

Temporal Color Video Demosaicking via Motion Estimation and Data Fusion

Xiaolin Wu, *Senior Member, IEEE*, and Lei Zhang, *Member, IEEE*

Abstract—Color demosaicking of charge-coupled device (CCD) data has been thoroughly studied for single-sensor still digital cameras. However, there has seemingly been little research on color demosaicking techniques for single-sensor video digital cameras. The temporal dimension of a color mosaic image sequence can reveal new information on the missing color components due to the mosaic subsampling, which is otherwise unavailable in the spatial domain of individual frames. This paper proposes a temporal approach to color demosaicking. A pixel of the current frame is matched to another in a reference frame via motion analysis, such that the CCD sensor samples different color components of the same object position in the two frames. The resulting inter-frame estimates of missing color components are fused with suitable intra-frame estimates to achieve a more robust color restoration. Our experimental results demonstrate clear advantages of the presented temporal color demosaicking approach over its intra-frame counterparts in reducing the color artifacts.

Index Terms—Bayer pattern, color demosaicking, data fusion, motion estimation, single-sensor digital video cameras.

I. INTRODUCTION

COLOR demosaicking of charge-coupled device (CCD) sensor data holds a key to the quality of color images reconstructed from single-sensor digital still and video cameras. Such digital cameras capture an image with a single-sensor array. At each pixel, only one of the three primary colors (red, green, and blue) is sampled. Fig. 1 shows the widely used Bayer color filter array (CFA) [3]. The full color image is reconstructed by interpolating the missing color samples. Color demosaicking has been extensively studied in spatial domain for still digital cameras [1], [2], [5]–[13], [15]–[21], [24], [25], [27], [28]. We call this class of color demosaicking techniques the intra-frame or spatial demosaicking.

The earlier spatial demosaicking methods, such as nearest-neighbor replication and bilinear and bicubic interpolation [15], can be simply implemented, but they suffer from many artifacts such as blocking, blurring, and zipper effect at edges. Lately developed demosaicking methods exploited the correlation between color channels. The smooth hue transition (SHT) methods [1], [6] assume images having slowly varying hue. SHT methods tend to cause large interpolation errors in the red and blue channels when green values abruptly change.

Manuscript received February 25, 2004; revised January 24, 2005. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grants IRCPJ 283011-01 and RGP45978-2000. This paper was recommended by Associate Editor S.-U. Lee.

The authors are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: xwu@mail.ece.mcmaster.ca; johnray@mail.ece.mcmaster.ca).

Digital Object Identifier 10.1109/TCSVT.2005.860938

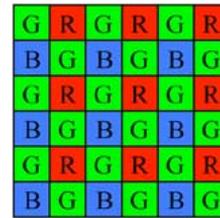


Fig. 1. Bayer pattern of color mosaic for digital cameras.

Since human visual systems are sensitive to the edge structures in an image, many adaptive demosaicking methods try to avoid interpolating across edges. In the second-order Laplacian filter proposed by Hamilton and Adams [2], [9], the second-order color gradients are used as the correction terms to interpolate the color channels. In the gradient-based scheme of Chang *et al.* [5], gradients in different directions are computed, and a subset of them is selected by adaptive threshold. The missing samples are estimated from the samples along the selected gradients. Recently, Zhang and Wu [28] proposed a linear minimum mean square-error (LMMSE) estimation-based demosaicking method and achieved very good results. They reconstructed the primary difference signals (PDS) between the green channel and the red or blue channel, instead of directly interpolating the missing color samples. In [18], Lukac and Plataniotis used a normalized color-ratio model in the color interpolation to suppress the color artifacts. They also proposed an edge-sensing method by using color correlation correction based on a difference plane model [19]. Some color demosaicking techniques are iterative schemes. Kimmel's two-step iterative demosaicking process consists of a reconstruction step and an enhancement step [13]. Another iterative demosaicking scheme was proposed by Gunturk *et al.* [7]. They reconstructed the color images by projecting the initial estimates onto so-called constraint sets. A wavelet-based iterative process was employed to update the high-frequency details of color channels. Other recently reported demosaicking methods include the method of adaptive homogeneity by Hirakawa and Parks [10], the primary-consistent soft-decision method of Wu and Zhang [27], the principal vector method of Kakarala and Baharav [11], and the bilinear interpolation of color difference by Pei and Tam [20].

However, if the image signal is highly discontinuous in both chrominance and luminance, spatial demosaicking techniques, including those recently developed sophisticated ones, are error prone, due to lack of correlation in both spectral and spatial domains. Some examples of color artifacts of demosaicking are presented in Fig. 2. The demosaicked images in Fig. 2 are produced by the recently proposed algorithm in [10], which is considered one of the best.

