Journal of Electronic Imaging 19(1), 013008 (Jan–Mar 2010)

Context-based adaptive image resolution upconversion

Guangming Shi
Weisheng Dong
Xidian University
Key Laboratory of Intelligent Perception
and
Image Understanding of the Ministry of Education
Xi'an, China 710071
wsdong@mail.xidian.edu.cn

Xiaolin Wu
McMaster University
Department of Electrical and Computer Engineering
Canada

Lei Zhang
The Hong Kong Polytechnic University
Department of Computing
Hong Kong, China

Abstract. *We propose a practical context-based adaptive image resolution upconversion algorithm. The basic idea is to use a low-resolution (LR) image patch as a context in which the missing high-resolution (HR) pixels are estimated. The context is quantized into classes and for each class an adaptive linear filter is designed using a training set. The training set incorporates the prior knowledge of the point spread function, edges, textures, smooth shades, etc. into the upconversion filter design. For low complexity, two 1-D context-based adaptive interpolators are used to generate the estimates of the missing pixels in two perpendicular directions. The two directional estimates are fused by linear minimum mean-squares weighting to obtain a more robust estimate. Upon the recovery of the missing HR pixels, an efficient spatial deconvolution is proposed to deblur the observed LR image. Also, an iterative upconversion step is performed to further improve the upconverted image. Experimental results show that the proposed context-based adaptive resolution upconverter performs better than the existing methods in both peak SNR and visual quality.* © *2010 SPIE and IS&T.*
[DOI: 10.1117/1.3327934]

1 Introduction

Image resolution upconversion, which aims to produce a clean and sharp high-resolution (HR) image from an associated degraded low-resolution (LR) image, has a wide range of applications: medical imaging, remote sensing, surveillance, computer vision, and consumer electronics. In particular, the requirements for upconverting video contents from standard definition to high definition have intensified the research on image resolution upconversion. If assuming Dirac downsampling, image resolution upconversion is commonly referred to as image interpolation. Existing image interpolation techniques fall into three categories: (1) image-independent linear interpolators, such as bilinear, cubic convolution[1] and cubic spline interpolators;[2] (2) adaptive linear interpolators, including some of the best performing interpolation techniques;[3–6] and (3) context-based interpolators[7–9] that perform the interpolation with an off-line trained filter, according to the local LR pixel structures. The image-independent linear interpolators have the lowest computational complexity, hence they are favored for real-time applications. But they suffered from blurred edges and visual artifacts, such as ringing, jaggies, and zippering. The adaptive linear interpolators have good performance, producing sharp edges and having less visual artifacts. However, the adaptive interpolators are designed using sample statistics of the LR image. Their performance is limited by the degree of agreement between HR and LR images in statistics. Also, the online design of adaptive filters is computationally expensive and unsuitable for real-time applications. The off-line-trained context-based interpolators have the advantage of incorporating prior knowledge of natural images into the design of a family of interpolation filters, one per context state. At the run time, a context-based interpolation technique chooses an interpolator according to a local LR pixel patch. This chosen interpolator adapts to the local 2-D waveform. Though the off-line training of the family of context-based interpolators is computationally expensive, the online adaptive interpolation is rather simple, making it suitable for real-time applications.

This paper is mainly concerned with context-based image resolution upconversion and the key associated problem: the design of the context-based estimators via off-line training. In addition to the already mentioned advantages,

Paper 09109RR received Jun. 16, 2009; revised manuscript received Dec. 22, 2009; accepted for publication Jan. 11, 2010; published online Feb. 24, 2010.

1017-9909/2010/19(1)/013008/9/$25.00 © 2010 SPIE and IS&T.

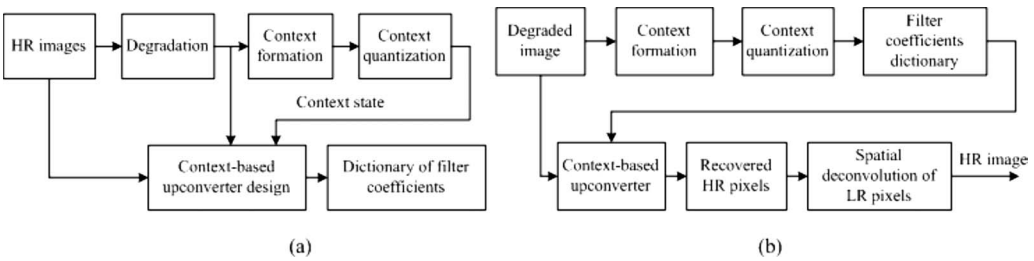

Fig. 1 Context-based image resolution upconversion: (a) training process of the context-based upconverters and (b) context-based adaptive filtering process.

the training is also a convenient way to factor in the effect of point spread function (PSF) in the design of upconversion filters. For this purpose, a set of LR images are generated by an image formation process of low-pass filtering and downsampling, which mimics the PSF of a camera. The upconversion filters are then designed by training on the pairs of LR images and corresponding HR images, learning the knowledge of the PSF. Hence, the context-based interpolators provide a joint solution for image resolution upconversion and deconvolution. Whereas almost all existing image interpolators ignore the effect of the PSF, and deconvolution must be done in a separate step after interpolation, the image deconvolution is highly susceptible to noise. The separate deconvolution tends to amplify the interpolation errors near the edges, producing severe artifacts.

The context-based image interpolator was first proposed in a patent,[7] that aimed at a practical video upconverter. In, Ref. 7, a 3×3 block of the input LR image is quantized into many binary edge patterns. According to the binary edge pattern, a set of filter coefficients is selected to perform the interpolation. Following the line of, Ref. 7 context-based filters have also been used in other applications, such as image/video artifacts reduction and deblocking.[9] However, the performance of the context-based interpolators in Refs. 7 and 9 is limited by their small 2-D supports, which is only a 3×3 template. Also, the context-based interpolators in Refs. 7 and 9 do not validate the trained filter in the input image. Thus, they perform poorly if there is a statistical mismatch between the input image and the training set. Note also that Freeman *et al.*[10] proposed to use a Markov network to learn the HR high-frequency image details from a training set. Jiji and Chaudhuri used a contorlet transform to facilitate the learning of high-frequency structures of HR images.[11] However, our work differs from the preceding papers in that we learn the interpolation filters that reconstruct HR images, rather than the high-frequency components directly.

