10-2005

# Partition-based Interpolation for Color Filter Array Demosaicking and Super-Resolution Reconstruction

Min Shao
*Philips Medical Systems*

Kenneth E. Barner
*University of Delaware*

Russell C. Hardie
*University of Dayton*, rhardie1@udayton.edu

# Partition-based interpolation for color filter array demosaicking and super-resolution reconstruction

**Min Shao**
Philips Medical Systems
Thousand Oaks, California 91320

**Kenneth E. Barner**
University of Delaware
Department of Electrical and Computer Engineering
Newark, Delaware 19716
E-mail: barner@udel.edu

**Russell C. Hardie**
University of Dayton
Department of Electrical and Computer Engineering
Dayton, Ohio 45469

**Abstract.** A class of partition-based interpolators that addresses a variety of image interpolation applications are proposed. The proposed interpolators first partition an image into a finite set of partitions that capture local image structures. Missing high resolution pixels are then obtained through linear operations on neighboring pixels that exploit the captured image structure. By exploiting the local image structure, the proposed algorithm produces excellent performance on both edge and uniform regions. The presented results demonstrate that partition-based interpolation yields results superior to traditional and advanced algorithms in the applications of color filter array (CFA) demosaicking and super-resolution reconstruction. *© 2005 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.2087428]

Subject terms: partition based interpolators; image interpolation; color filter array.

Paper 040486R received Jul. 20, 2004; revised manuscript received Mar. 30, 2005; accepted for publication Apr. 3, 2005; published online Oct. 31, 2005.

## 1 Introduction

High resolution images are often difficult to obtain due to the limitation of physical sensors. For instance, to reduce cost, most digital still cameras (DSCs) utilize only one charge-coupled device (CCD) sensor, covered with a color filter array (CFA), to capture all three (red, green, and blue) color channels. Figure 1 shows the most commonly used Bayer pattern CFA, where each pixel samples one of the three color channels, resulting in down-sampled color channels. The process to recover missing color channels at each pixel location is called CFA demosaicking. In infrared imaging systems, detector elements in the focal plane array cannot provide a sampling rate high enough to capture high frequency components in infrared images.[1] Super-resolution reconstruction is needed to reproduce high resolution images from a series of low resolution frames.

Image interpolation algorithms are widely used in applications such as image magnification (zooming) and color filter array demosaicking, and are a critical step in super-resolution reconstruction. The commonly used bilinear,[2] bicubic,[3,4] and B-spline[5] interpolation methods employ a simple weighted sum operation to estimate high resolution (HR) pixels. Although convenient to implement, they disregard local image structures and fail to effectively preserve edges, introducing artifacts such as blurring. To address this issue, edge-directed interpolation methods have been developed to avoid interpolating across edges.[6–9] Simple implementations, such as the edge-sensing method,[6] utilize an *ad hoc* thresholding of local variance to decide whether to average along the horizontal, vertical, or diagonal directions. More sophisticated methods, such as that proposed by Li and Orchard,[9] are capable of adapting themselves to arbitrarily directed edges. This method estimates the correlation between a pixel and its four neighboring pixels by applying the classical covariance method to a local $8\times8$ low resolution (LR) window. Then, based on geometric duality, the correlation is utilized to estimate pixels on a higher resolution grid using the observed neighboring pixels. This method, however, does not perform well in cases where multiple structures exist in the same local $8\times8$ LR window.

Images, although nonstationary in nature, are generally composed of a finite set of structures, predominantly uniform regions, and edges at different orientations and scales. These structures appear repeatedly throughout an image and can thus be exploited. Indeed, the partition-based filters were designed to process signals comprised predominantly of regularly, although not periodically, occurring structures.[10] Here, we apply this partition-based methodology to image interpolation.

| | | | | |
|---|---|---|---|---|
| G | R | G | R | G |
| B | G | B | G | B |
| G | R | G | R | G |
| B | G | B | G | B |
| G | R | G | R | G |

**Fig. 1** Bayer pattern color filter array.

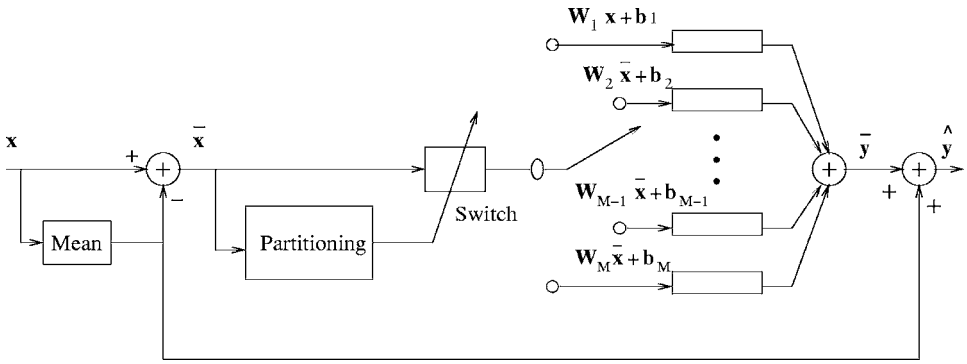



**Fig. 2** Partition-based interpolation scheme.

In this methodology, the observation space is partitioned into a finite number of regions, where each region contains the observation samples corresponding to a specific signal structure. The samples within a partition are processed by a filter designed to exploit the represented structure. This methodology has been successfully applied to the problems of ECG signal separation[11] and image signal-to-noise ratio (SNR) enhancement.[12]

The extension of the partition-based approach to interpolation utilizes vector quantization (VQ) to form structure partitions. The samples within each partition have similar statistics, which are exploited in partition-specific linear interpolators. Thus, each partition is associated with a structure-specific interpolator. This approach is similar to that utilized by Candocia and Principe[13] for image magnification. Here, we formulate the interpolation problem in the more rigorously analyzed partition-filtering framework, which provides analysis and optimization procedures, and apply the approach to the problems of color filter array demosaicking and super-resolution reconstruction.

The remainder of this work is arranged as follows. Section 2 describes the proposed partition-based interpolation algorithm and its specific formulations for the applications considered. Optimization procedures are given in Sec. 3. Section 4 compares, both quantitatively and subjectively, the performance of the proposed algorithm and others reported in the literature in the applications of CFA demosaicking and super-resolution reconstruction. Finally, conclusions are drawn in Sec. 5.

## 2 Partition-Based Interpolation Algorithm

Image interpolation aims to recover missing pixels from observed pixels. Specifically, in demosaicking, color pixels missing from a color filter ararry (CFA) sampling are estimated using pixels available in the CFA. Similarly, registration-interpolation methods are commonly used in super-resolution reconstruction (also referred as resolution enhancement). In this application, LR frames are calibrated (shifted) following registration to form a higher resolution image with nonuniformly spaced sample points.[14] Then, interpolation is applied to reconstruct all the pixels on the HR grid.

To formulate the general interpolation problem, it is commonly assumed that a natural image source can be modeled as a locally stationary Gaussian process.[9] Thus, there exists a linear operation that relates observed pixels and missing pixels in a local image structure window.