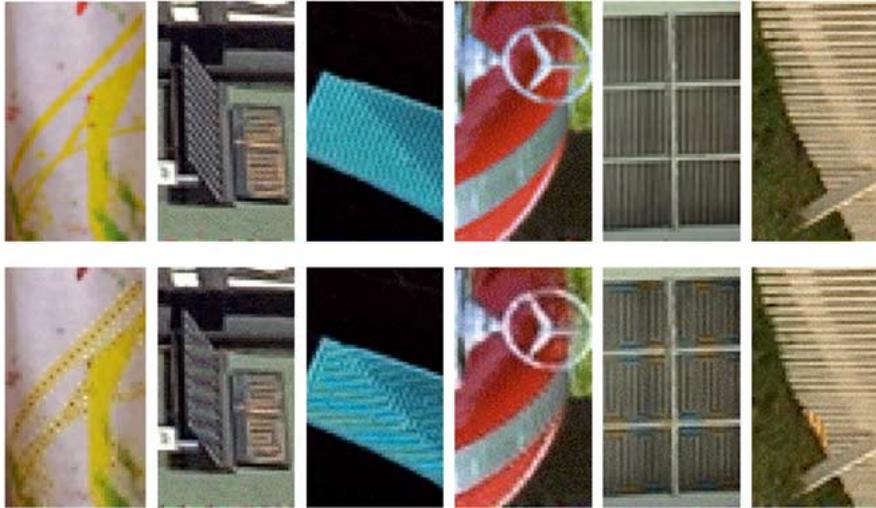


Fig. 2. Top row: original full color images. Bottom row: demosaicked images by the method in [10], one of the best spatial demosaicking methods in the literature. The color artifacts are clearly visible.

In order to overcome the limitation of spatial color demosaicking, additional knowledge and constraints of the original color signals are needed. For digital CCD video cameras, the temporal dimension of a sequence of color mosaic images often reveals more and new information on the color values that are not sampled by the CFA sensors. This potentially valuable information about the color composition of the scene would be unavailable in the spatial domain of individual mosaic frames. The correlation of adjacent frames can be exploited to aid the color demosaicking process if the camera and object motions can be estimated. We call this approach temporal color demosaicking.

However, there seems to be a lack of research reported on temporal color demosaicking, despite its obvious potential. In this paper, we present an effective temporal color demosaicking technique to enhance the color video quality. Without the loss of generality, we consider the Bayer CFA [3] that is widely used in digital color video cameras (see Fig. 1). The temporal demosaicking techniques to be developed in this paper can be readily generalized to other CFA patterns. In the Bayer pattern, the sampling frequency of the green channel is twice that of the red or blue channel. This is because the sensitivity of the human visual system peaks at the green wavelength, and the green channel contributes the most to the luminance of an image [4]. For natural images, there exists high spectral correlation between the red/blue and green channels. Once the green channel is interpolated with the help of the red/blue channel, it can then be used to guide the interpolation of the red/blue channel. The main idea of the proposed temporal demosaicking scheme is to match the CFA green sample blocks in adjacent frames in such a way that missing color samples in one frame can be inferred from available color samples of matched adjacent frames. Since the green channel has higher spatial resolution than the red/blue channel, it is naturally employed in the motion-estimation process of the proposed temporal demosaicking approach.

In order to feed the motion analysis with sufficient information, the green channels of all frames are first reconstructed individually by interpolating the missing green samples via intra-frame demosaicking. Motion estimation between ad-

acent frames for temporal color demosaicking is based on this reconstructed green image sequence. With the estimated motion vectors, adjacent frames are registered spatially. The best matched green samples in adjacent reference frames are then fused with the intra-frame estimates of the missing green samples of the current frame to improve the quality of the previously estimated green channel. The resulting improved green channel will serve as an anchor to reconstruct the red and blue channels by interpolating the missing red and blue samples using both the intra-frame and inter-frame information.

The paper is structured as follows. In Section II, we present a technique for temporal demosaicking of the green channel through motion estimation and data fusion. This temporal demosaicking technique is then extended to reconstruct the missing red/blue samples at the blue/red sample positions of CFA in Section III, and to reconstruct the missing red/blue samples at the green sample positions of CFA in Section IV. Section V proposes a new system workflow for digital video cameras to best realize the potential of temporal color demosaicking and to keep its computational complexity reasonable. Section VI presents experimental results, and Section VII concludes the paper.

