relevant LR pixels. Since most natural images have an exponentially decaying power spectrum [25], the aliasing artifacts only appear in a small proportion of pixels in high-frequency areas. Additionally, most image pixels in low-frequency areas can be interpolated well through initial bicubic interpolation. With the appropriate patch size, the neighboring low-frequency pixels can overcome the influence of aliasing artifacts when finding texture-relevant LR pixels. As an intuitive illustration in Fig. 3(c), the ground-truth value at pixel y_b can be found from pixel x_a, where the texture relevance is measured using a pixel-centered patch.
- *Iterative multiscale interpolation.* From the antialiasing standpoint, it is desirable to prefer larger patch sizes so

that more neighboring pixels can be used as the reference to find texture-relevant LR pixels. On the other hand, fine textures in a natural image require smaller patch sizes to avoid mismatching. However, it is difficult to determine the suitable patch size from the observed LR image with aliasing. Hence, an iterative multiscale interpolation procedure is proposed to integrate advantages from both the large- and small-scale patch matching.

III. SEMILOCAL INTERPOLATION IN SINGLE SCALE AND SINGLE ITERATION

Based on the observation that natural image has redundant similar patches and the analysis of antialiasing in Section II, we propose an iterative multiscale semilocal interpolation procedure for antialiasing and recovering high-quality HR image. Here, we will introduce semilocal interpolation of only a single iteration and a single patch scale. In the following, we will first describe the way to measure patch similarity, followed by the MAP estimation for missing pixel value calculation.

A. Measuring of Patch Similarity

As mentioned, the aliasing artifacts still remain in the initially interpolated HR image $\tilde{I}_h$. Our task is to find reliable texture-relevant LR pixels for each missing pixel y_j by pixel-centered patch matching on the HR grid. In this paper, in order to decrease the influence of initially inaccurate interpolation, we extend the l^2 distance and propose a masked l^2 distance. In the masked l^2 distance, we introduce mask H and only calculate the l^2 distance between the LR pixels in patch $u(y_j)$ and the corresponding pixels in patch $u(x_i)$ (as shown in Fig. 4), where $u(\cdot)$ represents pixel-centered patch operator on the HR grid. Mask H is determined by patch $u(y_j)$ and defined as follows:

$$H_q = \begin{cases} 1, & u(y_j)_q \in I_l \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where index q represents the pixel position in mask H. If the pixel in patch $u(y_j)$ belongs to the observed LR image I_l, the value of mask H at this position is set as 1. Otherwise, the corresponding value of H is set as 0. After applying mask H, and together with the Gaussian weighting, the weight value of semilocal neighboring LR pixel x_i for the missing pixel y_j is calculated as follows:

$$w(x_i, y_j) = \exp\left(-\frac{\|G \odot H \odot (u(x_i) - u(y_j))\|_2^2}{2\sigma_d^2}\right),$$
$$x_i \in \Delta(y_j) \quad (2)$$

where $\odot$ operates the element-by-element product. G is a Gaussian kernel with standard deviation σ_g. σ_d is the standard deviation of $w(x_i, y_j)$. $\Delta(y_j)$ allows one to restrict the search space of texture-relevant LR pixels x_i for each missing pixel y_j. $\Delta(y_j)$ can be as large as the whole image; this will find all possible appropriate LR pixels. In practice, in view of time complexity, we restrict the search space in the semilocal neighborhood. Based on the experimental results, we set the search window size as 17×17 that measured on the HR grid.

Standard deviation σ_g takes into account the distance between the central pixel and other pixels within a patch when
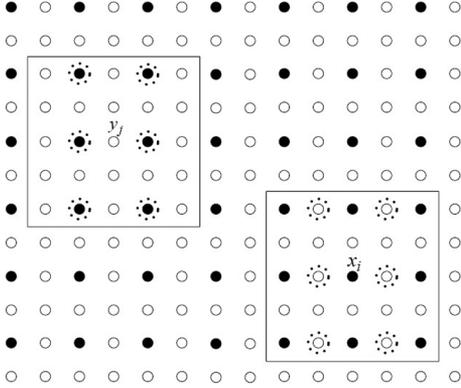
Fig. 4. Illustration of the similarity measurement between missing pixel y_j-centered patch and LR pixel x_i-centered patch. (Solid square) The pixel-centered patch and (dot circle) the valid pixel to measure patch similarity.

performing patch matching. In this paper, we set σ_g as the linear function of the patch size. Basically, the spatially closer pixels are more dependent, and the pixels closer to the center pixel should have larger weights. In addition, as subsequently described, we will calculate masked l^2 distance with multiscale patch sizes. In order to give full play of the multiscale, we set σ_g as the linear function of the patch size. Assume that the patch size is $p \times p$, and σ_g is set as $0.5p$ based on the experimental results.

Standard deviation σ_d refines the influence of masked l^2 distance on final weight w. In our algorithm, σ_d is set to be adaptive with patch size $p \times p$. For each missing pixel y_j, σ_d works as a soft Gaussian threshold to distinguish the texture-relevant LR pixels from all the neighboring LR pixels. Normally, the initially interpolated values of missing pixels are inaccurate and will bring the disturbance into the calculation of the masked l^2 distance. For the sake of analysis, we assume the difference between the initially interpolated value of y_j and the ground-truth one to obey a Gaussian distribution with the mean zero and standard variance σ_n. Then, the masked l^2 distance between two similar patches $u(y_j)$ and $u(x_i)$ obeys a noncentral chi-square distribution with mean $d^{\text{ideal}} + cp^2\sigma_n^2$, where d^{ideal} is the ground-truth distance between $u(y_j)$ and $u(x_i)$, and c is a constant value. In addition, as the patch size increases, the patch similarities within a local area will decrease, and d^{ideal} between similar patches will increase. Based on the above analysis, σ_d can be set as the linear function of p^2. In this paper, based on the experimental results, σ_d is set as $0.0048p^2$.

From (2), we can see that masked patch $H \odot u(y_j)$ is used as the standard patch, independent of the initial interpolation. All interpolated semilocal masked patches $H \odot u(x_i)$ are compared with the standard patch. As described later in our multiscale procedure, to incorporate the advantages of both large- and small-scale semilocal interpolation, we group the outputs of all previous iterations together as inputs for the next smaller scale iteration. The interpolated values of missing pixels among the inputs are different. Masked patch $H \odot u(y_j)$ contains only LR pixels and is constant; hence, $H \odot u(y_j)$ can be used as the standard patch to compare with all neighboring masked patches $H \odot u(x_i)$, no matter which input they are from.