Having identified the drawbacks of previous context-based interpolators, we set out to develop an improved context-based adaptive image resolution upconverter. We adopted a pair of context-based bidirectional interpolators instead of a single 2-D context-based interpolator. A missing pixel is estimated by two 1-D context-based upconversion filters in two mutually orthogonal directions. The two directional estimates are fused into a robust estimate. The idea of using two 1-D directional interpolators for image interpolation was studied by other authors.[12,13] But the directional interpolators used in these references are fixed (simply cubic), and they totally ignore the effect of the PSF. To provide a safeguard against possible statistical mismatches between the input image and those in the training set, we introduce a validation mechanism to assess the performance of two directional estimators in a local window and fuse the two estimates by a linear minimum mean-squares weighting approach. The other advantage of the new technique is that the image interpolation and deconvolution are performed jointly at a low complexity in a single unified framework. On the recovery of the missing HR pixels, an efficient spatial deconvolution method is used to deblur the input LR pixels for better visual quality. Also an iterative upconversion step is employed to improve the recovered HR image. Experimental results show that the proposed context-based upconverter performs better than the existing methods in both peak SNR (PSNR) and visual quality.

The rest of this paper is organized as follows. Section 2 presents the design of context-based upconverter in a training process. Section 3 presents the linear minimum mean-squares weighting of the two directional estimates generated by two 1-D context-based upconverters. Section 4 describes the spatial deconvolution and the followed iterative upconversion process. Experimental results and a comparison study are given in Sec. 5. Section 6 concludes the paper.

2 Context-Based Bidirectional Upconversion with Training

The proposed learning-based image resolution upconversion system has two aspects: off-line training and online adaptive upconversion. Figure 1(a) is a schematic description of the training process. A set of high-quality HR images are past through an image formation process of low-pass filtering and downsampling, which mimics the PSF of a camera. This generates the corresponding set of LR images, and hence builds statistical dependency between the HR and LR images. The sample statistics of the training data is used to learn a set of context-based linear filters (upconverters) that adapt to local image waveform. The context of a missing HR pixel, denoted by $\mathbf{x}\in\mathbb{Z}^{|W|}$, consists of a local window W of LR pixels, where $\mathbb{Z}$ is the set of possible pixel values. The context $\mathbf{x}$ is quantized or classified into K context states. Finally, K context-based upconverters are designed, one per context state.

Figure 1(b) depicts the work flow of the online adaptive resolution upconversion process. The context of a missing HR pixel is quantized into a context state, and the off-line-learned upconverter corresponding to the state is used to

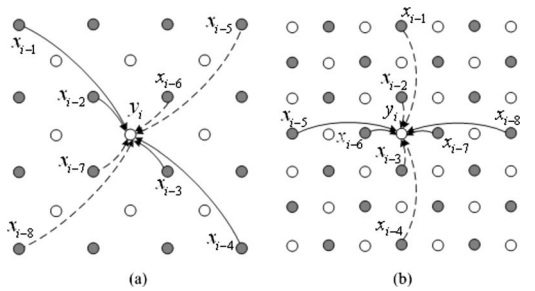

Fig. 2 Image resolution upconversion using two directional 1-D interpolators for (a) upconversion in the first pass and (b) upconversion in the second pass. Gray dots denote the known pixels of an LR image, and white dots denote the missing HR pixels.

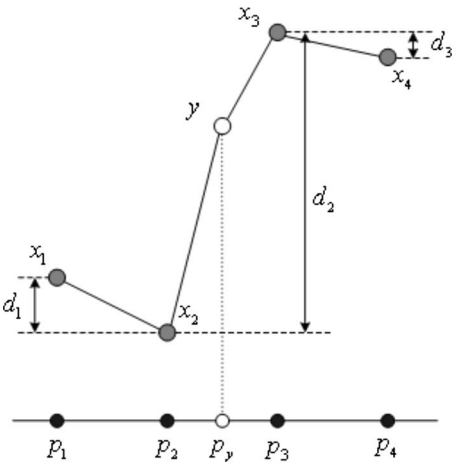

Fig. 3 One-dimensional signal classification. Black dots denote the known pixels around the missing pixel, which is denoted by the white dot. The distances between the gray dots and the black dots denote the magnitudes of the pixel values.

estimate the missing HR pixel. This estimate can be further adjusted or improved by statistics of the input LR image. This is a safeguard against any statistical mismatch between the input image and those in the training set. On the recovery of the missing HR pixels, the LR pixels themselves, which are blurred by PSF, will be restored by spatial deconvolution.

Since the image is a 2-D signal of high-order statistics, ideally a large 2-D context should be considered in resolution upconversion. But since the number of possible context states grows exponentially in context size, we face the problems of context dilution (the curse of dimensionality) and high computational complexity. To circumvent these difficulties, we use two 1-D contexts that are orthogonal to each other in the image space. For each 1-D context state, a context-based upconversion filter is designed using the statistics of the training set. Accordingly, we upconvert an LR image in two passes. In the first pass, those missing HR pixels, whose four 8-connected pixels are LR pixels, are estimated by fusing the results of two diagonal context-based upconverters, as shown in Fig. 2(a). Once the missing HR pixels are interpolated in the first pass, the remaining half of the missing HR pixels are interpolated by fusing the results of two axial (horizontal and vertical) interpolators in the second pass, as shown in Fig. 2(b).

Context formation, i.e., feature selection, is a key to the success of context-based adaptive upconverter. In our design, the context consists of four known LR pixels on a line, either diagonal or axial. As shown in Fig. 3, (p_1,p_2,p_3,p_4) is the context of the missing pixel p_y and x_1, x_2, x_3, and x_4 are the values of these four LR pixels in the context of p_y. We let $x_1 \leqslant x_4$ and reverse the sequence x_1, x_2, x_3, x_4 if $x_1 > x_4$. This serves to combine similar statistics through the symmetry of the signal waveform to reduce the number of context classes, which helps to prevent data overfitting. For the same reason, we merge the sample statistics in two diagonal directions, and design one context-based directional interpolator for diagonal cases. Similarly, one context-based directional interpolator is designed for horizontal and vertical direction cases. To characterize the waveform of signal x_1, x_2, x_3, x_4 in the interpolation direction compactly, we use three features $d_i = x_{i+1} - x_i$, $i=1,2,3$, as marked in Fig. 3. Similar features are also used in the predictive image coding technique CALIC (context-based, adaptive, lossless image coding) for context modeling of the prediction residuals.[14]

Now the task is to construct a context quantizer $Q:\Re^3 \rightarrow \{1,2,\ldots,K\}$. The quantization of feature vector $\mathbf{d}=(d_1,d_2,d_3)$ is NP-hard even for the much simpler L_2 metric in the feature space $\Re^3$ itself. To obtain a practical solution, the magnitudes of d_i, $i=1,2,3$, are quantized into L levels, respectively, with exponentially increasing size, since the magnitudes of d_i, $i=1,2,3$, obey exponential distributions. The magnitude levels of d_i, $i=1,2,3$, form $K=L^3$ contexts. This results a context quantizer $Q(\mathbf{x}):\Re^3 \rightarrow \{1,2,\ldots,K\}$. In practice, the number of L is set to be 8 for good performance.