II. TEMPORAL DEMOSAICKING OF THE GREEN CHANNEL

In order to perform motion estimation in maximum possible spatial resolution, we first demosaick the green channel of each frame separately. This can be done by any of the intra-frame demosaicking methods. In this paper, we adopt the directional second-order Laplacian interpolation filter proposed by Hamilton and Adams [9] for its good performance and low complexity. The green estimates by intra-frame demosaicking are to be improved by motion estimation and sample registration in adjacent frames.

Referring to Fig. 3, we denote the original green samples by G and the interpolated green samples through intra-frame demosaicking by \hat{G} . Obviously, due to the sampling structure of the Bayer CFA, for an $N_1 \times N_2$ image, there are $N_1 \times N_2/2$ original G samples, and $N_1 \times N_2/2$ interpolated \hat{G} samples,

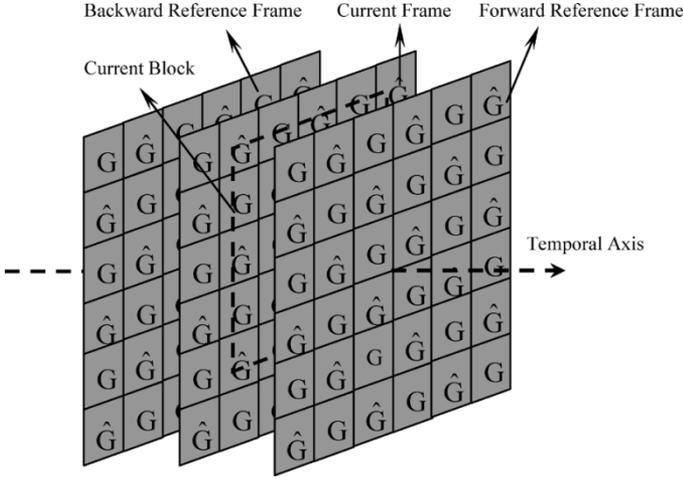


Fig. 3. Current green frame and its backward and forward neighboring frames.

which are to be temporally updated. In the temporal improvement of these samples in the current frame, we employ K adjacent reference frames: $K/2$ backward and $K/2$ forward. For each sample \hat{G} in the current frame, we use a block centered at it to find the best matched sample from the K reference frames by block matching. Fig. 4 illustrates the reference sample matched to the considered sample $\hat{G}(0)$ from the k th reference frame, $k = 1, \dots, K$. We denote original and interpolated green samples in the k th reference frame as G^k and \hat{G}^k , respectively. The dashed green squares on a red/blue background represent the green estimates at these positions that are generated by intra-frame interpolation.

Referring to Fig. 4(a), consider the 5×5 block centered at $\hat{G}(0)$, the sample to be temporally updated, in the current frame, and denote this block as \mathbf{M} . Let a corresponding block in the k th reference frame with displacement $(\Delta x, \Delta y)$ be $\mathbf{M}^k(\Delta x, \Delta y)$. Fig. 4(b) shows, for example, the relationship between \mathbf{M} and $\mathbf{M}^k(1, 0)$. Due to the structure of the Bayer pattern, if the motion vector $(\Delta x, \Delta y)$ satisfies

$$(\Delta x, \Delta y) = (2i, 2j + 1) \text{ or } (\Delta x, \Delta y) = (2i + 1, 2j) \quad (2-1)$$

where i, j are integers, then $\hat{G}(0)$ can be matched to an original green sample $G^k(0)$ in the k th reference frame. Denote by Θ the set of all motion vectors $(\Delta x, \Delta y)$ determined by (2-1) in a suitable search range $\{-Z \leq \Delta x \leq Z, -Z \leq \Delta y \leq Z\}$, where Z is an integer.

Considering block \mathbf{M} as a vector consisting of the 25 green samples in a 5×5 window, no matter original or estimated by intra-frame demosaicking; likewise, block $\mathbf{M}^k(\Delta x, \Delta y)$ is the corresponding vector in the k th reference block. Applying the block matching of \mathbf{M} to all the K reference frames, we select the best matched block of \mathbf{M} from all the $\mathbf{M}^k(\Delta x, \Delta y)$, $k = 1, \dots, K$. Let

$$(\kappa, dx, dy) = \arg \min_{\substack{k \in \{1, 2, \dots, K\} \\ (\Delta x, \Delta y) \in \Theta}} \|\mathbf{M} - \mathbf{M}^k(\Delta x, \Delta y)\| \quad (2-2)$$