Suppose there are $p$ observed pixels and $q$ missing pixels in a local observation window. (Missing pixels on the boundary of the observation window are not considered, because some of their surrounding pixels fall outside the observation window.) If we form the observed pixels into a $p\times 1$ vector $\mathbf{x}$ and the missing pixels into a $q\times 1$ vector $\mathbf{y}$, then this linear operation can be presented as

$$\bar{\mathbf{y}} = \mathbf{W}\bar{\mathbf{x}}, \tag{1}$$

where $\mathbf{W}$ is the weight matrix of the weighted sum linear operation and $\bar{\mathbf{x}}$ (or $\bar{\mathbf{y}}$) is $\mathbf{x}$ (or $\mathbf{y}$) with zeros bias, i.e., the mean is subtracted from all elements. The mean of $\mathbf{y}$ must be added back to form the final estimation $\hat{\mathbf{y}}$. In practice, however, the mean of $\mathbf{y}$ is rarely known *a priori*, but, as noted earlier, is usually close to that of $\mathbf{x}$. In the implementation of Eq. (1), we thus form $\bar{\mathbf{y}}$ by subtracting the mean of $\mathbf{x}$ from $\mathbf{y}$. To compensate for the possible bias introduced in this approximation, we introduce an additional bias term $b$, and rewrite Eq. (1) as

$$\bar{\mathbf{y}} = \mathbf{W}\bar{\mathbf{x}} + \mathbf{b}, \tag{2}$$

where $b$ is the bias between the mean of $\mathbf{x}$ and the mean of $\mathbf{y}$ during the training process described before. Finally, this equation can be simplified as,

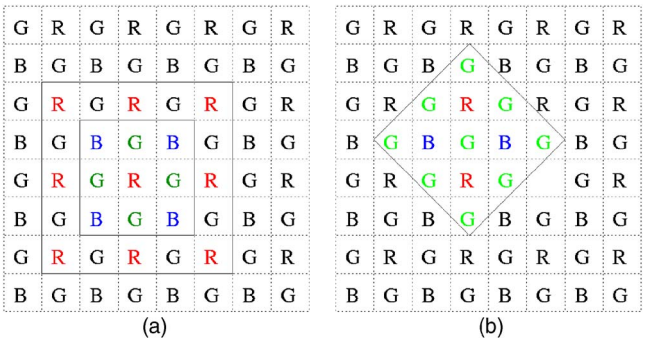



**Fig. 3** (a) A $3\times 3$ observation window for interpolating R pixels. (b) A diamond $3\times 3$ observation window for interpolating G pixels.

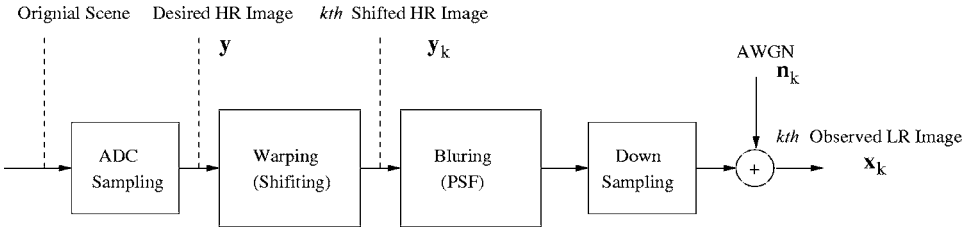



**Fig. 4** Super-resolution observation model.

$$\bar{\mathbf{y}} = \tilde{\mathbf{W}}\mathbf{v}, \tag{3}$$

where $\tilde{\mathbf{W}}$ is the augmented matrix $[\mathbf{W}|\mathbf{b}]$ and $\mathbf{v}$ is $[\bar{\mathbf{x}}^T|1]^T$. The zero-mean $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ vectors are said to be the observed and missing sample *structures*, respectively. The vectors are mean shifted to eliminate structures that differ solely due to DC offsets, e.g., intensity shifts. Note that the collection of possible observation structures forms the observation space comprising the zero mean subset of $R^p$.

As it is not realistic to enumerate all local structures in a natural image, we divide the observation space into a finite set of partitions and assign a common linear operation $\tilde{\mathbf{W}}$ to all observation structures within a unique partition. It is shown in Ref. 10 that VQ is an effective method for segmenting the observation space into representative structure-based partitions. Moreover, VQ partitioning yields improved Gaussianity for the observed structures within each partition,[10] further motivating and justifying the use of linear interpolation *within* a partition.

The proposed algorithm can, thus, be divided into two steps: 1. partitioning and 2. interpolation. In the VQ partitioning step, the observation space is divided into a set of $M$ mutually exclusive partitions, $\Omega_1, \Omega_2, \ldots, \Omega_M$, that form Voroni regions,

$$\Omega_i = \{\bar{\mathbf{x}} \in R^p : \|\bar{\mathbf{x}} - \mathbf{C}_i\|^2 \leqslant \|\bar{\mathbf{x}} - \mathbf{C}_j\|^2\} \text{ for } j = 1, \ldots, M,\ j \neq i, \tag{4}$$

where each partition is represented by a vector $\mathbf{C}_i \in R^p$, the set of which forms the codebook $\mathbf{C} = \{\mathbf{C}_i, i = 1, \ldots, M\}$. The codewords that form the codebook can be set as the centroids of the samples within each partition,

$$\mathbf{C}_i = \frac{1}{\|\Omega_i\|} \sum_{\bar{\mathbf{x}} \in \Omega_i} \bar{\mathbf{x}}, \tag{5}$$

where $\|\Omega_i\|$ denotes the number of observation vectors in the $i$'th partition.

In the interpolation step, a linear operation $\tilde{\mathbf{W}}_i$ is applied to all $\bar{\mathbf{x}} \in \Omega_i$,

$$\hat{\bar{\mathbf{y}}} = \tilde{\mathbf{W}}_i \mathbf{v}, \tag{6}$$

where $\hat{\bar{\mathbf{y}}}$ is the estimate of $\bar{\mathbf{y}}$. This relation can be alternatively written as

$$\hat{\bar{\mathbf{y}}} = \sum_{i=1}^{M} \tilde{\mathbf{W}}_i \mathbf{v} I(\bar{\mathbf{x}} \in \Omega_i), \tag{7}$$

where

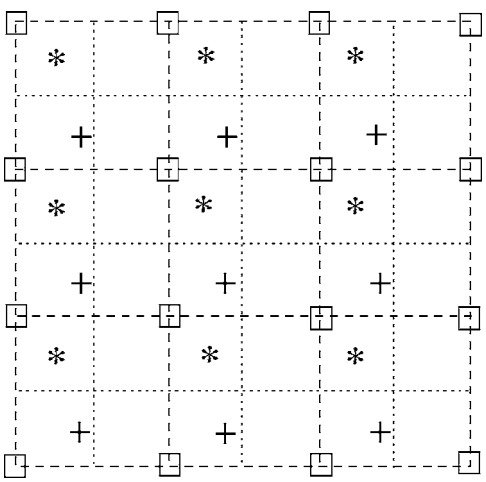


**Fig. 5** Higher resolution image with nonuniformly spaced samples. [(□) pixels from the reference frame; (*) pixels from the second frame; (+) pixels from the third frame; (– – –) LR grids of the reference frame; (…) HR grids of the reference frame.]

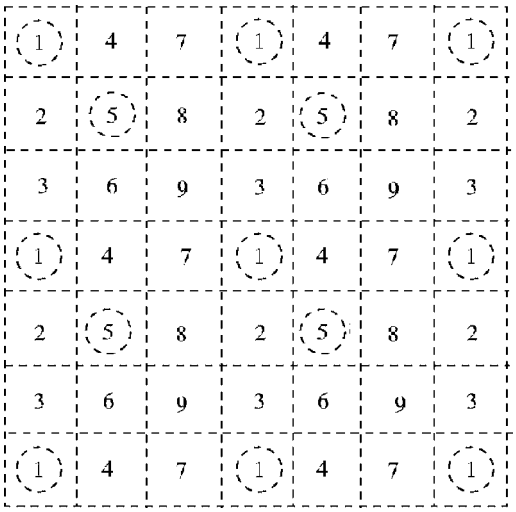



**Fig. 6** An observation window for super-resolution reconstruction. [1: pixels from the reference LR frame; 2 to 9: pixels from other LR frames. ○: pixels from available (observed) LR frames.]

(a) (b)

**Fig. 7** Original nonmosaicked images: (a) Building (400×532), (b) Bike (526×758).

$$I(\text{event}) = \begin{cases} 0 & \text{if event is false} \\ 1 & \text{if event is true} \end{cases}. \quad (8)$$

Finally, since the mean of $\mathbf{y}$ is rarely known *a priori*, but is usually close to that of $\mathbf{x}$, we estimate $\hat{\mathbf{y}}$ as $\hat{\bar{\mathbf{y}}}$ with the mean of $\mathbf{x}$ added back to its elements.

The block diagram of this partition-based interpolation scheme is shown in Fig. 2. First, the LR observation vector $\mathbf{x}$ is normalized to obtain the observation structure $\bar{\mathbf{x}}$. Next, $\bar{\mathbf{x}}$ is assigned to one of the linear operations, based on the results of partitioning. Finally, the mean of $\bar{\mathbf{x}}$ is added back to form the estimated $\hat{\mathbf{y}}$.