To design context-based adaptive filters, a large training set consisting of pairs of simulated degraded images and corresponding HR images is used. The context-based interpolator $\mathbf{a}_t$ is computed as

$$\mathbf{a}_t = \arg \min_{\mathbf{a} \in \Re^4} \sum_{Q(\mathbf{d}_i)=t} (y_i - \mathbf{a} \cdot \mathbf{x}_i)^2, \quad 1 \leqslant t \leqslant K. \tag{1}$$

At the time of online context-based upconversion, a feature vector $\mathbf{d}$ can be easily mapped to $Q(\mathbf{d})=t$. Finally, the corresponding missing pixel will be interpolated by interpolator $\mathbf{a}_t$.

The design of the context-based adaptive upconversion filters is completed off line. At the run time, we first compute the context of the missing pixel and then choose the corresponding upconverter to perform the upconversion. This new filter design approach lends a high degree of adaptability to the resolution upconversion process, while having two distinct advantages over the current least-squares interpolation algorithms: (1) computationally expensive online filter design process is eliminated and (2) the preknowledge of HR images and the PSF is integrated into the context-based adaptive filters.

3 Linear Minimum Mean-Squares Fusion of Two Directional Estimates

As already pointed out, the 1D context-based directional interpolator has the advantage of low complexity. Now our task is to design a 2-D image interpolator of low complex-

ity using the proposed 1-D adaptive context-based directional interpolator. The key to the success of an image interpolator is its ability to adapt to varying pixel structures, such as edges and textures. Thus, it is crucial that we interpolate the image signal along the directions of edges and textures. However, it is difficult to estimate edge and line directions of an HR image using the LR image samples, particularly when the signal frequency is close to or beyond the Nyquist limit. The penalty of interpolating in a wrong direction is high in both subjective and objective quality. For robust estimation of edge/line directions, two 1-D context-based adaptive interpolators in perpendicular directions are used to construct a 2-D interpolator. The two 1-D interpolators can be roughly considered as adaptive basis functions for the 2-D image signal. The two directional estimates produced by the two 1-D interpolators are adaptively fused into a robust estimate. In this section, to design the 2-D image interpolator with low complexities, a linear minimum mean-squares error estimate (LMMSE) technique is used for adaptive data fusion.

As already suggested, two estimates of a missing pixel are made in two orthogonal directions, and then fused into a robust estimate. To make the data fusion process efficient enough for real-time applications, we propose an LMMSE-based data fusion approach. Referring to Fig. 2(a), for each missing pixel y_i, two directional estimates are produced by two context-based interpolators along two orthogonal diagonal directions:

$$\hat{y}_i^+ = \mathbf{a}_k^+ \cdot \mathbf{x}^+, \quad \mathbf{x}^+ = (x_{i-5}, x_{i-6}, x_{i-7}, x_{i-8}),$$

$$\hat{y}_i^- = \mathbf{a}_k^- \cdot \mathbf{x}^-, \quad \mathbf{x}^- = (x_{i-1}, x_{i-2}, x_{i-3}, x_{i-4}), \tag{2}$$

where $\hat{y}_i^+$ and $\hat{y}_i^-$ are the two directional estimates, and $\mathbf{a}_k^+$ and $\mathbf{a}_k^-$ are the two corresponding context-based interpolators. The corresponding estimation errors are denoted by e_i^+ and e_i^-, respectively, i.e., $y_i = \hat{y}_i^+ + e_i^+$ and $y_i = \hat{y}_i^- + e_i^-$. For a more accurate estimate of y_i, $\hat{y}_i^+$ and $\hat{y}_i^-$ are fused into a combined estimate by weighting:

$$\hat{y}_i = w_i^+ \hat{y}_i^+ + w_i^- \hat{y}_i^-, \tag{3}$$

where the weights $w_i^+ + w_i^- = 1$ are determined by minimizing the mean squares estimation error of y_i. We assume that e_i^+ and e_i^- are both zero mean and are approximately uncorrelated. Then the weights can be given by

$$w_i^+ = \frac{(\sigma_i^-)^2}{(\sigma_i^+)^2 + (\sigma_i^-)^2}, \quad w_i^- = \frac{(\sigma_i^+)^2}{(\sigma_i^+)^2 + (\sigma_i^-)^2}, \tag{4}$$

where $(\sigma_i^+)^2$ and $(\sigma_i^-)^2$ are the variances of e_i^+ and e_i^- in a local window centered at the location of the missing pixel y_i. In this paper, a validation mechanism is proposed to estimate $(\sigma_i^+)^2$ and $(\sigma_i^-)^2$. Also, the interpolation error variance can be learned from a training set in an off-line process. The online and off-line estimates of the interpolation error variances can be fused for a more robust estimate.

3.1 *Estimation of Interpolation Error Variance via Validation*

Assuming that image signal statistics is stationary in the locality of a missing pixel, we estimate $(\sigma_i^+)^2$ and $(\sigma_i^-)^2$ by the interpolation errors of neighboring known LR pixels. To this end, two 1-D context-based interpolators in two orthogonal directions are used to estimate the neighboring known pixels, using the estimates of the missing pixels along the same direction, and the resulting errors can be measured.

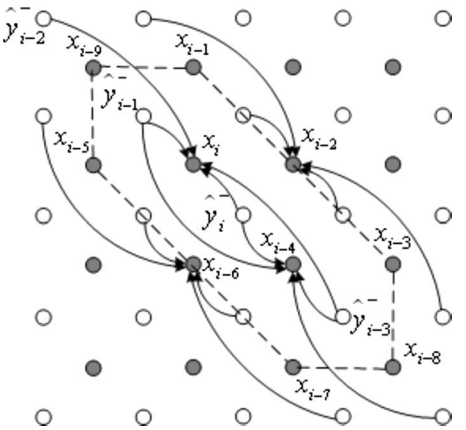

Fig. 4 Estimation of the interpolation error of the interpolator along the diagonal 135-deg direction.