B. MAP Estimation

After the patch similarity is calculated, L pixels $x_l^j (l = 1, 2, \ldots, L)$ with the largest weights will be selected as the texture-relevant candidates for interpolating missing pixel y_j. In the following, we will first describe two different terms for interpolating and then describe the combination of these two terms in a MAP framework for a final interpolating in a single iteration.

1) Data Fidelity Term: As mentioned, we select top L pixels $x_l^j (l = 1, 2, \ldots, L)$ as the texture-relevant candidates for missing pixel y_j and expect that the pixel value of y_j can agree with selected candidates x_l^j. In general, this process can be modeled as the data fidelity term used in the MAP estimation, defined as

$$U_{\text{data}}(y_j) = \sum_{l=1}^{L} \bar{w}_l^j \left\| y_j - x_l^j \right\|_2^2 \qquad (3)$$

where $\bar{w}_l^j$ is the normalized weight of candidate x_l^j and the calculation is $\bar{w}_l^j = w_l^j / \sum_{l=1}^{L} w_l^j$. Note that the error norm is critical for final solution. For example, the l^1 norm will lead the reconstructed pixel value toward the median of candidate pixels, and it is verified to be robust with respect to outliers and noise [21]. In our paper, pixels $x_l^j (l = 1, 2, \ldots, L)$ do not exactly have the same value with the ground truth of y_j and always slightly fluctuate around the ground truth. Therefore, the l^2 norm is adopted to lead y_j to the mean value of $x_l^j (l = 1, 2, \ldots, L)$. L defines the number of texture-relevant candidates for each missing pixel in the data fidelity term. To achieve the smoothness of fluctuations, L cannot be small. On the other hand, the number of similar fragments is limited within the semilocal neighborhood; thus, L cannot be large. Based on the experimental results, we set L as 8 in this paper.

2) Regularization Term: Given an inaccurate initial input, texture-relevant LR pixels for all missing pixels do not always exist. If the candidate values for one missing pixel fluctuate too much, the data fidelity term at this position is not credible. Therefore, it is desirable to adopt a regularization or prior term that worked on a local neighborhood to help this pixel toward stable solution. In addition, we also want to keep the sharp edges through the regularization term.

In order to meet with these properties, we employ the bilateral TV [21] as the regularization term. The gradient of bilateral TV works as a weighted median filter. This can not only preserve sharp edges but also alleviate unwanted artifacts. The regularization term is represented as

$$U_{\text{regular}}(y_j) = \sum_{m=-N}^{N} \sum_{n=0}^{N} \alpha^{|m|+|n|} \left\| y_j - F_j(S^m S^n I_h) \right\|_1 \qquad (4)$$

where $m + n \geq 0$ and matrices S^m and S^n shift I_h by m and n pixels in horizontal and vertical directions, respectively. Operator F_j extract the pixel value at original position j from shifted image $S^m S^n I_h$. Coefficient α $(0 < \alpha < 1)$ is the spatially decaying factor. For pixel y_j, the bilateral TV term is actually the weighted sum of absolute difference between y_j and its local neighboring pixels. The gradient of bilateral TV works as a weighted median filter. N determines the local working area

size of the bilateral TV regularization, and the size of its local working area is $(2N+1) \times (2N+1)$. N cannot be large, otherwise the median filter will remove the fine textures. We set N as 2 in this paper so that it can achieve the ideal results.

3) MAP Estimation Combining Both Terms: After rewriting the data fidelity and regularization term into a matrix form, we can reconstruct the HR image by minimizing the following energy:

$$
\begin{aligned}
\tilde{I}_h &= \arg\min_{I_h} \{U_{\text{data}}(I_h) + U_{\text{regular}}(I_h)\} \\
&= \arg\min_{I_h} \left\{ \sum_{l=1}^{L} \left\| \bar{W}_l^{\frac{1}{2}} \odot (Y - X_l) \right\|_2^2 \right. \\
&\quad \left. + \lambda \sum_{m=-N}^{N} \sum_{n=0}^{N} \alpha^{|m|+|n|} \|I_h - S^m S^n I_h\|_1 \right\}
\end{aligned}
$$
(5)

where X_l and $\bar{W}_l$ are the matrix form of the lth candidates for missing pixels Y and their corresponding normalized weights, respectively. Operator $\odot$ is the element-by-element product of two matrices. λ is the tradeoff parameter to tune the ratio of the data fidelity term and the regularization term. The bilateral TV term serves as supplements to the data fidelity term, such as sharpening edges and leading the unexpected pixel value toward stable solution. It cannot be given a large weight, otherwise it will remove the fine details within its local working area $(2N+1) \times (2N+1)$. Based on the experimental results, we set λ as 0.002 in this paper. We use gradient descent to carry out the optimization, and the desired interpolated HR image is updated as

$$
\begin{aligned}
\tilde{I}_h^{k+1} = \tilde{I}_h^k - \eta M \\
\times \left\{ \sum_{l=1}^{L} \bar{W}_l \odot \left(\tilde{Y}^k - X_l \right) \right. \\
+ \lambda \sum_{m=-N}^{N} \sum_{n=0}^{N} \alpha^{|m|+|n|} (I - S^{-m} S^{-n}) \\
\left. \times \text{sign}\left(\tilde{I}_h^k - S^m S^n \tilde{I}_h^k \right) \right\}
\end{aligned}
$$
(6)

where M is the mask that sets the energy to zero for noninterpolated pixels. If the pixel belongs to the observed LR image I_l, the value of mask M at the corresponding position is set as 0. Otherwise, the value of mask M at the corresponding position is set as 1. η controls the step size in the direction of the gradient. I is the identity matrix. S^{-m} and S^{-n} transpose and shift in the opposite direction as matrices S^m and S^n, respectively.