Note that $\hat{\mathbf{y}}$ contains multiple elements, and missing pixels may thus be estimated from multiple observation windows. In such cases, we simply average the multiple estimates to form the final result, although more sophisticated combinations could certainly be employed. As noted in Ref. 13, if all of the multiple estimates are reliable, this averaging does not produce the low pass filtering effect commonly associated with averaging.

This general partition-based interpolation formulation can be customized to address the specific needs of particular applications. In the following, we present implementations specific to the applications of image color filter array demosaicking and super-resolution reconstruction.

### 2.1 *Demosaicking*

Color filter array demosaicking is fundamentally an interpolation problem that can be addressed through partition-based means. In this case, simple variations on the window shape are introduced. Although we only discuss demosaicking of the most popular Bayer pattern,[15] the approach easily lends itself to other CFA patterns.

A Bayer pattern CFA is shown in Fig. 1, where the LR locations of the captured red, blue, and green pixels are noted with reference to the desired HR samples. Note that the captured R and B LR pixels represent a $2\times2$ down sampling. The G pixels, however, are considered more important, and are thus captured at a higher density on the LR grid. As mentioned in Sec. 2, the demonsaicking process aims at recovering the missing color pixels, which in this special case are the missing R and B pixels at G pixel locations, and missing G and B pixels at R pixel locations. Define $L_1$ and $L_2$ as the up sampling rate along the vertical and horizontal directions, respectively. The observation vector $\mathbf{x}$ is formed by the pixels in an $H_1\times H_2$ LR observation window and the elements of $\mathbf{y}$ are the missing pixels, on the HR grid, within this observation window.

The HR R and B pixels are obtained by a straightforward $L_1=L_2=2$ interpolation, illustrated in Fig. 3(a) for the

**Table 1** PSNR of R pixel demosaicking using bilinear interpolation in the original color and color difference space, and edge-directed and partition-based interpolation in the color difference space.

| | Image | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Method (color space) | Bridge | Lake | Building | Trees | Parrot | Yard | Bike | Woman |
| Bilinear (original) | 23.72 | 25.33 | 16.70 | 27.95 | 30.72 | 21.51 | 18.84 | 19.99 |
| Bilinear (difference) | 29.54 | 30.05 | 22.31 | 21.83 | 30.76 | 26.02 | 25.21 | 25.71 |
| Edge-directed (difference) | 30.31 | 30.27 | 23.12 | 22.06 | 31.08 | 26.73 | 26.03 | 24.20 |
| Partition-based (difference) | 31.52 | 32.15 | 24.81 | 22.97 | 32.59 | 28.26 | 29.43 | 28.28 |

**Table 2** PSNR of G pixel demosaicking using bilinear interpolation in the original color and color difference space, and edge-directed and partition-based interpolation in the color difference space.

| | Image | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Method (color space) | Bridge | Lake | Building | Trees | Parrot | Yard | Bike | Woman |
| Bilinear (original) | 27.62 | 28.33 | 19.98 | 20.40 | 34.20 | 24.82 | 22.15 | 22.60 |
| Bilinear (difference) | 32.71 | 32.81 | 24.31 | 24.05 | 34.93 | 29.34 | 27.50 | 27.40 |
| Edge-directed (difference) | 34.04 | 32.76 | 24.46 | 23.89 | 35.66 | 29.03 | 27.47 | 26.48 |
| Partition-based (difference) | 34.72 | 34.02 | 25.80 | 24.77 | 36.02 | 30.56 | 30.08 | 30.83 |

R case with $H_1=H_2=3$. Specifically, the nine colored R pixels form **x**, and the eight missing R pixels at the location of colored G and B pixels form **y**.

The G pixels, because of their higher sampling rate, require $L_1=L_2=\sqrt{2}$ interpolation. To achieve this interpolation, we utilize a diamond-shaped window, depicted in Fig. 3(b), for $H_1=H_2=3$. In this case, the nine colored LR G pixels inside the window constitute **x**, and the four missing G pixels form **y**.

### 2.2 *Super-Resolution Reconstruction*

In super-resolution reconstruction, multiple same scene LR image frames are used to reconstruct HR images. An observation model relating the original HR and the observed LR images[14] is shown in Fig. 4. This model is equivalent to the true imaging process where the observed images are formed before ADC sampling, provided that the HR images are sampled at or above the Nyquist rate.[16]

During an imaging process, the original (continuous) scene may be subject to a series of operations, including ADC sampling, warping, blurring, down sampling, and noise contamination.[14] Warping includes rotation and other translations, such as shifting. Blurring processes may include optical (e.g., out of focus), motion, and sensor point spread function (PSF) blur.[14] This is a very general model that shows that the desired HR image is related to multiple LR observed images in a nontrivial fashion. We address a specific realization of this model consisting of global shift warping, a spatial averaging PSF, and additive Gaussian noise.

The HR image in this application can be obtained through a registration-interpolation approach, which typically involves three steps: registration, interpolation, and restoration.[1,14,17] The registration process extracts the relative motion between the LR frames and the reference LR frame. (The reference frame is the LR image for which the motion shift is defined to be zero.) Using the estimated motion information, LR images can be calibrated (shifted) and super-imposed to form a higher resolution image with nonuniformly spaced samples,[14] as illustrated in Fig. 5. Next, an interpolation algorithm is utilized to reconstruct missing samples on the HR grids.[1,18–20] And finally, a restoration method, such as the Wiener filter, is employed to reconstruct the HR image, cancelling blurring and additive noise.

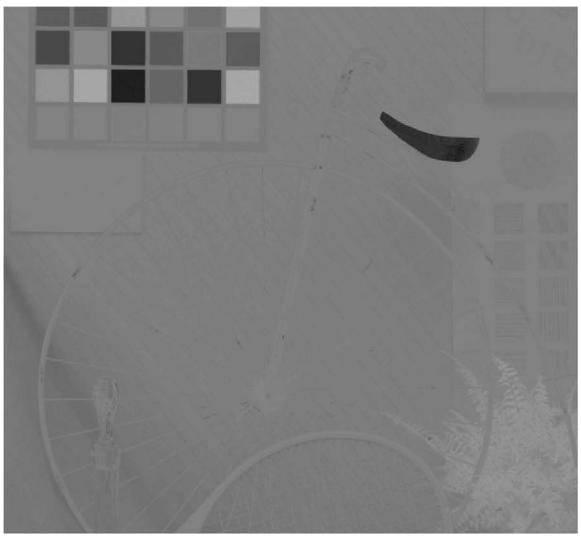
Partition-based interpolation can be readily incorporated into the general registration-interpolation framework. In our implementation, however, we combine the interpolation and restoration operations into a single step. The proposed method utilizes gradient-based registration.[21]

Once all the LR images are registered, they are superimposed to form a higher resolution image, as in Fig. 5. Since motion shift may be fractional, we first round LR frames to their nearest HR grid locations. If multiple frames are rounded to the same HR grid, we take a weighted sum of these frames. The weight applied to each LR frame in the weighted sum operation is inversely proportional to the distance between the LR frame and the HR grid it is rounded to. (Note that the weights are normalized such that their summation is one.) Once all the LR pixles are set to

**Table 3** PSNR of B pixel demosaicking using bilinear interpolation in the original color and color difference space, edge-directed and partition-based interpolation in the color difference space.

| | Image | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Method (color space) | Bridge | Lake | Building | Trees | Parrot | Yard | Bike | Woman |
| Bilinear (original) | 24.50 | 25.29 | 16.94 | 18.60 | 32.05 | 21.61 | 18.56 | 20.12 |
| Bilinear (difference) | 29.45 | 30.06 | 21.26 | 22.22 | 32.29 | 26.02 | 25.13 | 25.59 |
| Edge-directed (difference) | 30.23 | 30.05 | 21.82 | 22.54 | 32.80 | 26.38 | 25.76 | 24.88 |
| Partition-based (difference) | 31.53 | 32.05 | 22.74 | 23.33 | 33.85 | 27.38 | 28.34 | 28.68 |

(a)



(b)

**Fig. 8** (a) G channel image and (b) $K_R$ image of Bike.

their appropriate HR grid locations, they are considered as one effective frame in the following operations. If a sufficient number of frames is available to fill all the HR grids, interpolation is not necessary. This situation, however, rarely occurs in practice.