As shown in Fig. 4, the known LR pixels x_{i-t}, $t=0,1,\ldots,7$, are interpolated by a 1-D context-based interpolator along diagonal 135-deg direction, using the estimates of the missing HR pixels generated along the same direction. For instance, a directional estimate of the known pixel x_i is produced along the diagonal 135-deg direction:

$$\hat{x}_i^- = \mathbf{a}_k^- \cdot \hat{\mathbf{y}}^-, \quad \hat{\mathbf{y}}^- = (\hat{y}_{i-2}^-, \hat{y}_{i-1}^-, \hat{y}_i^-, \hat{y}_{i-3}^-). \tag{5}$$

This uses the degree of fit of the interpolator $\mathbf{a}_k^-$ to estimate $(\sigma_i^-)^2$. An estimation of $(\sigma_i^-)^2$ can be obtained by averaging the mean-squares interpolation errors of the known LR pixels in a local window:

$$(\hat{\sigma}_i^-)^2 = \frac{1}{8} \sum_{0 \le t \le 7} (x_{i-t} - \hat{x}_{i-t}^-)^2. \tag{6}$$

By symmetry, one can analogously derive the estimate of $(\sigma_i^+)^2$.

3.2 *Learning Interpolation Error Variance by Training*

The validation mechanism already described requires the local stationarity of image signal. However, natural images often contain nonstationary features (e.g., edges and textures). Hence, the estimation of interpolation error variances may not be accurate. To make more robust estimates, we propose to learn the error variances from a training set. For this purpose, we condition a missing pixel in a large context that consists of a local window W of LR pixels, denoted by $C \in \mathbb{Z}^{|W|}$. The context C is quantized into J context states. Finally, J interpolation error variances are calculated, one per context state.

We now describe the learning of the interpolation error variance for each context. The context C of a missing HR pixel consists of its neighboring LR pixels. For instance, the context of y_i, $C_i = (x_i, x_{i-1}, \ldots, x_{i-9})$, is shown in Fig. 4.

Let $\{C_k, y_k\}$, $k=1,2,\ldots,N$, be a sequence of pairs of contexts and the HR pixels extracted from the training set. The collected contexts C_k, $k=1,2,\ldots,N$, are partitioned into J classes:

$$\min_{\bar{C}_1,\bar{C}_2,\cdots,\bar{C}_J} \sum_{j=1}^{J} \sum_{C_k \in S_j} \|C_k - \bar{C}_j\|, \tag{7}$$

where $\{S_1, S_2, \ldots, S_J\}$ is a J partition of the contexts, $\bar{C}_j$ is the centroid of the partition cell S_j, and $\|\cdot\|$ is the l_2 norm. We use the well-known K-means algorithm to generate the J partition. The centroids $\bar{C}_j$ of S_j are calculated by

$$\bar{C}_j = \frac{1}{|S_j|} \sum_{C_k \in S_j} C_k, \quad j = 1,2,\ldots,J. \tag{8}$$

With the classification of the contexts, the interpolation error variance of each partition cell S_j is calculated by

$$(\tilde{\sigma}_j)^2 = \frac{1}{|S_j|} \sum_{C_k \in S_j} (y_k - \mathbf{a}_k \cdot \mathbf{x}_k)^2, \quad j = 1,2,\cdots,J, \tag{9}$$

where $\mathbf{a}_k$ is the context-based interpolator corresponding to context C_k, and $|S_j|$ is the number of contexts in partition S_j.

The online estimated error variance $\hat{\sigma}^2$ is good for stationary image regions. In this case, we simply adopt the online estimates of error variances. In areas of edges and texture, $\hat{\sigma}^2$ is fused with the offline learned error variance $\tilde{\sigma}^2$, and the combined variance estimate is

$$\sigma^2 = (1-\omega)\hat{\sigma}^2 + \omega\tilde{\sigma}^2, \tag{10}$$

where ω is the weight, which can be optimized for each context. In practice, we perform the weighting only when $\hat{\sigma}^2$ is larger than a predefined threshold.

After the completion of the first pass, half of the HR pixels are obtained, which will be used to interpolate the other half of the missing HR pixels in the second pass. The remaining half of the missing pixels will be interpolated in the second pass, which is essentially same as the first pass, as shown in Fig. 2(b). Note that the 1-D context-based interpolators used in the second pass are different from those in the first pass, since the spatial distances between the neighboring pixels used for interpolation are smaller than those in the first pass. Therefore, the context-based interpolators for the two passes are designed separately.

4 Spatial Deconvolution and Iterative Image Resolution Upconversion

Upon the recovery of the missing HR pixels, the observed LR pixels themselves, which are blurred by the PSF, must be deblurred for better visual quality. We assume that the PSF is known, or can be estimated. In our simulation, the PSF is assumed to be a 2-D Gaussian filter. For the low complexity, we propose a spatial deconvolution method to deblur the LR pixels. After the observed LR pixels are deblurred, an iterative upconversion process is carried out to further improve the upconverted HR image.

Let I_l and $I_h^{(1)}$ denote the observed LR image and the upconverted HR image, respectively. Let G be a 2-D filter for the PSF with β_0 being the centered coefficient. Let G_0 be a 2-D filter such that $G_0(i,j)=G(i,j)$, $(i,j)\neq(0,0)$, and $G_0(0,0)=0$. And let the context-based upconverter be $\mathbb{F}(\cdot)$. With the constraint of the PSF, we obtain

$$D(I_h * G) = D(I_h * G_0) + \beta_0 I_\delta = I_l, \tag{11}$$

where $D(\cdot)$ denotes the 2-D downsampling operator, $*$ denotes the convolution operator, and I_δ denotes the LR image that is generated by Dirac sampling of the orignal HR image (i.e., without the PSF). From Eq. (11), we have

$$I_\delta = \frac{1}{\beta_0}[I_l - D(I_h * G_0)]. \tag{12}$$

Then, an estimate of I_δ can be obtained by substituting $I_h^{(1)}$ for I_h:

$$I_\delta^{(1)} = \frac{1}{\beta_0}[I_l - D(I_h^{(1)} * G_0)]. \tag{13}$$

Since the missing HR pixels are interpolated and deblurred jointly in a single framework, the spatial deconvolution can effectively deblur the observed LR pixels. In our experiments, the spatial deconvolution can increase the PSNR of the observed LR images by 2.0 to 6.0 dB.