IV. ITERATIVE MULTI-SCALE SEMI-LOCAL INTERPOLATION

From the antialiasing standpoint, it is desirable to prefer larger patch sizes so as to find more reliable LR pixels. The larger the patch size, the more neighboring pixels are used as the reference to find ground-truth texture-relevant LR pixels and further alleviate the influence of aliasing artifacts. On the other hand, the fine textures in the natural image require smaller patch sizes to recover vivid dense details. Otherwise, the pectination

artifacts will appear caused by patch mismatching. However, it is difficult to determine the suitable patch size based on the observed LR image with aliasing. In this paper, considering these two opposite items, we propose an iterative multiscale interpolation procedure that repeat the semilocal interpolation from large to small patch size iteratively.

Algorithm 1 Iterative Multiscale Semilocal Interpolation

Input LR image I_l.

Initialize semilocal neighborhood radius r; maximum and minimum patch sizes $p_{\max} \times p_{\max}$ and $p_{\min} \times p_{\min}$, respectively; and total iteration number $T = (p_{\max} - p_{\min} + 1)$,

for iteration $t = 1$ to T **do**

    if $t == 1$, **then**

        • Initial bicubic interpolation $\tilde{I}_h^0$.

        • Set the patch size to $p_{\max} \times p_{\max}$.

    else

        • Group $\tilde{I}_h^0$ and outputs of previous iterations $\{\tilde{I}_h^1, \ldots, \tilde{I}_h^{t-1}\}$ as inputs of current iteration.

        • Set the patch size to $(p_{\max} - t + 1) \times (p_{\max} - t + 1)$.

    end if

    • Calculate the weights of semilocal LR pixels from all inputs for each missing pixel y_j (Equation (2)): $w(x_i, y_j) = \exp(-\|G \odot H \odot (u(x_i) - u(y_j))\|_2^2 / 2\sigma_d^2)$.

    • Select top L texture-relevant LR pixels that are located at different positions of the LR image.

    • Obtain HR image $\tilde{I}_h^t$ through MAP estimation (5).

end for

Output final HR image $\tilde{I}_h^T$.

In this procedure, the semilocal interpolation is repeated from large to small patch sizes iteratively. It should be noted that the outputs of all previous iterations are grouped together as inputs for the next smaller scale iteration, which is different from traditional iterative procedure. In each iteration, given the patch size and inputs that covered the outputs of all previous iterations, each interpolated semilocal masked patch $H \odot u(x_i)$ centered by LR pixel x_i, no matter which input they are from, is compared with missing pixel y_j centered masked patch $H \odot u(y_j)$. As mentioned, $H \odot u(y_j)$ contains the LR pixels only and is constant, it works as a standard to measure the texture relevance of neighboring LR pixels from all the inputs. Furthermore, the most texture-relevant LR pixels located at different positions of the LR image are selected. Meanwhile, the irrelevant ones are filtered out. Then, the MAP estimation, as mentioned, is used to recover the HR image of current iteration. As iteration increases, the number of inputs will increase. However, the patch size that measures the texture relevance will decrease. This can not only guarantee that the robustness of antialiasing from large-scale
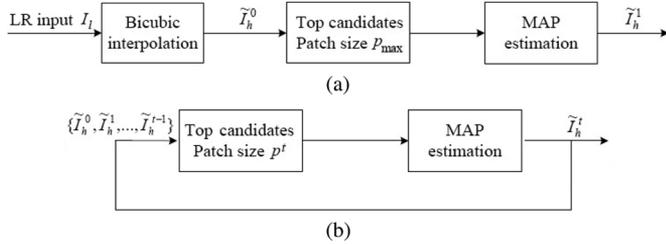
Fig. 5. Block diagram of iterative multiscale semilocal interpolation. (a) Initialization. (b) Iteration t.

semilocal interpolation is inherited but also gradually recover the dense and fine textures.

The whole process of the proposed method is described in Fig. 5 and Algorithm 1. More specifically, in the first iteration, bicubic interpolation is used as the input $\tilde{I}_h^0$. Then, the semilocal interpolation is executed with maximum patch size $p_{\max} \times p_{\max}$ to overcome the aliasing artifacts of $\tilde{I}_h^0$, and the output is represented as $\tilde{I}_h^1$. In the second iteration, $\tilde{I}_h^0$ and $\tilde{I}_h^1$ are used as inputs, and the same semilocal interpolation is executed with patch size $(p_{\max} - 1) \times (p_{\max} - 1)$. In the following iterations, the inputs include the outputs of all previous iterations and $\tilde{I}_h^0$, and the patch size gradually decreases. In final iteration T, the top L texture-relevant LR pixels measured with patch size $p_{\min} \times p_{\min}$ are selected from inputs $\{\tilde{I}_h^0, \tilde{I}_h^1, \ldots, \tilde{I}_h^t, \ldots, \tilde{I}_h^{T-1}\}$, and output $\tilde{I}_h^T$ is regraded as the final recovered HR image.

In our multiscale procedure, $p_{max} \times p_{max}$ defines the maximum patch size. As mentioned, from the antialiasing standpoint, it is desirable to prefer larger patch sizes so as to alleviate the influence of aliasing more. On the other hand, as the patch size increases, the chance to find exactly the same patches within a semilocal neighborhood will decrease. Considering these two aspects and the time complexity, we set the maximum patch size as 16×16. Additionally, it is difficult to determine the exact patch size for each missing pixel to achieve a good tradeoff between alleviating the influence of aliasing and guaranteeing the existence of similar patches within the semilocal neighborhood. Therefore, in our multiscale procedure, we perform the semilocal interpolation at each scale between the maximum and minimum patch sizes. $p_{\min} \times p_{\min}$ defines the minimum patch size in our multiscale procedure. Given patch size $p \times p$, the number of valid pixels for patch matching is between $\lfloor (p/2) \rfloor \times \lfloor (p/2) \rfloor$ and $\lceil (p/2) \rceil \times \lceil (p/2) \rceil$. To avoid the loss of patch representation ability in the minimum scale, we set the minimum patch size as 6×6. Then, the total number of iterations in our algorithm is 11.