Note that there are totally $(2^{L_1L_2}-1)$ possible configurations of available LR frames. Usually, training of partition-based interpolation needs to be carried out after the registration step has determined the specific configuration. Due to the complexity of partition-based interpolation, the proposed super-resolution reconstruction method is not well suited to real-time applications, such as video processing. For small $L_1$ and $L_2$ cases, however, it is possible to have the codebooks and weights for all the configurations precomputed, allowing a real-time implementation of partition-based interpolation.

As an illustrative example, consider the case $L_1=L_2=3$. In this example, there are nine possible LR grid locations, as shown in Fig. 6. The nine LR images have been rounded to their closest HR grids using the previously mentioned procedure. In this example, only frame 1 (the reference frame) and frame 5 are available. The frames numbered 2 to 4 and 6 to 9 represents empty samples on the HR grid.

In this application, we base the observation window on the HR image. All available LR samples in the window form the observation vector $\mathbf{x}$. Note that since we combine interpolation *and* restoration into a single step, all the HR samples constitute $\mathbf{y}$. Thus, $\mathbf{y}$ contains the restored (deblurred and denoised) versions of samples at the observed LR locations, as well as interpolated and restored samples at the missing data locations. The samples shown in Fig. 6 can now be interpretated as a $7\times7$ observation window, the observed samples within which comprise $\mathbf{x}$. In this case, $\mathbf{y}$ contains restored versions of the 1 and 5 frame samples as well as interpolated and restored versions of 2 to 4 and 6 to 9 frame samples. Thus, the weight vector $\tilde{\mathbf{W}}$ is optimized to jointly perform interpolation and restoration. Last, note that in this example $p=13$ and $q=25$.

## 3 Optimization of Partition-Based Interpolation

The partitioning and interpolation stages are nonlinearly coupled. This nonlinear coupling makes global optimization difficult. While methods have been developed to globally optimize partition-based filters,[10] they are extremely computationally intensive and slow to converge. Moreover, a simple suboptimal procedure that performs a two-stage optimization, first optimizing the partitioning operation followed by filtering optimization, has been shown to produce near optimal results.[10] Thus, we adopt this simpler two-stage approach. Specifically, the VQ codebook is generated using the Linde-Buzo-Grey (LBG) algorithm.[22] The VQ partitioning obtained using the LBG algorithm minimizes partitioning errors and effectively captures the underlying spatial structures in observation signals and images.[12,23]

Given the defined partitioning scheme, optimization of interpolation coefficients is carried out minimizing the mean squared interpolation error

$$J=E\left\|\bar{\mathbf{y}}-\sum_{i=1}^{M}\tilde{\mathbf{W}}_i\mathbf{v}I(\bar{\mathbf{x}}\in\Omega_i)\right\|^2. \tag{9}$$

The optimial weight matrices are obtained by setting the gradient of $J$, with respect to each weight matrix $\tilde{\mathbf{W}}_i$, to zero, yielding

$$\tilde{\mathbf{W}}_i=E[\bar{\mathbf{y}}\mathbf{v}^T](E[\mathbf{v}\mathbf{v}^T])^{-1}. \tag{10}$$

Intuitively, optimizing a weight matrix for each partition reduces to obtaining the Wiener solution[24] for observation vectors within that partition.

The determination of the weight vector in Eq. (10) requires the statistics of the test image, which can be obtained from training data if it is not known *a priori*. Training data can be obtained in two ways: 1. selecting a representative image that shares similar statistics with the test image, or 2. utilizing the observed test image itself. We refer to these methods as training schemes 1 and 2, respectively. Candocia and Principe[13] applied scheme 1 to the image zooming applications, where $\tilde{\mathbf{W}}$ needs only to represent correlations


**Fig. 9** Results of demosaicking: (a) original nonmosaicked image and results of (b) bilinear method operating in the original color space, (c) bilinear, (d) edge-directed, and (e) partition-based method operating in the color difference space.

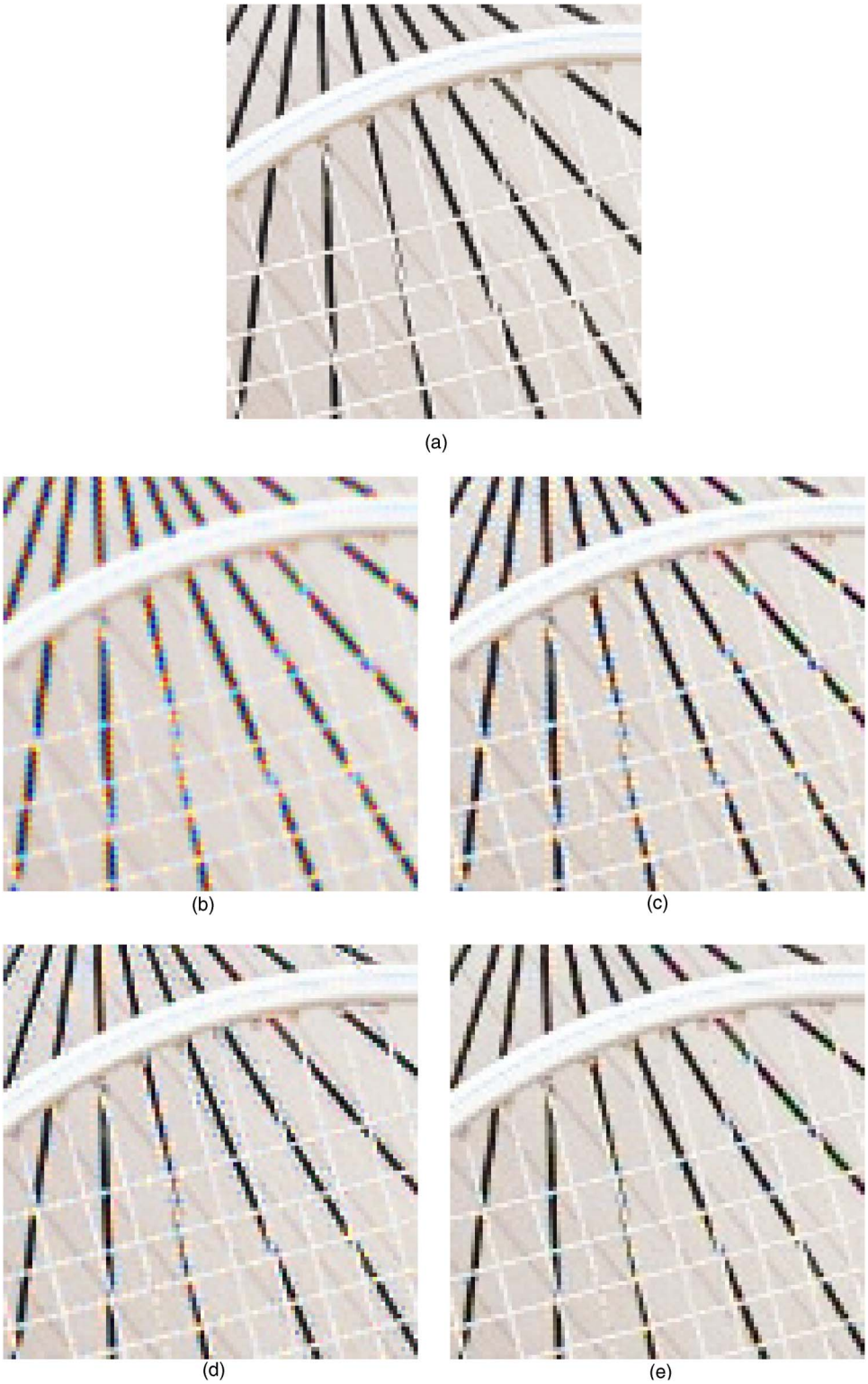



**Fig. 10** Results of demosaicking: (a) original nonmosaicked image and results of (b) bilinear method operating in the original color space, (c) bilinear, (d) edge-directed, and (e) partition-based method operating in the color difference space.