$$\varepsilon = \|\mathbf{M} - \mathbf{M}^\kappa(dx, dy)\|. \quad (2-3)$$

If the difference value ε is no more than a preset threshold T , i.e., $\varepsilon \leq T$, then block $\mathbf{M}^\kappa(dx, dy)$ is taken as the best matched reference block of current block \mathbf{M} . If the scenes in

adjacent frames change sharply, the redundancies between adjacent frames are low, and ε will be of high value. In this case, the unreliable reference block found by (2-2) can be ruled out by the threshold T . The value of T can be determined in a training process. In this paper, we set it as $T = 0.1\|\mathbf{M}\|$. The block-based motion-estimation scheme is similar to that of MPEG [14]. The differences are that the motion vectors are confined to the set Θ , and multiple reference frames are searched. The objective here is to find the existing green sample $G^\kappa(0)$ in the best matched block $\mathbf{M}^\kappa(dx, dy)$ that provides a good measurement of the missing green sample $\hat{G}(0)$ in the current frame.

This technique works well if the following three conditions are met: the object/camera motion is translation; the luminance does not change in a small time window; and a matched block $\mathbf{M}^\kappa(dx, dy)$ can be found. For relatively high frame rates, the first two conditions can be met satisfactorily, because the motion vectors are small and the illumination in the scene changes little, if at all. Regarding the last requirement, there is a probability of $1/2$ that the motion vector is of the form in (2-1), i.e., the offsets in the x and y directions have different parities. Under the assumptions that motion vectors are measured up to the precision of integer pixels, and motions are equally probable in all directions, there is equal chance for the parities of the motion offsets $(\Delta x, \Delta y)$ to be one of the four combinations (even, even), (odd, odd), (even, odd), (odd, even). Thus, if one reference frame is used, the probability of not finding an existing reference sample for the considered sample $\hat{G}(0)$ (i.e., not satisfying (2-1)) is $1 - (1/2) = 1/2$. If two reference frames are used, the probability of finding a reference sample for $\hat{G}(0)$ from any of the two reference frames is $1 - (1/2)^2 = 3/4$. Therefore, if one searches K adjacent frames, the probability of finding a reference sample from any one of the K frames is very high, namely, $1 - 2^{-K}$. On the rare occasion when the matching criterion $\|\mathbf{M} - \mathbf{M}^\kappa(dx, dy)\| < T$ cannot be met, then no temporal estimate $G^\kappa(0)$ will be generated, and we resort to intra-frame demosaicking only.

If the motion estimation yields the best matched green sample $G^\kappa(0)$ that spatially corresponds to $\hat{G}(0)$, we fuse $\hat{G}(0)$ and $G^\kappa(0)$ to get a more robust estimate of $G(0)$, the unknown true green value at the position of $\hat{G}(0)$. Both $\hat{G}(0)$ and $G^\kappa(0)$ can be viewed as the noisy measurements of true value $G(0)$, and can be written as

$$\begin{cases} \hat{G}(0) = G(0) + v_1 \\ G^\kappa(0) = G(0) + v_2 \end{cases} \quad (2-4)$$

where terms v_1 and v_2 are the measurement errors of $\hat{G}(0)$ and $G^\kappa(0)$. Generally, v_1 and v_2 are zero mean and uncorrelated with each other.

We employ the weighted-average strategy to fuse $\hat{G}(0)$ and $G^\kappa(0)$

$$G_f(0) = a_1\hat{G}(0) + a_2G^\kappa(0). \quad (2-5)$$

To make $G_f(0)$ an unbiased estimator of $G(0)$, i.e., $E[G_f(0)] = E[a_1\hat{G}(0) + a_2G^\kappa(0)] = (a_1 + a_2)E[G(0)] = E[G(0)]$, we let $a_1 + a_2 = 1$. The optimal weights to minimize the mean square error (MSE) of fused green value $G_f(0)$ are given by

$$(a_1, a_2) = \arg \min_{a_1 + a_2 = 1} E \left[(G_f(0) - G(0))^2 \right]. \quad (2-6)$$