In order to give an intuitive impress about the necessity of adopting multiscale procedure, we set up an experiment on an LR image that is downsampled with factor 2 from the image "Butterfly" of Berkeley Segmentation Database (BSD) [5]. In the experiment, we perform both our multiscale algorithm and our algorithm with only one iteration at patch size 7×7. The experimental results are shown in Fig. 6. We can see that the semilocal interpolation with only one iteration at medium patch size can only recover sharper edges than bicubic interpolation, but it does not alleviate aliasing. On the other hand, the multiscale interpolation can recover the ground truth well from the
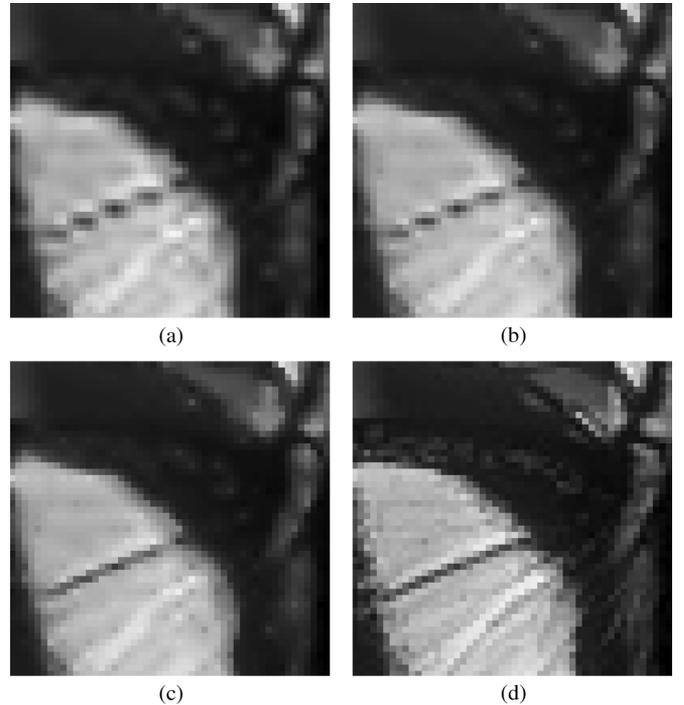

Fig. 6. Simulation results of different methods on the image "Butterfly." (a) Bicubic interpolation. (b) Our method with only one iteration at patch size 7×7. (c) Our multiscale method. (d) Ground-truth HR image.
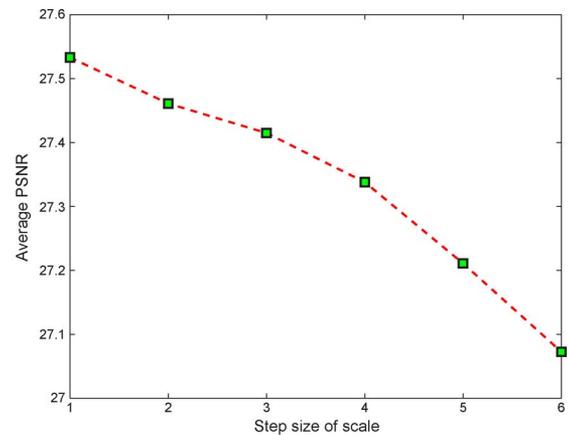

Fig. 7. Average PSNR value as the function of the step size of scale.

LR image with aliasing, such as the stripe in the image "Butterfly" [as shown in Fig. 6(c)].

In order to illustrate the necessity of computing each scale between the maximum patch size $p_{\max} \times p_{\max}$ and the minimum patch size $p_{\min} \times p_{\min}$, we set up another experiment. We first randomly select 80 images from BSD [5]. The Y channel of YUV color space of these images are separated and downsampled with factor 2 to get the LR images. Then, we perform our algorithm with different step sizes of scale between 16×16 and 6×6. For example, as the step size is 1, 5, and 6, the scales that we performed will be (16, 15, 14, ..., 7, 6), (16, 11, 6) and (16, 6), respectively. At last, we compute the average PSNR values of the results at each step size, respectively. The average PSNR curve as the function of the step size is plotted in Fig. 7. We can
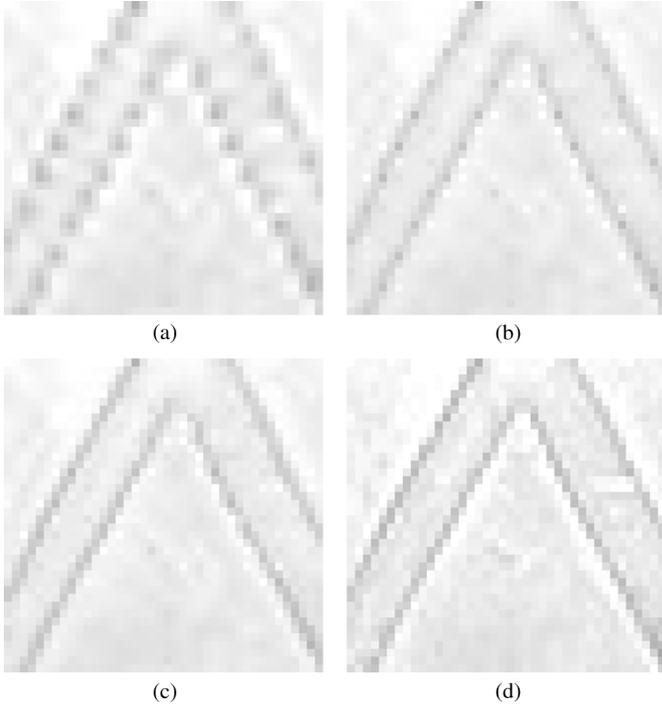
(a)   (b)

(c)   (d)

Fig. 8. Simulation results of different methods on the image "Vase." (a) Bicubic interpolation. (b) Our method with Tikhonov regularization. (c) Our method with bilateral TV regularization. (d) Ground-truth HR image.

see that the PSNR value decreases as the step size increases. Therefore, to get the best visual quality, the step size is set as 1.

In our algorithm, to reduce the time complexity, we only apply the procedure on the highly activated areas, whereas the values in smooth areas are obtained by bicubic interpolation. To guarantee the completeness of antialiasing, the pixels near the highly activated areas are also similarly handled by our algorithm. The time complexity of our algorithm is roughly about twice that of new edge-directed interpolation (NEDI) method [6]. One possible improvement, as mentioned, is that we may perform our multiscale algorithm with scale step size 2, instead of 1, at a little expense of performance, and the time complexity can be potentially reduced by half.