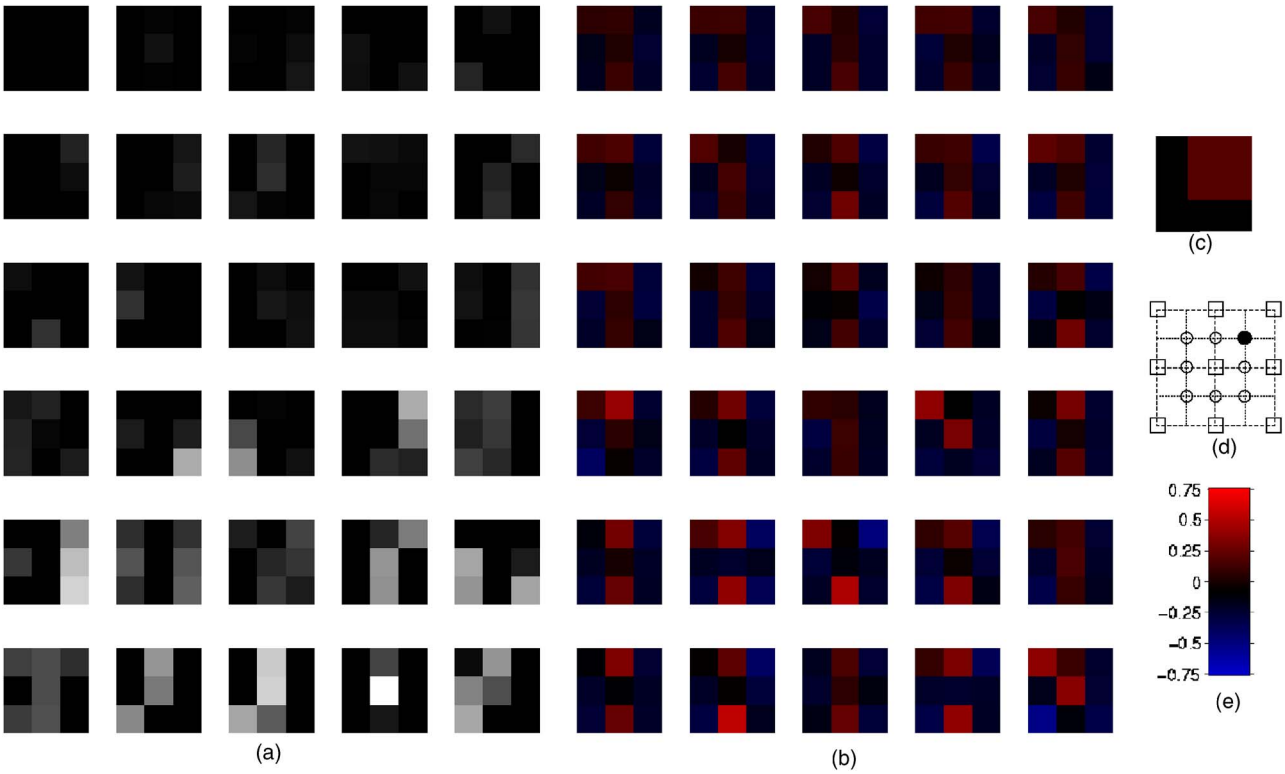



**Fig. 11** The $M=30$ (a) codebook and (b) weight matrix [for the reconstruction of the black pixel in (d)]. For comparison, the bilinear weights are shown in (c). The color map used to plot (b) and (c) is given in (e).

across scales. This approach does not exploit same-scale correlations, as is required, for instance, in the super-resolution application. Also note that in scheme 1, interpolation performance is dependent on the statistics of the representative image. This motivates the use of training scheme 2, which utilizes geometric duality[9] by assuming the statistics remain consistent across scales. This is the approach adopted here to address demosaicking applications. In super-resolution reconstruction, however, we choose scheme 1 because the observed images are not large enough to generate sufficient training data in this application.

Note that in the super-resolution reconstruction case, the goal is to estimate the original scene. Thus, the weight matrix is trained not only to interpolate, but to also reverse blurring and minimize noise contamination. Once the optimization is complete and the method is applied to an observation image, the interpolation and restoration steps are carried out in single matrix multiplication. This results in a much simpler architecture and hardware implementation.

**Table 4** Combination of LR frames utilized in the simulation.

| $r$ | Frames selected |
|---|---|
| 1 | 1 |
| 2 | 1,5 |
| 3 | 1,5,9 |
| 4 | 1,3,5,9 |
| 5 | 1,3,5,7,9 |
| 6 | 1,2,3,5,7,9 |
| 7 | 1,2,3,5,7,8,9 |
| 8 | 1,2,3,4,5,7,8,9 |
| 9 | 1,2,3,4,5,6,7,8,9 |

## 4 Simulation and Results

In this section, we demonstrate the effectiveness of the proposed algorithm in the applications of demosaicking and super-resolution reconstruction. The partition-based methods are compared with traditional and advanced algorithms used in practice and reported in the literature.

### 4.1 *Demosaicking*

Consider first the application of demosaicking, where we compare bilinear, edge-directed[9] and partition-based interpolation operating on images Bike and Woman, which are JPEG2000 test images, and images Bridge, Lake, Building, Trees, Parrot, and Yard, which are Kodak photo samples. The original nonmosaicked Building and Bike images are shown in Fig. 7. All seven images are sampled using a Bayer pattern, Fig. 1, to generate the test images.

As shown in Fig. 8, color difference images have much less details than images in the original color space.[25] Thus,

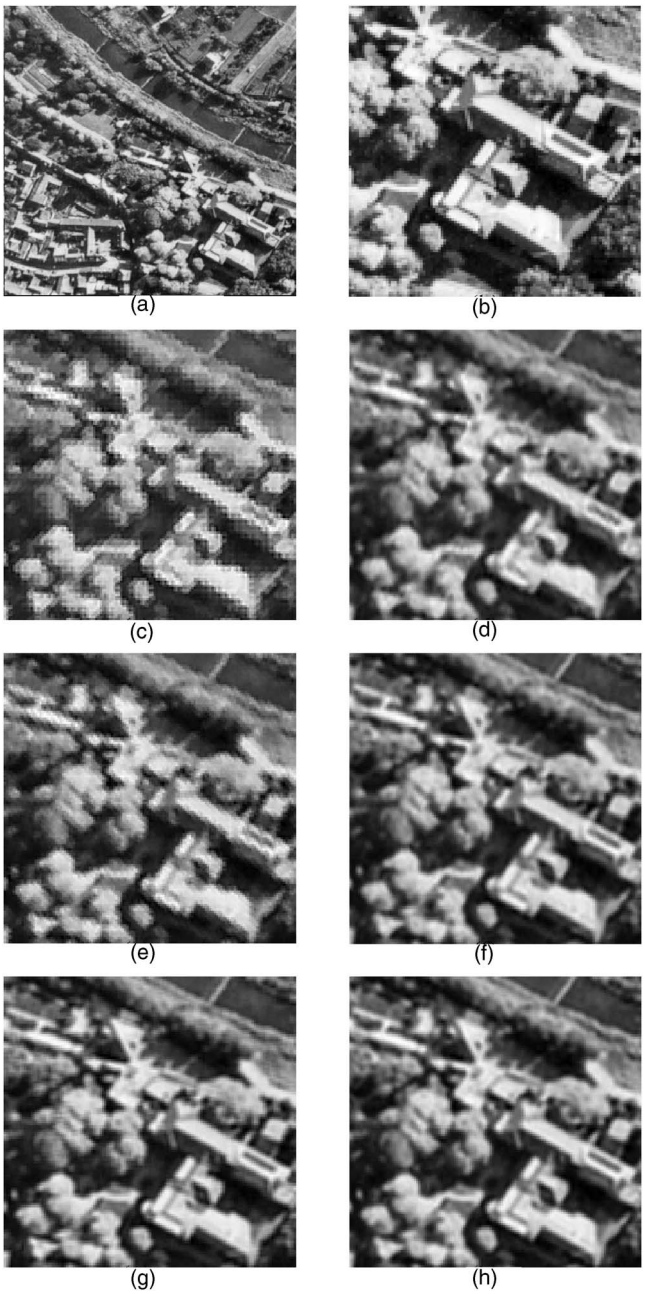



**Fig. 12** Results of super-resolution reconstruction ($L_1=L_2=3$): (a) full-size original aerial image; (b) enlarged portions of the original image; and that from WNN-based reconstruction in the case (c) $r=2$, (e) $r=4$, and (g) $r=9$; and partition-based interpolation in the case (d) $r=2$, (f) $r=4$, and (h) $r=9$.