Since the reconstructed HR image should ideally match the observed LR image, we have

$$I_l = D(I_h * G). \tag{14}$$

Let us consider the iterative back-projection technique[15] to minimize the norm of the reconstruction error image

$$E = I_l - D(I_h * G). \tag{15}$$

In iteration n, the error image $E^{(n)}$ is computed, and then back projected onto the nth estimated HR image $I_h^{(n)}$, producing an improved HR image:

$$I_h^{(n+1)} = I_h^{(n)} + \lambda \cdot \mathbf{F}[E^{(n)}], \tag{16}$$

where $\mathbf{F}$ is a back-projection kernel, and λ is a scalar. In our case, $\mathbf{F}$ is the context-based upconverter, and we have

$$I_h^{(2)} = I_h^{(1)} + \lambda \cdot \mathbb{F}[E^{(1)}] = I_h^{(1)} + \lambda \cdot \mathbb{F}\{I_l - D[I_h^{(1)} * G]\}. \tag{17}$$

Since $\mathbb{F}[\cdot]$ incorporates the blurring effects, it can be roughly regarded as an inverse kernel of $\mathbf{G}$. Hence, we let $\lambda=1$ and terminate the back-projection in one iteration.

We now show that the improved HR image $I_h^{(2)}$ can be achieved by combing $I_h^{(1)}$ and the upconverted $I_\delta^{(1)}$. Using the relationship between $\mathbf{G}$ and $\mathbf{G}_0$, $I_\delta^{(1)}$ can be reexpressed as

$$\begin{aligned} I_\delta^{(1)} &= \frac{1}{\beta_0}\{I_l - D[I_h^{(1)} * G_0]\} \\ &= \frac{1}{\beta_0}\{I_l - D[I_h^{(1)} * G - \beta_0 I_h^{(1)}]\} \\ &= I_l + \frac{1}{\beta_0}\{I_l - D[I_h^{(1)} * G]\}. \end{aligned} \tag{18}$$

Thus, $E^{(1)}$ can be obtained by

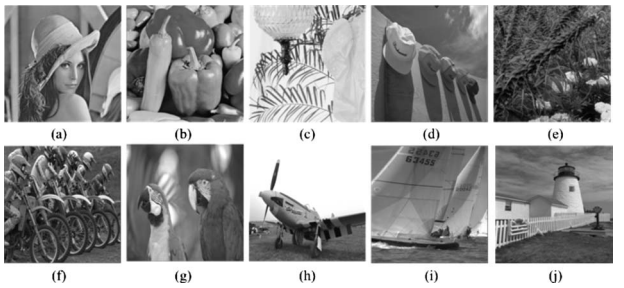

Fig. 5 Ten sample images in the test set: (a) "Lena," (b) "Peppers," (c) "Leaves," (d) "Caps," (e) "Bush," (f) "Bikes," (g) "Parrot," (h) "Plane," (i) "Sailboat," and (j) "Tower."

$$E^{(1)} = \beta_0 I_\delta^{(1)} - \beta_0 I_l. \tag{19}$$

Substituting Eq. (19) and $\lambda = 1$ into Eq. (17) yields

$$I_h^{(2)} = I_h^{(1)} + \mathbb{F}[\beta_0 I_\delta^{(1)} - \beta_0 I_l] = (1-\beta_0) I_h^{(1)} + \beta_0 \mathbb{F}[I_\delta^{(1)}]. \tag{20}$$

This reveals that $I_h^{(2)}$ can be achieved by weighting $I_h^{(1)}$ and $\mathbb{F}[I_\delta^{(1)}]$. Finally, the pixels of $I_h^{(2)}$ at LR pixel locations are replaced by those of the deblurred LR image $I_\delta^{(1)}$. In our simulations, the iterative upcoversion can substantially improve the initially upconverted HR images, increasing the upconversion quality by 0.3 to 1.3 dB.

5 Experimental Results

In this section, we present experimental results of the proposed context-based adaptive image resolution upconversion (CAIRU) method, and compare our results to those of four well-known interpolation techniques: the bicubic interpolator, the new edge-directed interpolation (NEDI) technique of Li and Orchard,[3] the edge guided interpolation[13] (EGI), and Kondo *et al.*'s context-based upconversion method in Ref. 7. All these methods other than CAIRU and that of Kondo *et al.* do not perform deconvolution. For fair comparisons, we report the results of the other methods with a deconvolution step using Wiener filtering, assuming that the PSF is known. Ten test images used for this comparison study are listed in Fig. 5. The Kondo *et al.* method and the proposed method use the same training images, which are not included in the set of test images. Fifteen high-quality natural images are used to learn the context-based interpolators and the interpolation error variances.

To have sufficient training samples for each context instance, we collected 49, 626, and 99 training samples (i.e., pairs of the context and the HR pixel) from pairs of the degraded and the corresponding high-quality images. These training samples were then classified into $8^3 = 512$ classes using the method described in Sec. 2. When creating the training data, the Gaussian PSF of standard deviation 0.6 was used.

The LR images in our simulations were generated by first applying the PSF and then downsampling by a factor of 2. The PSNR results of the reconstructed HR images by the tested methods are listed in Table 1. For these methods do not include deconvolution, we also report the PSNR results after Wiener deconvolution. The Wiener deconvolution increases the PSNR for the test images that contain large smooth regions, such as the image "Lena." But for images containing large high-frequency regions, e.g., the image "Plane," the PSNR drop can be up to 2.83 dB for the NEDI method. This is because Wiener deconvolution is highly sensitive to interpolation noises. The proposed CAIRU outperforms the other methods on all test images with no exception. The PSNR gain of the CAIRU method over the Kondo *et al.* method, which is generally the second best in the comparison group, is 0.88 to 2.26 dB. To assess the visual qualities of tested methods, the cropped HR images produced by them are presented in Figs. 6–10.

Table 1 PSNR results of the interpolated HR images by different approaches.

Images	Bicubic	Bicubic with Wiener	NEDI	NEDI with Wiener	EGI	EGI with Wiener	Kondo *et al.*	CAIRU
"Lena"	33.99	34.74	33.76	34.69	33.88	34.74	34.37	**35.60**
"Peppers"	33.60	32.84	33.68	34.30	33.71	34.30	34.00	**35.06**
"Leaves"	28.55	28.84	28.45	28.94	28.47	28.82	30.17	**32.43**
"Caps"	33.69	34.02	34.01	34.41	33.83	34.19	34.22	**35.23**
"Bush"	28.61	29.30	27.81	28.53	28.08	28.70	29.15	**30.10**
"Bike"	25.87	26.45	25.84	26.56	25.86	26.52	26.80	**27.73**
"Parrot"	34.61	35.11	34.81	35.46	34.48	35.02	35.30	**36.64**
"Plane"	31.12	28.57	31.23	28.78	31.24	28.84	31.88	**32.76**
"Sailboat"	31.62	31.73	32.18	32.33	31.76	31.91	32.34	**33.45**
"Tower"	27.51	27.71	27.11	27.43	27.40	27.67	27.83	**28.72**
Average	30.92	30.93	30.89	31.14	30.87	31.07	31.61	**32.77**