V. EXPERIMENTS

Here, we first conduct experiments of our method with different types of regularizer so as to justify the necessity of adopting the bilateral TV regularization. Then, we compare the proposed method with existing interpolation methods on synthetic LR images that are downsampled from corresponding HR images with factors 2 and 4 to justify the capabilities of the proposed method in relieving aliasing effects. At last, we compare the proposed method with existing interpolation methods on real-world images to validate the effectiveness of the proposed method on the LR images with a more complicated downsampling process.

In the first experiment, we apply our algorithm with bilateral TV and Tikhonov regularization [26] to the synthetic LR images, respectively. The LR image is obtained by downsampling from the image "Vase" of BSD [5] with factor 2. The experimental results are shown in Fig. 8. Fig. 8(a) shows a local re-

Fig. 9. Eight images in the test set. (From left to right in top row) Butterfly, Sail, Dragonfly, and Vase. (Bottom row) Woman, Plane, Church, and Penguin.
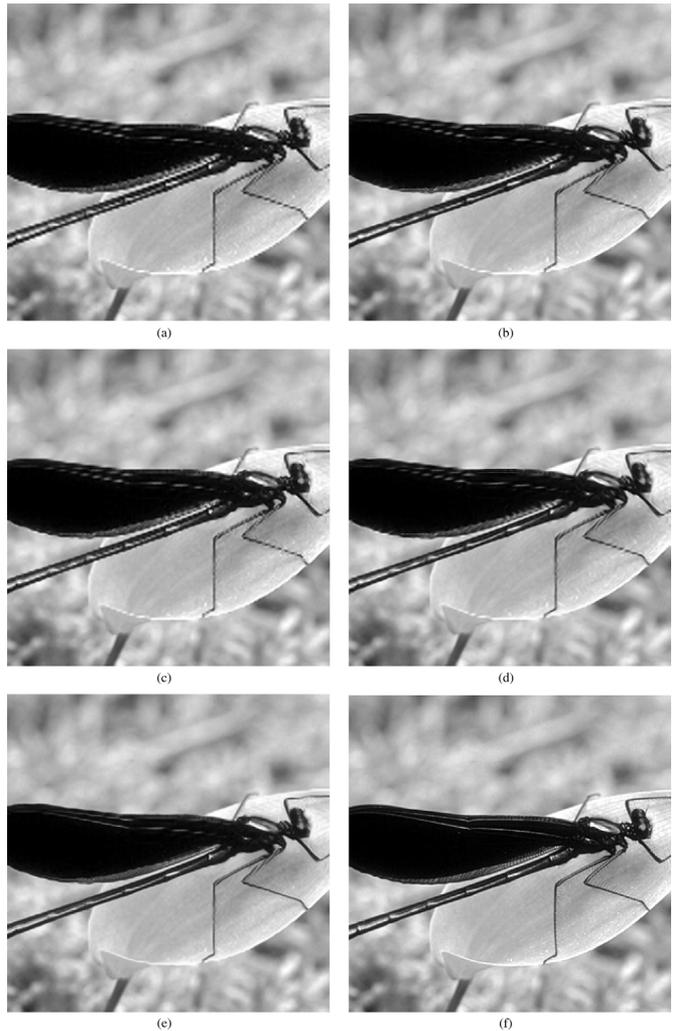

(a)   (b)

(c)   (d)

(e)   (f)

Fig. 10. Comparison of different methods on the image "Dragonfly." (a) Bicubic interpolation. (b) NEDI method in [6]. (c) Method in [7]. (d) SAI method in [8]. (e) Our method. (f) Ground-truth HR image.

gion of the LR image after bicubic interpolation, which has severe aliasing artifacts. Fig. 8(b) shows the result of our method with Tikhonov regularization. We can see that the shape of the image structure is recovered well, but the edges are blurred, and the LR pixels stand out against the interpolated pixels. This is because the Tikhonov regularization forces the spatial smoothness, and this blurring accumulates in our multiscale procedure.

TABLE I
PSNR AND SSIM COMPARISON OF DIFFERENT METHODS. FOR OUR ALGORITHM, ITS GAIN OVER THE SECOND-BEST METHOD IS ALSO GIVEN

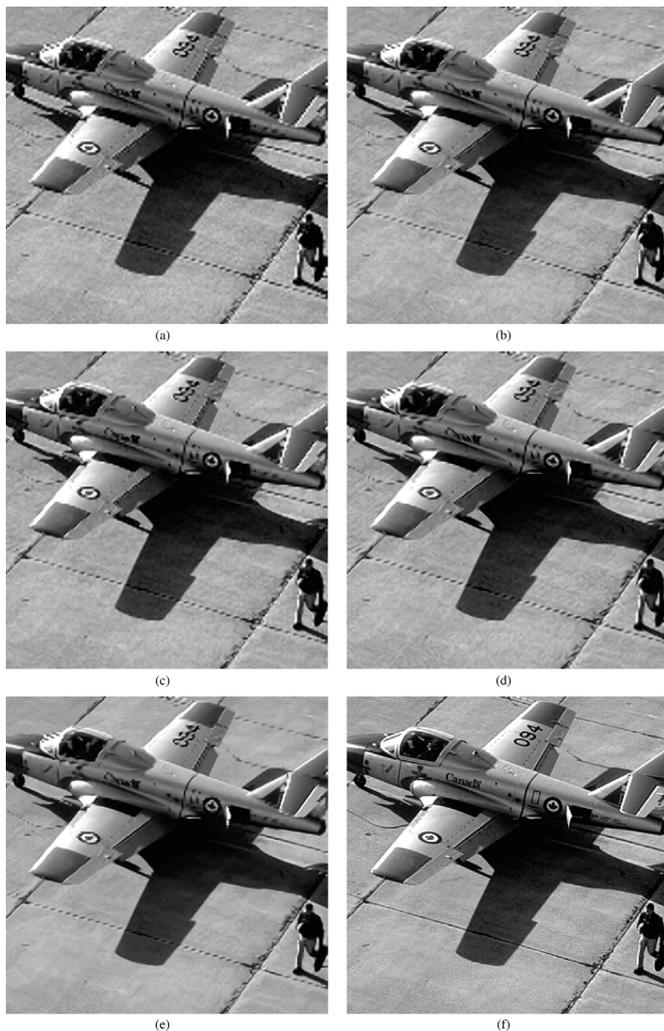| Image	Bicubic		NEDI[6]		Method[7]		SAI[8]		Our method	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Butterfly	27.01	0.896	27.50	0.900	27.23	0.897	27.83	0.909	27.96(0.13)	0.910(0.001)
Sail	25.53	0.863	25.60	0.867	25.58	0.869	25.73	0.871	26.58(0.85)	0.881(0.010)
Dragonfly	29.54	0.938	30.41	0.938	30.26	0.940	30.66	0.943	31.83(1.17)	0.954(0.011)
Vase	27.75	0.843	28.11	0.847	28.02	0.845	28.34	0.852	28.58(0.24)	0.855(0.003)
Woman	21.97	0.838	22.07	0.843	21.57	0.837	22.28	0.845	27.03(4.75)	0.887(0.042)
Plane	24.41	0.720	24.88	0.729	24.72	0.727	25.00	0.730	25.46(0.46)	0.746(0.016)
Church	22.35	0.792	22.12	0.787	22.49	0.794	22.45	0.795	22.77(0.28)	0.799(0.004)
Penguin	31.32	0.924	31.79	0.922	31.67	0.922	31.95	0.926	32.24(0.29)	0.929(0.003)
Average	26.24	0.852	26.56	0.854	26.44	0.854	26.78	0.859	27.81(1.03)	0.870(0.011)