interpolating in the color difference space, utilizing interchannel correlation, is more effective than interpolating in the original color space.[25] We apply the three demosaicking algorithms to the color difference space, $K_R=R-G$ and $K_B=B-G$. Note that missing pixels on the right-hand side of the formulas are obtained through bilinear interpolation as in Ref. 25. In the color difference space demosaicking processes, the G pixels at R pixel locations in the CFA array are estimated as $\hat{G}=R-K_R$, while those at the B pixel locations are estimated as $\hat{G}=B-K_B$. The R and B pixels are recovered by $\hat{R}=K_R+G$ and $\hat{B}=K_B+G$, respectively. (Please refer to Ref. 25 for more details on interpolating in difference color spaces.) Bilinear interpolation on the original color space is also applied as a benchmark. Optimization in this application, under scheme 1, requires a training image that has color difference images representative of the color difference in the observed image. Thus, the higher dimensionality of the problem makes selecting a representative image for demosaicking optimization difficult. We therefore utilize only training scheme 2 in this application. Specifically, the observed LR image is further mosaicked (down-sampled) by a Bayer pattern CFA. The original observed LR image thus becomes the desired HR image, whose pixels contribute to $\bar{\mathbf{y}}$ in Eq. (10). Obviously, pixels from the further mosaicked image form the observation vector $\mathbf{v}$. The size of the observation window is chosen as $3\times3$. An example of the observation window is shown in Fig. 3. Extensive image processing applications in Ref. 10 indicate that vector quantization utilizing 30 partitions is sufficient to capture all important underlying image structures in the $3\times3$ and $5\times5$ observation window cases. Experiments in Ref. 13 also indicate that 30 partitions are an appropriate choice for the image zooming application. Therefore, we choose $M=30$ for both the demosaicking and the super-resolution reconstruction applications.

Tables 1–3 list the PSNRs of demosaicking for each of the compared methods. As can be seen in each table, operating in the color difference space substantially improves demosaicking performance. Note that while edge-directed interpolation does not always yield higher PSNR than bilinear interpolation, it does produce images with better visual quality (Figs. 9 and 10). Partition-based interpolation outperforms bilinear interpolation both quantitatively and subjectively. In comparison with edge-directed interpolation, it yields a 1 to 4 dB gain in PSNR as well as improved visual quality. To aid in the visual comparison, enlarged portions of the images are shown in Figs. 9 and 10. As can be seen in the figures, significant visible color artifacts, such as the Zipper effect, are introduced by the bilinear interpolator. This can be seen, for instance, along the spokes of the bike in Fig. 10. Edge-directed and partition-based interpolation produce images with much better visual quality. Note that the Zipper effect is almost completely suppressed by partition-based demosaicking, and the false colors in the two windows at the bottom of the image are better suppressed (Fig. 9). The partition-based demosaicked image is indeed very close to the original nonmosaicked image. In the edge-directed demosaicking result, the Zipper effect, although suppressed, is still visible.

To gain further insight into the proposed interpolation method, the codebook and the weight matrices generated by the partition-based interpolation of the $K_R$ image of Bike, using training scheme 2 with $M=30$, are given in Figs. 11(a) and 11(b), respectively. Note that only the weight matrices used to reconstruct the $K_R$ pixel shaded in Fig. 11(d) are shown. The other pixel location matrices have similar structures. The bilinear weights in the form of a $3\times3$ weight matrix are also included for comparison [Fig. 11(c)]. As an inspection of the partitioning vectors shows, the codebook captures underlying image structures, including uniform regions and steps and corners oriented at

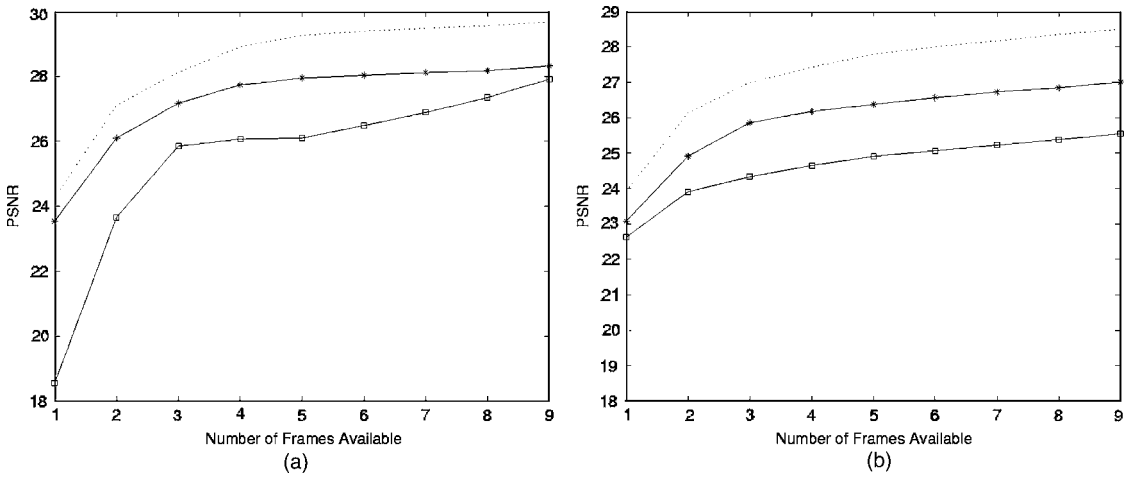



**Fig. 13** Reconstruction PSNR from partition-based and WNN-based reconstruction in the case where (a) no noise and (b) 20 dB noise is added in the observation model. *: partition-based reconstruction; □: WNN-based reconstruction; …: upper bound of partition-based interpolation (trained by HR test image).

different directions. Inspection of the weight matrices shows that the optimization produces weights that avoid interpolating across edges. Bilinear interpolation takes the average of the four top-right LR pixels. Thus, all other elements in the $3\times 3$ window are zero. The partition-based interpolation, in contrast, utilizes weights tuned to the underlying structure. This results, for instance, in weights similar to the bilinear weights in the uniform regions, and weights that, in edge regions, interpolate along, rather than across, edges.

### 4.2 *Super-Resolution Reconstruction*

Consider now the application of super-resolution reconstruction. Note that the proposed method combines partition-based interpolation with the gradient-based registration.[21] It is tested using both simulated LR frames and a true LR infrared sequence. This algorithm is compared with another registration-interpolation method that utilizes gradient-based registration, weighted nearest neighbor (WNN) interpolation,[1] and Wiener filtering (for restoration). WNN interpolation estimates each missing HR pixel using a weighted sum of the three nearest frames. The proposed algorithm is also compared with the MAP reconstruction introduced in Ref. 26 through experiments on real data.

Consider first the results for simulated LR sequences generated from the aerial image in Fig. 12(a), using the observation model in Fig. 4. The down sampling rate is $L_1=L_2=3$, resulting in nine LR frames, each numbered according to Fig. 5. The partitioning and WNN-based methods are tested for the cases in which $r=1,2,\ldots,9$ LR frames are available (observed). The combination of LR frames utilized in each case is given in Table 4. Note that the reference frame (frame 1) is always selected. The selection scheme guarantees that observed frames are spatially spread apart, so that both of the algorithms achieve their maximum performance under the same $r$. In the case $r=1$, the problem reduces to image magnification. Conversely,

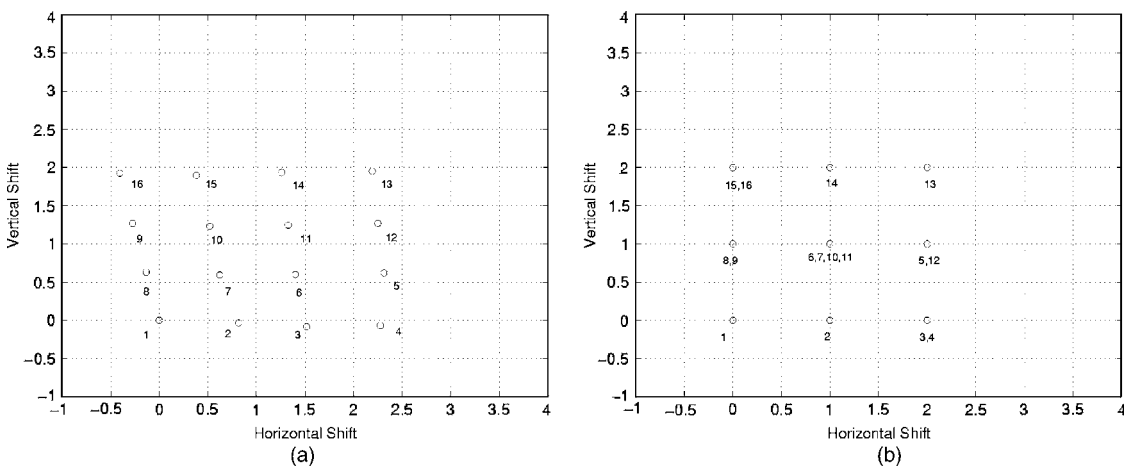



**Fig. 14** (a) Result of registration. (b) Motion shift after rounding. [The numbers represent the order in true image sequences. Shift is measured in terms of HR grids (dotted lines).]