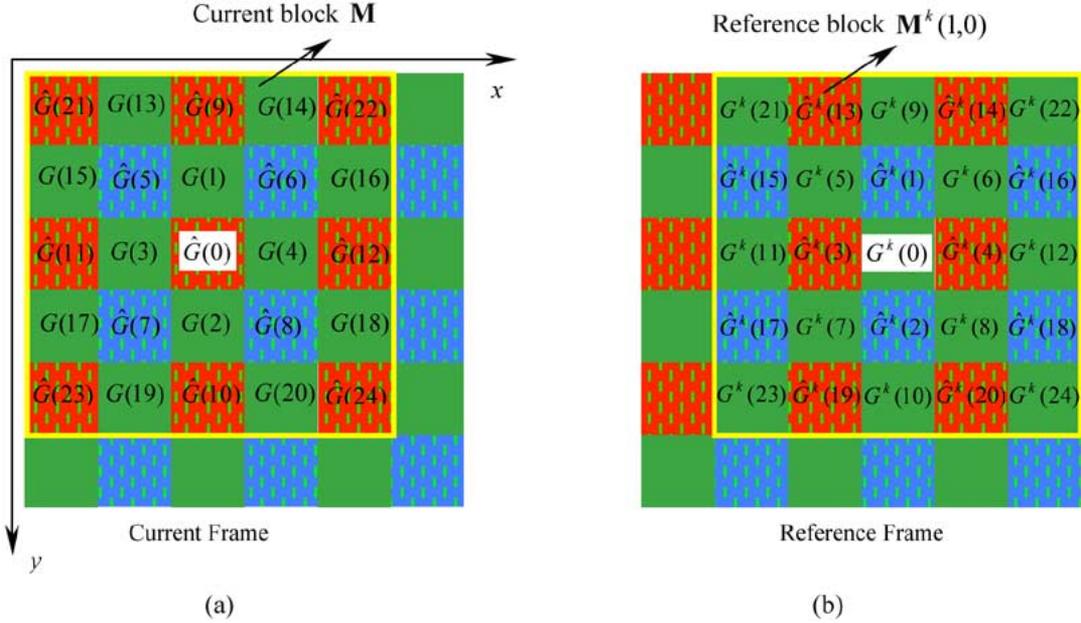
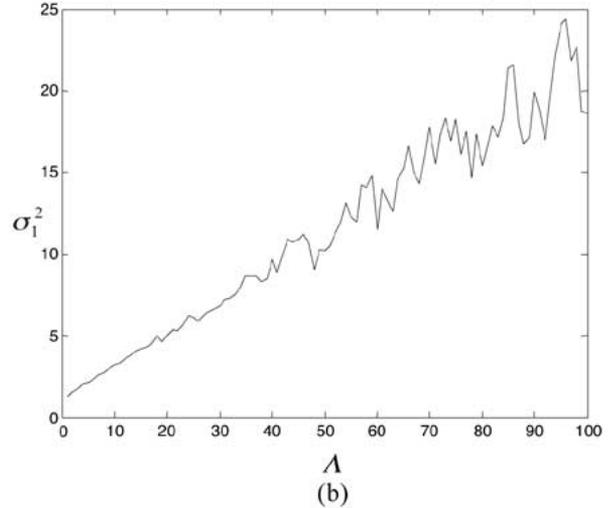


Fig. 4. Registration of green samples of the Bayer pattern in current and reference frames. Solid green squares are the green pixel positions of CFA; the dashed green squares on red/blue background are the red/blue pixel positions of CFA, where the missing green samples are interpolated by intra-frame demosaicking.



(a)



(b)

Fig. 5. (a) A test image. (b) Curve of σ_1^2 versus λ for the test image.

Denote by σ_1^2 and σ_2^2 the variances of noises v_1 and v_2 . Since v_1 and v_2 are uncorrelated, we have

$$\begin{aligned}\Omega &= E \left[(G_f(0) - G(0))^2 \right] \\ &= E \left[(a_1 v_1 + (1 - a_1) v_2)^2 \right] \\ &= a_1^2 v_1^2 + (1 - a_1)^2 v_2^2.\end{aligned}$$

Differentiating Ω with respect to a_1 and letting it be zero, we have

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad \text{and} \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}. \quad (2-7)$$

σ_1^2 and σ_2^2 can be estimated as follows. Referring to Fig. 4(a), we see that the four nearest neighbors of $\hat{G}(0)$ are the original

green pixels $G(1)$, $G(2)$, $G(3)$, and $G(4)$. Write the variance of $\hat{G}(0) - G(i)$, $i \in \{1, 2, 3, 4\}$ as

$$\Lambda = \frac{1}{4} \sum_{i=1}^4 (\hat{G}(0) - G(i))^2. \quad (2-8)$$

We empirically observed that σ_1^2 is nearly linearly proportional to Λ . This observation reflects the fact that demosaicking artifacts typically accompany high-frequency contents. Fig. 5(a) shows a test image, and Fig. 5(b) plots the curve of σ_1^2 versus Λ for the test image. The curves for other images are similar. These curves suggest that $\sigma_1^2 \approx \lambda \cdot \Lambda$, and $\lambda = 0.2$.