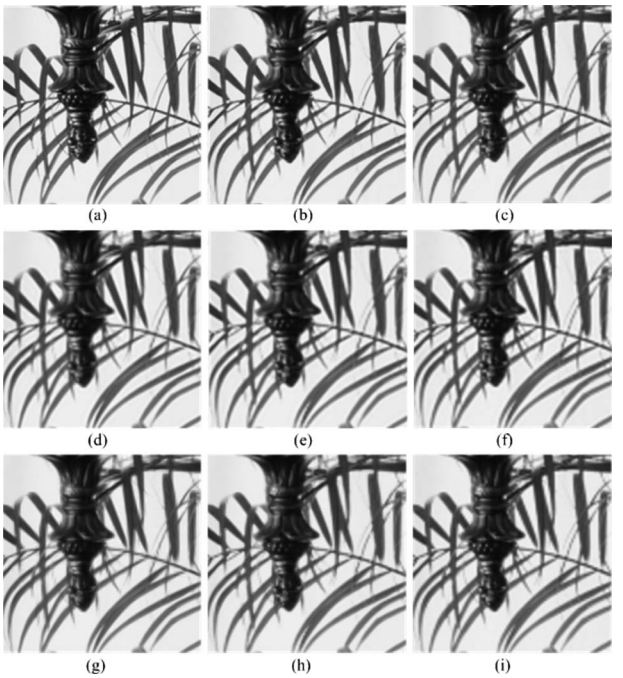

Fig. 6 Comparison of different methods on the "Leaves" image: (a) original HR image, (b) CAIRU, (c) Kondo *et al.*,[7] (d) NEDI with deconvolution, (e) EGI with deconvolution, (f) bicubic with deconvolution, (g) NEDI, (h) EGI, and (i) bicubic.

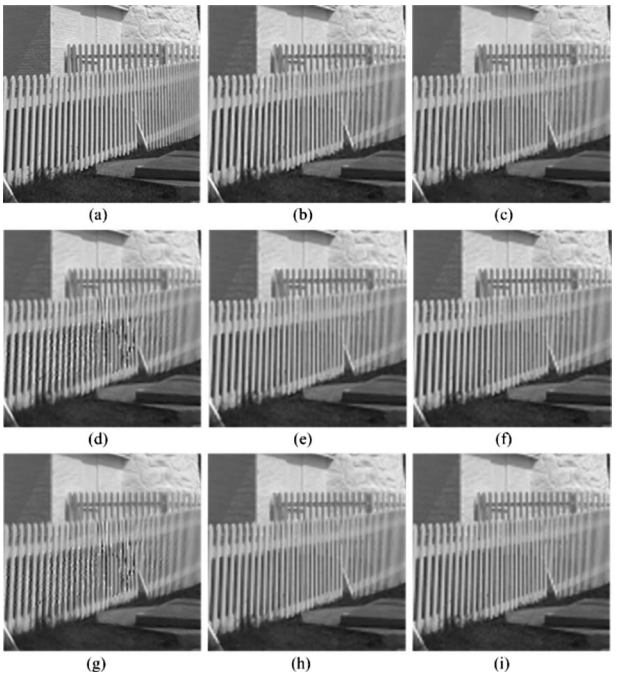

Fig. 7 Comparison of different methods on the "Tower" image: (a) original HR image, (b) CAIRU, (c) Kondo *et al.*,[7] (d) NEDI with deconvolution, (e) EGI with deconvolution, (f) bicubic with deconvolution, (g) NEDI, (h) EGI, and (i) bicubic.

The images reconstructed by the bicubic interpolator suffer from blurred edges, jaggies, and annoying ringing artifacts. The NEDI method with Wiener deconvolution can reconstruct sharp large-scale edges. But it has difficulty with small edges and textures, producing ringing artifacts and spurious small edges. The EGI method with Wiener deconvolution is slightly inferior to the NEDI method with Wiener deconvolution. The Kondo *et al.* method in Ref. 7 shows improvements over the NEDI method with Wiener deconvolution in the regions of small-scale edges and textures, eliminating the visual defects of the NEDI method. However, ringing artifacts can still be observed along edges. The proposed CAIRU technique produces the most visually pleasant results. The edges produced by the proposed upconverter are clean and sharp. Most visual artifacts that appeared in the results of the other methods, such as jaggies and ringings, are eliminated in the proposed CAIRU method.

In the preceding experiments, we assumed that the PSF used in the deconvolution step and the training of the context-based upconverters is exactly same as that used for generating the LR images. However, in practice the PSF is estimated, and we examine the robustness of the proposed method in the presence of the estimation error in PSF mismatch. We simulated a set of LR versions of the "Parrot" image using PSFs of standard deviations 0.4, 0.5, 0.6, 0.7, and 0.8. The proposed method is used to upconvert these LR images assuming a PSF of standard deviation 0.6. The restored HR images by the proposed method for the five LR versions are presented in Fig. 11. By comparing these images, we can see that the reconstructed HR images appear sharper (or smoother) when the real PSF is more peaked (or flatter) than that assumed. The perceptual quality varies most in the degree of sharpness when there is an error in the PSF, even though the PSNRs of the output images may drop by up to 2.1 dB.

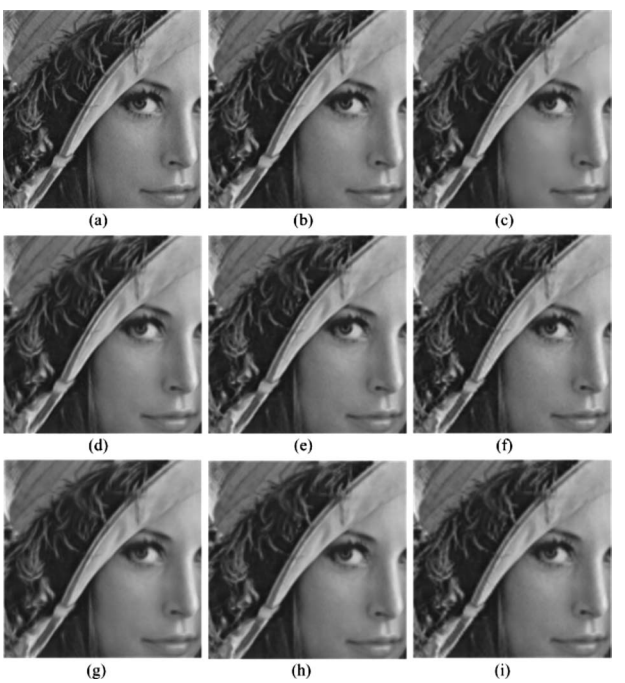

Fig. 8 Comparison of different methods on the "Lena" image: (a) original HR image, (b) CAIRU, (c) Kondo *et al.*,[7] (d) NEDI with deconvolution, (e) EGI with deconvolution, (f) bicubic with deconvolution, (g) NEDI, (h) EGI, and (i) bicubic.