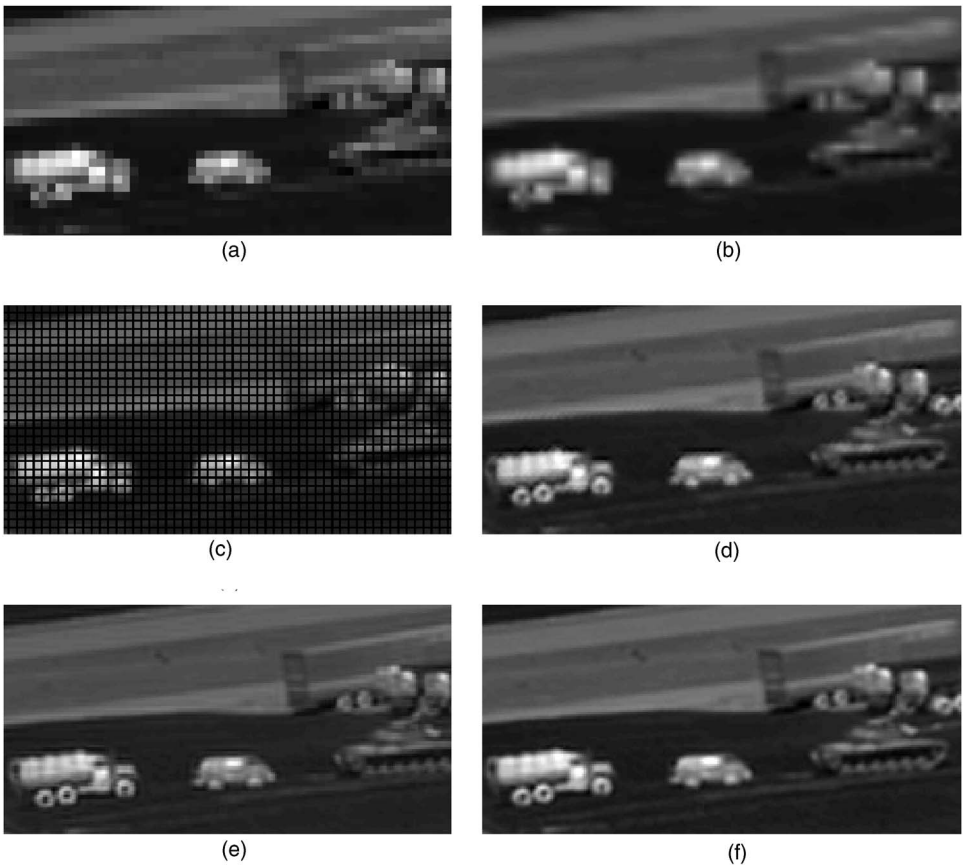



**Fig. 15** Results of super-resolution reconstruction: (a) nearest-neighbor and (b) bilinear method on one LR frame; (c) overlaid LR frames; (d) nonuniform interpolation followed by Wiener filtering; (e) MAP reconstruction; (f) partition-based interpolation.

when $r=9$, the problem becomes restoration only. For each $r$, we consider the noise-free and 20-dB additive noise cases. Also, we are constrained by the small pixel count of the observation frames, which makes self-similarity training (scheme 2) impractical. Additionally, for the experiment on simulated data, the registration step is not necessary, as the motion shift information is known. Thus, we utilize the well-known image Lena to train (scheme 1) the partition-based interpolation and the Wiener filter weights in the WNN-based methods. During the training of the partition-based interpolation, the image Lena is considered the desired HR image. It is down sampled and shifted to form a series of observed LR images. Then, $\bar{\mathbf{y}}$ and $\mathbf{v}$ are formed as described in Sec. 4, and the optimal $\tilde{\mathbf{W}}$ is obtained using Eq. (10).

An example of an observation window ($r=2$) is given in Fig. 6. In this case, the observation window covers a $3\times3$ window of pixels in the LR reference image as well as additional observed pixels in the HR grid $7\times7$ window. Thus, it is irregularly shaped. Moreover, the observation window is populated by different samples when different LR frames are observable. To simplify the optimization of the partition-based interpolation, the LBG algorithm is applied to an observation space formed by the full HR $7\times7$ window. This is possible since, during the training step, all samples inside the window are known. Thus, a single codebook of size $7\times7\times M$ is obtained for all cases of $r$. Also, when applying Eq. (4) to the test image in the partitioning step, the observation vector $\bar{\mathbf{x}}$ formed by $r$ observed samples, rather than the HR $7\times7$ window, is compared with the corresponding elements in the codewords, i.e., the 49 sample $(7\times7)$ codewords are appropriately subsampled and compared to the $r$ element observation window. This simplification does not have a material impact on the interpolation performance, but substantially reduces the computation complexity of the training process. Note that the training of weight matrix $\mathbf{W}$ in Eq. (10) is still based on $\bar{\mathbf{x}}$.

The reconstruction PSNR curves are shown, as a function of $r$, in Fig. 13, where the upper-bound PSNR of partition-based interpolation is obtained by training on the HR aerial image itself [Fig. 12(a)]. As expected, increasing $r$ leads to improved performance of both reconstruction algorithms. The partition-based algorithm produces more than a 2-dB gain over the WNN-based method for most $r$ values, and is only 0.7 to 1.5 dB below the upper bound.

The reconstructed HR images for $r=2$, 4, and 9 are compared in Fig. 12. For each $r$, partition-based interpolation yields the best performance, as an inspection of uniform and edge regions in the reconstructions shows. This is particularly true for small $r$ values. Note that because WNN interpolation disregards local image structures, a HR grid point may be estimated from samples far away, especially if

$r$ is small. In contrast, the partition-based method, by exploiting local pixel correlations, is able to produce satisfactory results with only a few observed LR frames.

Consider now true infrared images obtained from a staring infrared imaging system. 16 LR frames are obtained, each of size $64\times 64$. We elect to perform a $L_1=L_2=4$ reconstruction to obtain a visually acceptable resolution. The gradient-based registration[21] is employed to calculate the relative motion shift between each LR frame and the reference LR frame (Fig. 14). Each LR frame is rounded to its closest HR grid location as described in Sec. 2, after which the 16 LR frames fill 9 of the 16 HR grids [Fig. 14(b)]. In the following, we refer to the frames rounded to the same HR grid as one frame. Overlaying all the rounded LR frames yields Fig. 15(c), where the black grids represent missing HR samples. For the purpose of display, only the portion of the images that contain all the foreground objects is shown. Next, partition-based interpolation is applied to reconstruct the missing HR pixels and perform deconvolution of the system PSF, where it is assumed that no noise is present in the images. A codebook and weight matrix trained using the aerial image, Fig. 12(a), are utilized. In the WNN method, interpolation is applied on the fractional motion shift (without rounding). This is followed by a Wiener filter blur removal optimized using the same image.

Nearest-neighbor and bilinear interpolation images are shown for reference in Figs. 15(a) and 15(b), respectively. Both are interpolated from the LR reference frame to the HR grid, i.e., simple zooming is performed as a reference benchmark. A visual inspection of the HR results shows that the partition-based interpolator yields the best results, especially on important features such as the tank, trailers, and wheels of the truck. Although the proposed reconstruction method does not require a separate Wiener filter, there is an increase in computational cost associated with using the VQ partitioning. Thus, the cost for improved performance is a slight increase in computation complexity over traditional registration-interpolation methods.