Having the estimated σ_1^2 , we proceed to estimate σ_2^2 . Based on (2-4) and the assumption that v_1 and v_2 are uncorrelated, we have

$$E \left[(G^k(0) - \hat{G}(0))^2 \right] = \sigma_1^2 + \sigma_2^2. \quad (2-9)$$

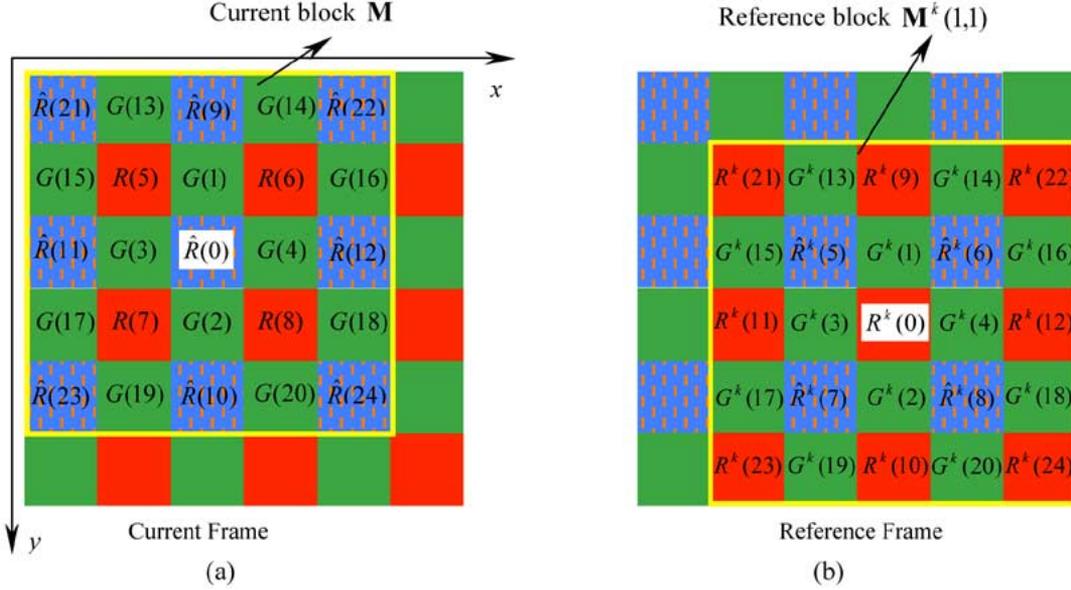


Fig. 6. Registration of red samples of the Bayer pattern in current and reference frames. Solid red squares are the red pixel positions of CFA; the dashed red squares on a blue background are the blue pixel positions of CFA, where the missing red samples are interpolated by intra-frame demosaicking.

Referring to Fig. 4(b), the four nearest original green pixels of $G^\kappa(0)$ in the chosen reference frame are $G^\kappa(5)$, $G^\kappa(6)$, $G^\kappa(7)$, and $G^\kappa(8)$, and we estimate σ_2^2 as

$$\sigma_2^2 = \frac{1}{5} \sum_{i=0,5,6,7,8} \left(G^\kappa(i) - \hat{G}(i) \right)^2 - \sigma_1^2. \quad (2-10)$$

Now both σ_1^2 and σ_2^2 are obtained, and then the fusing weights (a_1, a_2) are determined by (2-7). Finally, the missing green samples are recovered by the fused green estimates $G_f(0)$ made by (2-5).

III. RECOVERY OF THE MISSING RED/BLUE SAMPLES AT BLUE/RED PIXEL POSITIONS

An inherent drawback of the Bayer pattern is its inferior sampling scheme for the blue and red color components. Not only is the sampling frequency of red and blue only half of the sampling frequency of green, the 2-D sampling grid for red and blue also has a poor shape of square lattice, which deviates greatly from the optimal hexagonal lattice. By contrast, the green color component is preferentially sampled with twice as many samples as red and blue, and a far more efficient sampling grid of checkerboard lattice. Naturally, we seek ways of using the denser and better-shaped green sample grid to increase the sampling frequency of red and blue via temporal color demosaicking. Specifically, in this section, we discuss the temporal demosaicking technique to recover the missing red/blue samples at the blue/red sample positions in CFA, and then in the next section, the technique to recover the remaining missing red/blue samples at the green sample positions.