Fig. 11. Comparison of different methods on the image "Plane." (a) Bicubic interpolation. (b) NEDI method in [6]. (c) Method in [7]. (d) SAI Method in [8]. (e) Our method. (f) Ground-truth HR image.
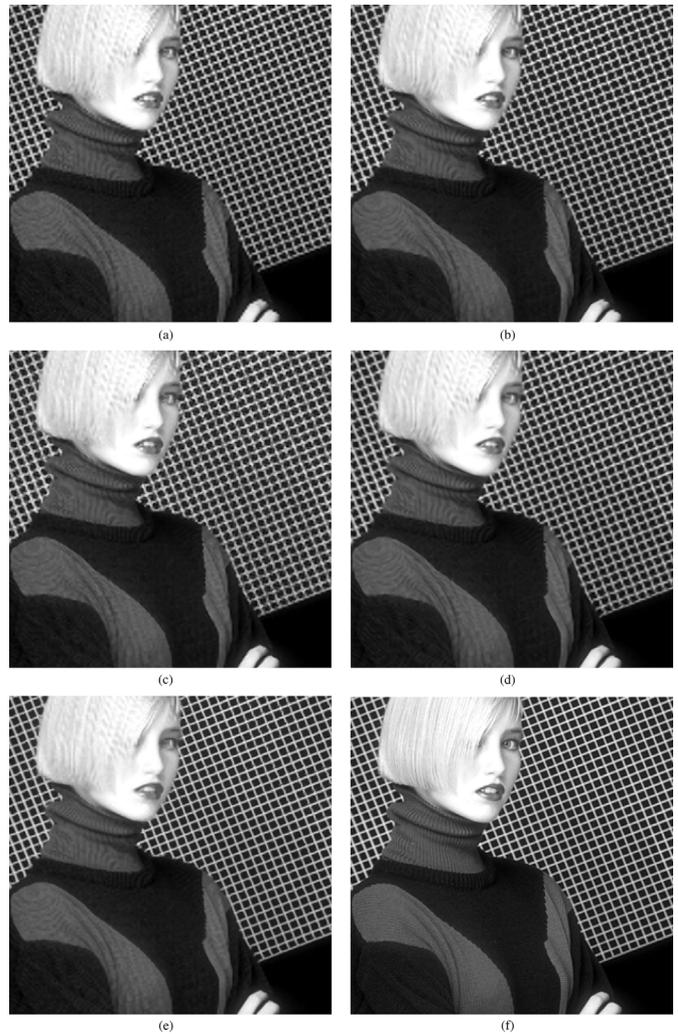

Fig. 12. Comparison of different methods on the image "Woman." (a) Bicubic interpolation. (b) NEDI method in [6]. (c) Method in [7]. (d) SAI Method in [8]. (e) Our method. (f) Ground-truth HR image.

Then, the LR pixels stand out against the blurred interpolated ones. Fig. 8(c) shows the result of our method with bilateral TV regularization. The clear structures and sharp edges testify the advantage of our method with bilateral TV regularization. Fig. 8(d) shows the ground-truth HR image.

In the second experiment, we compare the performance of the multiscale semilocal interpolation algorithm proposed in this paper with the existing image interpolation methods as follows:

1) the bicubic interpolation [4]; 2) the NEDI method [6]; 3) the weighted directional interpolation method [7]; and 4) the soft-decision adaptive interpolation (SAI) method [8]. Note that the last three methods are state-of-the-art methods to date. The code of three edge-directed methods were downloaded from author homepages, and their available parameter values were best selected based on the experimental results. The comparisons are conducted on BSD about natural scenes [5]. Fig. 9 lists eight example images in our test set. We work with Y channel of YUV