## 5 Conclusions

Partition-based nonlinear signal processing is an effective method for processing nonstationary signals such as images.[10] The proposed partition-based interpolators capture a finite number of local image structures in images using VQ partitioning. Next, a linear operation, optimized for each specific structure, is utilized to obtain missing HR pixels using their neighboring observation pixels. A convenient two-stage optimization procedure is utilized to train partition-based interpolators. By exploiting local image structures, this optimization method adapts weights to avoid interpolating across edges. The presented results show that, customized to address the specific applications of CFA demosaicking and super-resolution reconstruction, partition-based interpolators produce superior results compared with other methods reported in the literature.

## References


1. M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda, "Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames," *IEEE Trans. Instrum. Meas.* **49**(3), 915–923 (2000).
2. W. K. Prat, *Digital Image Processing*, John Wiley & Sons, New York (1991).
3. A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation, Compression and Standards*, Plenum, New York (1995).
4. R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-29**(6), 1153–1160 (1981).
5. M. Unser, A. Aldroubi, and M. Eden, "Fast B-spline transforms for continuous image representation and interpolation," *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 277–285 (1991).
6. J. E. Adams, "Interactions between color plane interpolation and other image processing functions in electronic photography," *Proc. SPIE* **2416**, 144–151 (1995).
7. S. Battiato, G. Gallo, and F. Stanco, "A locally adaptive zooming algorithm for digital images," *Image Vis. Comput.* **20**(11), 805–812 (2002).
8. S. Carrato and L. Tenze, "A high quality $2\times$ image interpolator," *IEEE Signal Process. Lett.* **7**(6), 132–134 (2000).
9. X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.* **10**, 1521–1527 (Oct 2001).
10. M. Shao, "Partition-based weighted sum filtering theory with applications to image processing and biomedical engineering," PhD Thesis, Univ. of Delaware, Newark (Aug. 2004).
11. M. Shao and K. E. Barner, "Optimization of partition based weighted sum filters," *Proc. IEEE-EURASIP Nonlinear Sig. Image Process. (NSIP) Workshop* (2001).
12. K. E. Barner, A. M. Sarham, and R. C. Hardie, "Partition-based weighted sum filters for image restoration," *IEEE Trans. Image Process.* **8**, 740–745 (May 1999).
13. F. M. Candocia and J. C. Principe, "Superresolution of images with learned multiple reconstruction kernels," in *Multimedia Image and Video Processing*, L. Guan, S. Kung, and J. Larsen, Eds., CRC Press, New York (2001).
14. S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.* **20**(3), 21–36 (2003).
15. B. E. Bayer, "Color imaging array." U.S. Patent 3,971,065 (1976).
16. R. C. Hardie and K. E. Barner, "Nonlinear filters," in *Encyclopedia of Electrical and Electronics Engineering*, J. G. Webster, Ed., pp. 570–585, John Wiley & Sons, New York (1998).
17. J. Gillette, T. M. Stadmiller, and R. C. Hardie, "Aliasing reduction in staring infrared imagers utilizing subpixel techniques," *Opt. Eng.* **34**(11), 3130–3137 (1995).
18. S. P. Kim and N. K. Bose, "Reconstruction of 2-D bandlimited discrete signals from nonuniform samples," *IEE Proc. F, Radar Signal Process.* **137**(3), 197–204 (1990).
19. J. J. Clark, M. R. Palmer, and P. D. Laurence, "A transformation method for the reconstruction of functions from nonuniformly spaced samples," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-34**(5), 1151–1165 (1985).
20. K. D. Sauer and J. P. Allebach, "Interative reconstruction of bandlimited images from nonuniformly spaced samples," *IEEE Trans. Circuits Syst.* **CAS-34**(12), 1497–1506 (1987).
21. R. C. Hardie, K. J. Bernard, J. G. Bognara, E. E. Armstrong, and E. A. Watson, "High resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," *Opt. Eng.* **37**(1), 247–260 (1998).
22. Y. Linde, A. Buzo, and P. M. Gray, "An algorithm for vector qantization," *IEEE Trans. Commun. Theory* **28**, 84–95 (Jan. 1980).
23. M. Shao, K. E. Barner, and M. H. Goodman, "An interference cancellation algorithm for non-invasive extraction of TaFEEG," in *Chicago 2000 World Congress of Medical Physics and Biomedical Engineering* (Jul. 2000).
24. M. Hayes, *Adaptive Filter Theory*, John Wiley & Sons, New York (1996).
25. S. Pei and I. Tam, "Effective color interpolation in CCD color filter arrays using signal correlation," *IEEE Trans. Circuits Syst. Video Technol.* **13**(6), 503–513 (2003).
26. R. C. Hardie, K. J. Bernard, and E. E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.* **6**(12), 1621–1633 (1997).


**Min Shao** received his Bachelor of Engineering degree in electrical engineering from Tsinghua University, Beijing, China, in 1995. He received his PhD degree in electrical engineering from the University of Delaware, Newark, in 2004. He worked as a research assistant in the Department of Electrical and Computer Engineering, University of Delaware, from 1999 to 2004. From February to May 2003, he was a research intern with Siemens Corporate Research, Princeton, New Jersey, where he worked on medical image analysis and 3-D visualization. He is currently a research scientist with Philips Medical Systems. His research interests include signal processing and pattern recognition for ECG interpretive algorithms.

**Kenneth E. Barner** received a BSEE degree (*magna cum laude*) from Lehigh University, Bethlehem, Pennsylvania, in 1987. He received his MSEE and PhD degrees from the University of Delaware, Newark, in 1989 and 1992, respectively. For his dissertation "Permutation filters: A group theoretic class of nonlinear filters," he received the *Allan P. Colburn Prize in Mathematical Sciences and Engineering* for the most outstanding doctoral dissertation in the engineering and mathematical disciplines. He was the duPont Teaching Fellow and a visiting lecturer at the University of Delaware in 1991 and 1992, respectively. From 1993 to 1997 he was an assistant research professor in the Department of Electrical and Computer Engineering at the University of Delaware and a research engineer at the duPont Hospital for Children. He is currently a professor in the Department of Electrical and Computer Engineering at the University of Delaware. He is the recipient of a 1999 NSF Career award. He was the cochair of the *2001 IEEE-EURASIP Nonlinear Signal and Image Processing (NSIP) Workshop* and a guest editor for a special issue of the *EURASIP Journal of Applied Signal Processing on Nonlinear Signal and Image Processing*. He is a member of the Nonlinear Signal and Image Processing Board and is coediting the book *Nonlinear Signal and Image Processing: Theory, Methods, and Applications*. He is also serving as an associate editor of the *IEEE Transactions on Signal Processing*, the *IEEE Signal Processing Magazine*, and the *IEEE Transaction on Neural Systems and Rehabilitation Engineering*. He is also a member of the Editorial Board of the *EURASIP Journal of Applied Signal Processing*. His research interests include signal and image processing, robust signal processing, nonlinear systems, communications, haptic and tactile methods, and universal access. He is a member of Tau Beta Pi, Eta Kappa Nu, and Phi Sigma Kappa.

**Russell C. Hardie** graduated magna cum laude from Loyola College in Baltimore, Maryland, in 1988 with a BS degree in engineering science. He obtained MS and PhD degrees in electrical engineering from the University of Delaware in 1990 and 1992, respectively. He served as a senior scientist at Earth Satellite Corporation in Maryland prior to his appointment at the University of Dayton in 1993. He is currently a professor in the Department of Electrical and Computer Engineering and holds a joint appointment with the electro-optics program. Along with several collaborators, he received the Rudolf Kingslake Medal and Prize from SPIE in 1998 for work on multiframe image resolution enhancement algorithms. In 1999, he received the School of Engineering Award of Excellence in Teaching at the University of Dayton. He received the first annual Professor of the Year Award in 2002 from the student chapter of the IEEE at the University of Dayton. His research interests include a wide variety of topics in the area of digital signal and image processing, with emphasis on signal and image restoration.