With the help of the already temporally recovered green channel, the missing red/blue samples at original blue/red positions are first spatially interpolated by using a bilinear strategy: 1) compute the average green/red or green/blue color difference with the four nearest neighbors along diagonal directions (45° and 135°); and 2) subtract the computed color

difference from the green sample to get the missing red or blue sample. Note that such a spatial interpolation of red/blue samples also exploits some temporal information, because the employed green channel has been temporally updated.

The spatially recovered red/blue samples are to be temporally enhanced. Due to the symmetry of the Bayer pattern, updating the red sample at the blue pixel position is the same problem as updating the blue sample at the red pixel position by simply switching the role of red and blue. Therefore, it suffices to only discuss the former case. Referring to Fig. 6, denote by R the original red sample and by \hat{R} the red estimate at the blue pixel position by spatial demosaicking. Correspondingly, let R^k and \hat{R}^k be the original and estimated red samples in the k th reference frame. Notice that at the original red and blue sample positions, we have temporally recovered the green samples, and we label these estimated green samples by \hat{G} and \hat{G}^k .

Again, we use \mathbf{M} to denote the 5×5 block centered at the interested sample $\hat{R}(0)$ in the current frame, and $\mathbf{M}^k(\Delta x, \Delta y)$ for the corresponding block in the k th reference frame with displacement $(\Delta x, \Delta y)$. It is easy to see from the sample layout of Bayer pattern that if the motion vector $(\Delta x, \Delta y)$ is of the form

$$(\Delta x, \Delta y) = (2i + 1, 2j + 1) \quad (3-1)$$

then $\hat{R}(0)$ can be matched to an original red sample $R^k(0)$ in the k th reference frame. The same motion-estimation technique of Section II can be used to search for the best matched block $\mathbf{M}^k(dx, dy)$ in the adjacent frames. Because the green channel is the most reliable one in the three color channels, we conduct motion estimation based on green samples, despite searching for the matching blocks of red/blue samples.

Fig. 6 shows that when the motion vector $(\Delta x, \Delta y)$ satisfies (3-1), an original green sample G in the current block \mathbf{M} will match an original green sample G^k in the reference block $\mathbf{M}^k(\Delta x, \Delta y)$, while an estimated green sample \hat{G} in \mathbf{M} will match an estimated green sample \hat{G}^k in $\mathbf{M}^k(\Delta x, \Delta y)$. We denote by \mathbf{M}_o the vector consisting of all 12 original

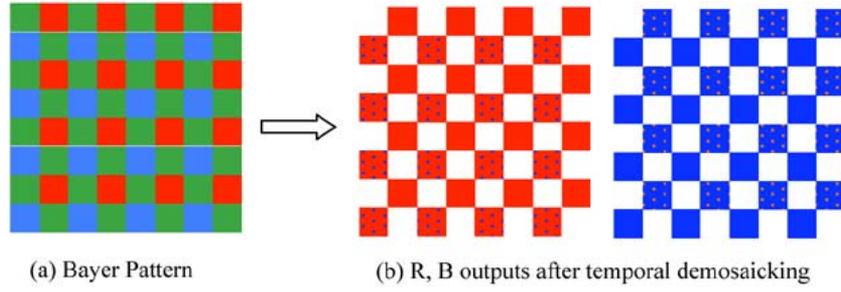


Fig. 7. Red and blue channel outputs after temporally recovering the missing red/blue samples at blue/red pixels. (a) Bayer pattern. (b) R, B outputs after temporal demosaicking.

green samples in block \mathbf{M} , and by \mathbf{M}_e the vector consisting of all 13 estimated green samples in block \mathbf{M} . Correspondingly, $\mathbf{M}_o^k(\Delta x, \Delta y)$ and $\mathbf{M}_e^k(\Delta x, \Delta y)$ denote the vectors of the original and estimated green samples in reference block $\mathbf{M}^k(\Delta x, \Delta y)$. When measuring the difference between \mathbf{M} and $\mathbf{M}^k(\Delta x, \Delta y)$, we place higher confidence on the original green samples than on the estimated ones, namely

$$(\kappa, dx, dy) = \arg \min_{\substack{\kappa \in [1, \dots, K] \\ (\Delta x, \Delta y) \in \Theta}} (c_o \cdot \|\mathbf{M}_o - \mathbf{M}_o^k(\Delta x, \Delta y)\| + c_e \cdot \|\mathbf{M}_e - \mathbf{M}_e^k(\Delta x, \Delta y)\|) \quad (3-2)$$

where confidence factors $c_o > 1$ and $c_e < 1$, and Θ is the set of motion vectors $(\Delta x, \Delta y)$ of (3-1) in a suitable search range $\{-Z \leq \Delta x \leq Z, -Z \leq \Delta y \leq Z\}$. Our experiments showed that $c_o = 1.2$ and $c_e = 0.8$ worked well. Nonetheless, the demosaicking results are not sensitive to the values of c_o and c_e .