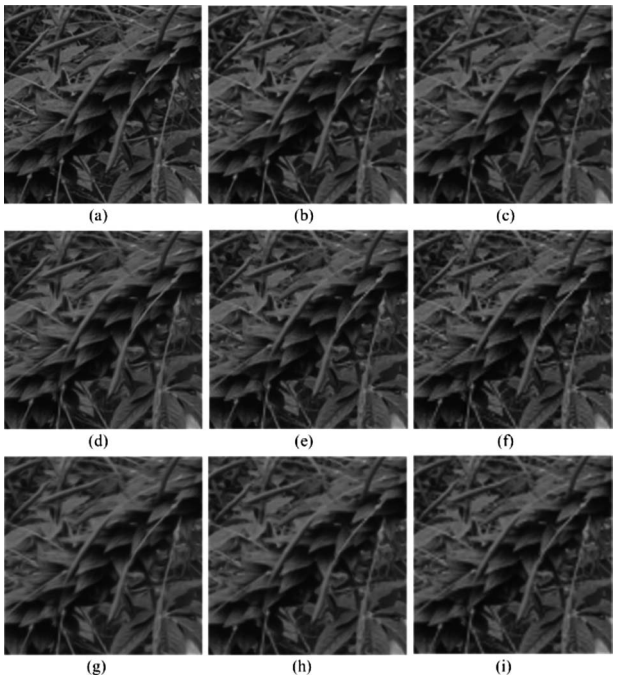

Fig. 9 Comparison of different methods on the "Bush" image: (a) original HR image, (b) CAIRU, (c) Kondo *et al.*,[7] (d) NEDI with deconvolution, (e) EGI with deconvolution, (f) bicubic with deconvolution, (g) NEDI, (h) EGI, and (i) bicubic.

Although the proposed technique is designed for a scaling factor of 2, it can be generalized to other scaling factors in principle. However, the generalization requires the scaling factor to be an additional design parameter, complicating the learning process. We leave this as future work. Instead, we adopt a simple method to extend the proposed CAIRU technique for other scaling factors. For scaling factor $S=2^n$, where n is a positive integer, we iteratively apply CAIRU n times. For other scaling factors $S \neq 2^n$, the upconversion can be carried out by first upconverting the LR image 2^n times such that $2^n < S < 2^{n+1}$, and then applying a conventional linear interpolation (e.g., bilinear or bicubic) to upconvert the output image of CAIRU with scaling factor m such that $2^n m = S$. An example of scaling factor of 3 is shown in Fig. 12, from which we can observe that the result by the CAIRU technique followed by the bicubic interpolator is visually more pleasant than that by the bicubic interpolator.

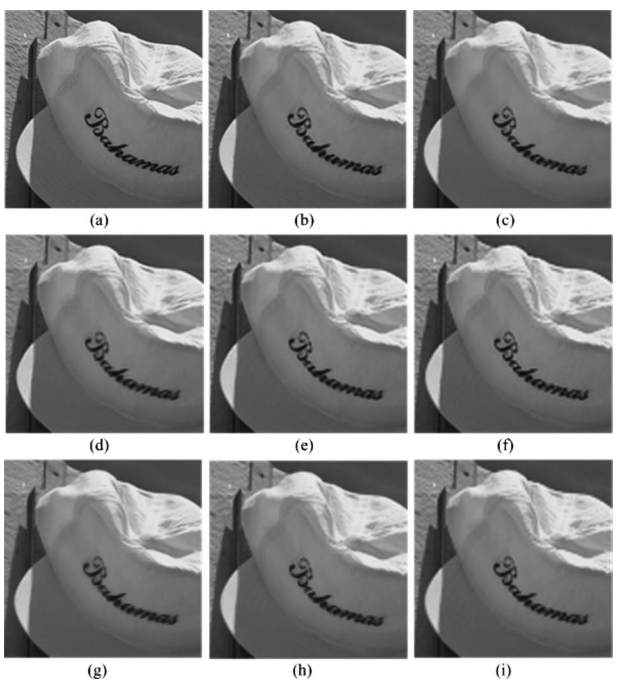

Fig. 10 Comparison of different methods on the "Caps" image: (a) original HR image, (b) CAIRU, (c) Kondo *et al.*,[7] (d) NEDI with deconvolution, (e) EGI with deconvolution, (f) bicubic with deconvolution, (g) NEDI, (h) EGI, and (i) bicubic.

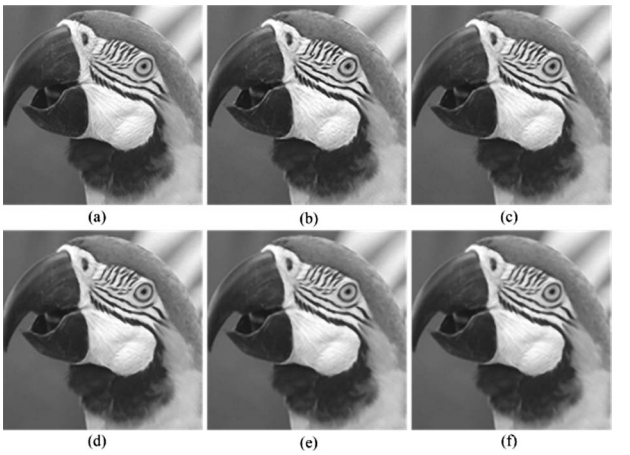

Fig. 11 Results by the CAIRU method: (a) original HR image and (b) to (f) reconstructed HR images from the LR images generated by PSFs of standard deviations 0.4, 0.5, 0.6, 0.7, and 0.8. The PSNRs of (b) to (f) are 28.88, 30.41, 31.02, 30.99, and 30.73 dB, respectively.