Fig. 13.   Comparison of different methods on the downsampled image "Lena," with magnification factor 4. (a) Bicubic interpolation. (b) NEDI method in [6]. (c) Method in [7]. (d) SAI Method in [8]. (e) Our method. (f) Ground-truth HR image.
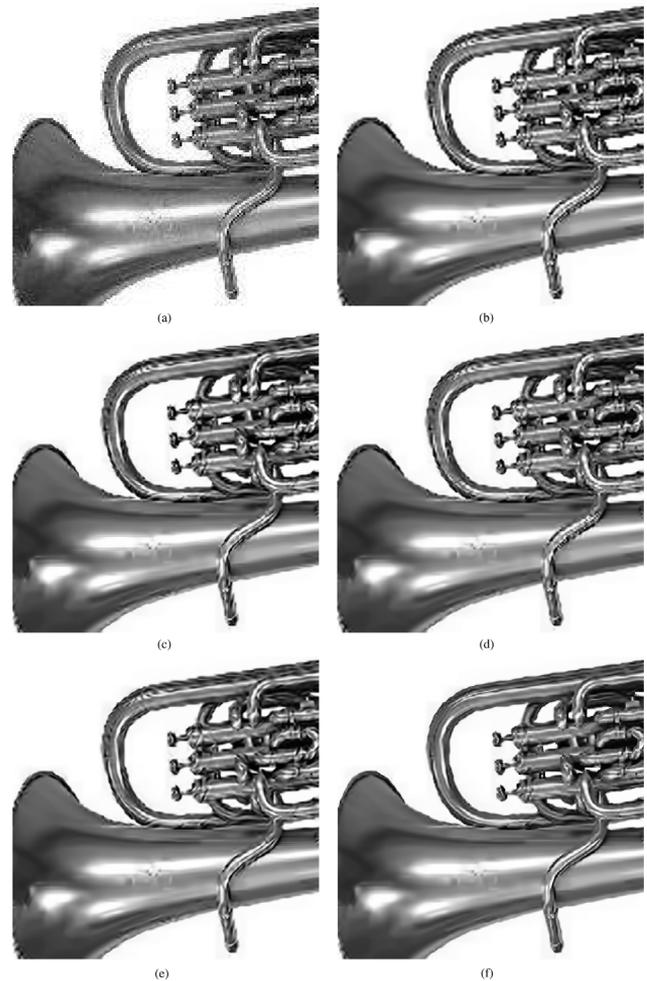

Fig. 14.   Comparison of different methods on the image "Euphonium." (a) LR image. (b) Bicubic interpolation. (c) NEDI method in [6]. (d) Method in [7]. (e) SAI Method in [8]. (f) Our method.

color space because humans are more sensitive to the brightness information. In our experiments, the observed LR images are obtained by downsampling from HR images with factor 2.

Table I lists the PSNR and the structural similarity (SSIM) index [27] values of these five different methods when applied to the eight test images in Fig. 9. In all instances, the proposed multiscale semilocal interpolation method outperforms other ones. Furthermore, for images with rich similar fragments, such as the image "Woman," the proposed algorithm exceeds the PSNR values of the second best method by 4 dB or more. The advantage of our algorithm measured with SSIM is not so significant as the one with the PSNR measurement. The main reason is that SSIM pays less attention to aliasing artifacts.

Figs. 1 and 10–12 compare the results of the five different image interpolation methods on test images "Sail," "Dragonfly," "Plane," and "Woman," respectively. From these figures, we can see that these methods exhibit different visual characteristics, particularly around the edge areas and the fine textures. Particularly, it is observed that the proposed method has obviously reduced aliasing artifacts and significantly recovered ground-truth content, such as the mast and jib in the image "Sail," the tail

and leg in the image "Dragonfly," the wing and empennage in the image "Plane," and the net and hair in the image "Woman." Furthermore, the proposed method can not only produce visually more pleasant HR images with sharp edges but also avoid lots of visual overfitting artifacts that usually exist in the other methods (e.g., the ground in the image "Plane" and the hair and sweater in the image "Woman"). This is because the large-scale semilocal interpolation is robust for antialiasing. This robustness can be inherited into the following smaller scale iterations, and these smaller scale iterations can further filter out the incorrectness caused by large-scale interpolation.

In the third experiment, we apply the multiscale semilocal interpolation algorithm proposed in this paper to a synthetic LR image that is downsampled from the image "Lena" with factor 4, to show the capabilities of the proposed method in relieving aliasing artifacts. In addition, we compare the result with the ones of four existing interpolation methods as mentioned. All of these methods are performed twice so as to get the HR image with magnification factor 4. After downsampling with factor 4, the severe aliasing artifacts appear in the LR image, and plenty of details are lost. Fig. 13 shows the results of these five different image interpolation methods. Compared with other interpolation methods, our method can recover more correct structures

and edges from the severe aliased LR images, such as the hat and the hair in Fig. 13(e).

The final experiment is conducted on a real-world LR image with aliasing artifacts to further validate the effectiveness of our method. We compare our method with four exiting interpolation methods, as mentioned on the image "Euphonium" from Caltech 101 image database [28]. The image "Euphonium" has severe aliasing artifacts, slight blurring, and noise. Therefore, we adopt the 3-D transform-domain collaborative filtering [29] to do denoising and sharpening on the LR image before performing these five interpolation methods. Fig. 14 shows the results of these five methods. Compared with other methods, our method can not only recover the correct structures and edges well from the aliased LR input but also avoid lots of overfitting artifacts, such as the edges of pipes and the trumpet. This example demonstrates that our algorithm can be readily extended on LR images generated by a more complicated downsampling process.

VI. CONCLUSION

In this paper, we have proposed an iterative multiscale semilocal interpolation method to recover high-quality HR images from texture-relevant LR pixels. This procedure can not only recover high-quality edges and structures in the image but also alleviate aliasing artifacts of LR images substantially. Experimental results validate the effectiveness of our proposed method.

There are still some drawbacks of our method. For example, under severe aliasing artifacts, the corresponding ground-truth content is still not fully recovered. The reason is that the reliable texture-relevant LR pixels are not found in the semilocal neighborhood. One approach to address this would extend the search space so as to increase the chance to find the texture-relevant LR pixels. However, it is time consuming and not practical. Hence, as a future work, it would be desirable to have a way of effectively improving the computational efficiency of our algorithm so as to comprise much larger search space for antialiasing.

ACKNOWLEDGMENT

REFERENCES

[1] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. Amer. Inst. Elect. Eng.*, vol. 47, pp. 617–644, Apr. 1928.

[2] C. E. Shannon, "Communication in the presence of noise," *Proc. Inst. Radio Eng.*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[3] P. Vandewalle, L. Sbaiz, M. Vetterli, and S. Susstrunk, "Super-resolution from highly undersampled images," in *Proc. Int. Conf. Image Process.*, Genova, Italy, Sep. 2005, pp. 889–892.

[4] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.

[5] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis.*, Vancouver, BC, Canada, Jul. 2001, pp. 416–423.

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

[7] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.

[8] X. Zhang and X. Wu, "Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 887–896, Jun. 2008.

[9] K. Jensen and D. Anastassiou, "Subpixel edge localization and the interpolation of still images," *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 285–295, Mar. 1995.

[10] S. Baker and T. Kannade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[11] J. Sun, N.-N. Zheng, H. Tao, and H. Shum, "Image hallucination with primal sketch priors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Madison, WI, Jun. 2003, pp. II-729–II-736.

[12] J. Yang, J. Wright, Y. Ma, and T. Huang, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, Jun. 2008, pp. 1–8.

[13] K. Dabov, A. Foi, V. Katkovnic, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[14] J. S. De Bonet, "Noise reduction through detection of signal redundancy," in *Rethinking Artificial Intelligence*. Cambridge, MA: MIT Press, 1997, pp. 1–2.

[15] S. Awate and R. Whitaker, "Higher-order image statistics for unsupervised, information-theoretic, adaptive image filtering," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, Jun. 2005, pp. 44–51.

[16] S. Kindermann, S. Osher, and P. W. Jones, "Deblurring and denoising of images by nonlocal functionals," *SIAM Multisc. Model Simul.*, vol. 4, no. 4, pp. 1091–1115, 2005.

[17] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, Jun. 2005, pp. 60–65.

[18] C. Kervrann and J. Boulanger, "Unsupervised patch-based image regularization and representation," in *Proc. Eur. Conf. Comput. Vis.*, Graz, Austria, May 2006, pp. 555–567.

[19] J. Boulanger, C. Kervrann, and P. Bouthemy, "Space-time adaptation for patch based image sequence restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1096–1102, Jun. 2007.

[20] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient nonlocal means for denoising of textural patterns," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1083–1092, Jul. 2008.

[21] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[22] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, May 2009.

[23] S. Ramani, D. V. D. Ville, T. Blu, and M. Unser, "Nonideal sampling and regularization theory," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1055–1070, Mar. 2008.

[24] A. Guevara and R. Mester, "Minimum variance image interpolation from noisy and aliased samples," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Austin, TX, May 2010, pp. 157–160.

[25] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 4, no. 12, pp. 2379–2394, Dec. 1987.

[26] N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient superresolution image reconstruction algorithm," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 573–583, Apr. 2001.

[27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[28] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," in *Proc. CVPR Workshop Generative Model Based Vis.*, Washington, DC, Jul. 2004, p. 178.

[29] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Joint image sharpening and denoising by 3-D transform-domain collaborative filtering," in *Proc. Int. TICSP Workshop Spectral Meth. Multirate Signal Process.*, Moscow, Russia, Sep. 2007.

Kai Guo received the B.E. degree from Shandong University, Jinan, China, in 2003, the M.E. degree from Shandong University, Jinan, China, in 2006, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2011, all in electrical engineering.

From August 2009 to March 2010, he was a Visiting Scholar at the College of Computing, Georgia Institute of Technology, Atlanta. Since 2011, he has been with at DMC Research and Development Center, Samsung Electronics, Suwon, Korea. His research interests include image and video processing, computer vision, machine learning, and pattern recognition.

Hongyuan Zha received the B.S. degree in mathematics from Fudan University, Shanghai, China, in 1984 and the Ph.D. degree in scientific computing from Stanford University, Stanford, CA, in 1993.

He was a Faculty Member of the Department of Computer Science and Engineering, Pennsylvania State University, University Park, from 1992 to 2006, and he was with Inktomi Corporation from 1999 to 2001. He is currently a Professor with the School of Computational Science and Engineering, College of Computing, Georgia Institute of Technology, Atlanta. His current research interests include Web search, recommendation systems, and machine learning applications.

Xiaokang Yang (M'00–SM'04) received the B.S. degree from Xiamen University, Xiamen, China, in 1994, the M.S. degree from Chinese Academy of Sciences, Shanghai, China, in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2000.

He is currently a Professor and the Deputy Director of the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University. From August 2007 to July 2008, he visited the Institute for Computer Science, University of Freiburg, Freiburg im Breisgau, Germany, as an Alexander von Humboldt Research Fellow. From September 2000 to March 2002, he was a Research Fellow in the Centre for Signal Processing, Nanyang Technological University, Singapore. From April 2002 to October 2004, he was a Research Scientist with the Institute for Infocomm Research (I2R), Singapore. He has published over 150 refereed papers and filed 20 patents. His current research interests include visual processing and communication, media analysis and retrieval, and pattern recognition. He actively participates in the International Standards of Motion Pictures Expert Group and JVT.

Dr. Yang He is currently a member of the Design and Implementation of Signal Processing Systems Technical Committee of the IEEE Signal Processing Society and a member of the Visual Signal Processing and Communications Technical Committee of the IEEE Circuits and Systems Society. He was the Special Session Chair of Perceptual Visual Processing of IEEE ICME2006. He was the Technical Program Cochair of the IEEE SiPS2007, and the Technical Program Cochair of 3DTV workshop in junction with 2010 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. He was the recipient of the National Science Fund for Distinguished Young Scholars in 2010, the Professorship Award of Shanghai Special Appointment (Eastern Scholar) in 2008, the Microsoft Young Professorship Award in 2006, the Best Young Investigator Paper Award at IS & T/SPIE International Conference on Video Communication and Image Processing in 2003, and awards from the Agency for Science, Technology and Research and Tan Kah Kee foundations.

Weiyao Lin received the B.E. degree from Shanghai Jiao Tong University, Shanghai, China, in 2003, the M.E. degree from Shanghai Jiao Tong University, in 2005, and the Ph.D. degree from the University of Washington, Seattle, in 2010, all in electrical engineering.

Since 2010, he has been an Assistant Professor with the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include video processing, machine learning, computer vision, and video coding and compression.

Songyu Yu received the degree from Shanghai Jiao Tong University, Shanghai, China, in 1963.

In 1963, he joined the Department of Electrical Engineering, Shanghai Jiao Tong University, where he is currently a Professor and the Vice-Director of the Institute of Image Communication and Information processing. From October 1979 to September 1981, he was a Visiting Scholar in the Department of Computer Science, University of Copenhagen, Denmark. From October 1989 to May 1990, he was a Senior Visiting Scholar in Berlin Institute of Technology, Berlin, German. He has extensively published over 80 refereed papers and two books. His research interests include image communication, digital television, and image processing.

Mr. Yu is currently the Secretary-General of Shanghai Graphics and Image Association.