In addition to its good performance, the computational complexity of the proposed CAIRU technique is low. For a missing HR pixel, we are only required to interpolate it two times along two orthogonal directions using the learned upconverters, and weight these two directional estimates. The selection of the context-based upconverter can be made simply by scalar quantization. To calculate the weights in Eq. (4), each known LR pixel is interpolated two times along two orthogonal directions. The computation of each missing pixel requires only 28 multiplications (including

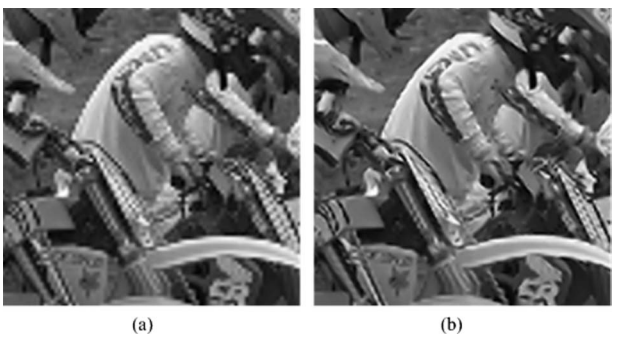

Fig. 12 Comparison of different methods on the "Bike" image with a scaling factor of 3: (a) bicubic and (b) CAIRU.

the iterative step). Hence, the computational complexity of the CAIRU method is low, making it suitable for real-time applications.

6 Conclusions

We proposed a practical context-based image resolution upconversion technique. It performs image interpolation and deconvolution jointly in a single estimation framework. To achieve low complexity, two 1-D context-based interpolators were used to estimate a missing HR pixel in two orthogonal directions. The resulting directional estimates were fused into a more robust estimate by LMMSE weighting. Upon the recovery of the missing HR pixels, a spatial deconvolution was performed to restore the observed LR pixels. The restored LR image was then used to further improve the estimated HR image in an iterative process. Experimental results demonstrated that the proposed technique can produce sharper and cleaner edges than existing techniques.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments that help to improve the presentation of this paper. This work is supported by the National Science Foundation (NSF) China (Grant Nos. 60736043, 60776795, 60805012, and 60902031), Research Fund for the Doctoral Program of Higher Education (RFDP) (No. 200807010004), and Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) (No. 20070701023).

References

1. R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-29**(6), 1153–1160 (1981).
2. H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Acoust., Speech, Signal Process.* **26**(6), 508–517 (1978).
3. X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.* **10**(10), 1521–1527 (2001).
4. X. Zhang and X. Wu, "Image interpolation by adaptive 2D autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Process.* **17**(6), 887–896 (2008).
5. T. Q. Pham, L. J. van Vliet, and K. Schuttle, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP J. Appl. Signal Process.* **2006**, 1–12 (2006).
6. N. K. Bose and N. A. Ahuja, "Superresolution and noise filtering using moving least squares," *IEEE Trans. Image Process.* **15**(8), 2239–2248 (2006).
7. T. Kondo, T. Fujiwara, Y. Okumura, and Y. Node, "Picture conversion apparatus, picture conversion method, learning apparatus and learning method," U.S. Patent No. 6,323,905, 2001.
8. L. Shao, "Adaptive resolution upconversion for compressed video using pixel classification," *EURASIP J. Adv. Signal Process.* **2007**, 71432 (2007).
9. L. Shao, H. Zhang, and G. Haan, "An overview and performance evaluation of classification-based least squares trained filters," *IEEE Trans. Image Process.* **17**(10), 1772–1782 (2008).
10. W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graphics Appl.* **22**(2), 56–65 (2002).
11. C. V. Jiji and S. Chaudhuri, "Single-frame image super-resolution through contourlet learning," *EURASIP J. Appl. Signal Process.* **2006**, 1–11 (2006).
12. D. D. Muresan, "Fast edge directed polynomial interpolation," in *Proc. IEEE Int. Conf. Image Process.*, Vol. 2, pp. 323–326 (2005).
13. L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.* **15**(8), 2226–2238 (2006).
14. X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Trans. Commun.* **45**(4), 437–444 (1997).
15. M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion and transparency," *J. Visual Commun. Image Represent* **4**, 324–335(1993).

Guangming Shi received his BS degree in automatic control in 1985, his MS degree in computer control in 1988, and his PhD degree in electronic information technology in 2002, all from Xidian University, where he joined the School of Electronic Engineering in 1988. From 1994 to 1996, as a research assistant, he cooperated with the Department of Electronic Engineering at the University of Hong Kong. Since 2003, he has been a professor in the School of Electronic Engineering at Xidian University, and in 2004 he headed the National Instruction Base of Electrician & Electronic. From June to December 2004, he studied with the Department of Electronic Engineering at the University of Illinois at Urbana-Champaign. He is currently the deputy director of the School of Electronic Engineering, Xidian University, and the academic leader in the subject of circuits and systems. His research interests include compressed sensing, the theory and design of multirate filter banks, image denoising, low-bit-rate image/video coding, and implementation of algorithms for intelligent signal processing (using digital signal processing and field-programmable gate arrays).

Weisheng Dong received his BS degree in electronic engineering from the Hua Zhong University of Science and Technology, Wu Han, China, in 2004 and he is currently pursuing his PhD degree in circuits and systems at Xidian University, Xi'an, China. From September to December 2006 he was a visiting student at Microsoft Research Asia, Bejing, China. His research interests include image compression, denoising, interpolation, and inverse problems.

Xiaolin Wu received his BSc degree from Wuhan University, China, in 1982 and his PhD degree from the University of Calgary, Ontario, Canada, in 1988, both in computer science. He began his academic career in 1988, and has since been on the faculty at the University of Western Ontario and the New York Polytechnic University, Brooklyn. He is currently with McMaster University, Hamilton, Ontario, where he is a professor in the Department of Electrical and Computer Engineering and holds the NSERC-DALSA Industrial Research Chair in Digital Cinema. His research interests include image processing, multimedia compression, joint source-channel coding, multiple description coding, and network-aware visual communication.

Lei Zhang received his BS degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, China, his MS and PhD degrees in automatic control theory and engineering from Northwestern Polytechnical University, Xi'an, China, in 1998 and 2001, respectively. From 2001 to 2002, he was a research associate with the Department of Computing, the Hong Kong Polytechnic University. From 2003 to 2006 he was a postdoctoral fellow with the Department of Electrical and Computer Engineering, McMaster University, Canada. Since 2006, he has been an assistant professor in the Department of Computing, the Hong Kong Polytechnic University. His research interests include image and video processing, biometrics, pattern recognition, multisensor data fusion, and optimal estimation theory. Dr. Zhang is an associate editor of *IEEE Trans. on Systems, Man, and Cybernetics, part C: Applications and reviews*.