Let $\mathbf{M}^\kappa(dx, dy)$ be the best matched block to \mathbf{M} over all searched frames, and let $R^\kappa(0)$ be the original red sample at the center of $\mathbf{M}^\kappa(dx, dy)$ that spatially corresponds to $\hat{R}(0)$. We fuse $R^\kappa(0)$ and $\hat{R}(0)$ to achieve a better estimation of $R(0)$: $R_f(0) = a_1 \hat{R}(0) + a_2 R^\kappa(0)$. As in Section II, the weights (a_1, a_2) are determined by (2-7). The sample set $\{R(5), R(6), R(7), R(8)\}$ is used to estimate the measurement error of $\hat{R}(0)$, i.e., the variance σ_1^2 ; similarly, the sample set $\{R^\kappa(0), R^\kappa(9), R^\kappa(10), R^\kappa(11), R^\kappa(12)\}$ is used to estimate the measurement error of $R^\kappa(0)$, i.e., the variance σ_2^2 .

The above scheme reconstructs the missing red sample at the blue pixel position and the missing blue sample at the red pixel position in the current frame. Consequently, the resolution of red and blue components can be increased to be as high as that of the green component. Fig. 7 depicts the outcomes of the red and blue channels after such temporal demosaicking, in which the dotted cells represent those recovered missing color components. The temporal demosaicking of red and blue channels generates a checkerboard pattern of either of these two color components. This facilitates spatial interpolation of the other missing red and blue samples.

IV. RECOVERY OF THE MISSING RED/BLUE SAMPLES AT THE GREEN PIXEL POSITIONS

Having doubled the number of red and blue samples in the previous section, there are still half of the red and blue samples missing at the positions of original green samples. From Fig. 7(b), we observe that there are four neighbors of the same

color around each missing red or blue sample at the green pixel positions, two original and two estimated. With the help of these neighbors and the already recovered green channel, the missing red and blue samples are first spatially interpolated by the bilinear strategy described in Section III (any good spatial interpolation method will do). Then the spatially estimated red/blue samples are to be improved by additional information from adjacent frames. Next, we describe the temporal demosaicking process for the red channel. The case for the blue channel can be treated analogously.

As shown in Fig. 8, the temporally demosaicked red samples (by the scheme in Section III) at the blue pixels are represented by red cells with blue dots. The spatially demosaicked red samples at the green pixels are represented by dashed red cells with a green background. The current block \mathbf{M} is defined as the 5×5 windows centered at the considered sample $\hat{R}(0)$ in the current frame. It is to be matched to a reference block $\mathbf{M}^k(\Delta x, \Delta y)$, whose center is an original red sample $R^k(0)$. Again, it should be stressed that the green channel is employed in searching for the best matched block.

It can be seen that for a spatially interpolated red sample that lies to the right/left of an original red sample, if the motion vector is of form $(\Delta x, \Delta y) = (2i, 2j + 1)$, then $\hat{R}(0)$ can be matched to an original red sample $R^k(0)$ in the k th reference frame. For those spatially interpolated red samples that are above/below the original red pixels, the desired motion vector is $(\Delta x, \Delta y) = (2i + 1, 2j)$. The motion-estimation process is similar to that in Section II. After finding the best matched block $\mathbf{M}^\kappa(dx, dy)$ of \mathbf{M} , $\hat{R}(0)$ and $R^\kappa(0)$ are fused to generate $R_f(0) = a_1 \hat{R}(0) + a_2 R^\kappa(0)$. The calculation of weights (a_1, a_2) is done in the same way as in Sections II and III.

Summarizing all the steps described in the preceding sections, we present the proposed temporal demosaicking algorithm in the following pseudocode.

- 1) Spatially interpolate individual green frames. (The second-order Laplacian filter in [9] is used in this paper.)
- 2) Temporally update the green channel via the procedure developed in Section II.
- 3) Spatially interpolate the missing red/blue samples at the original blue/red pixel positions by using the temporally updated green channel in step 2) to bilinearly interpolate the green/red or green/blue color difference.
- 4) Temporally update the red/blue samples interpolated in step 3), as described in Section III.
- 5) Spatially interpolate the missing red/blue samples at original green pixel positions, similar to step 3).

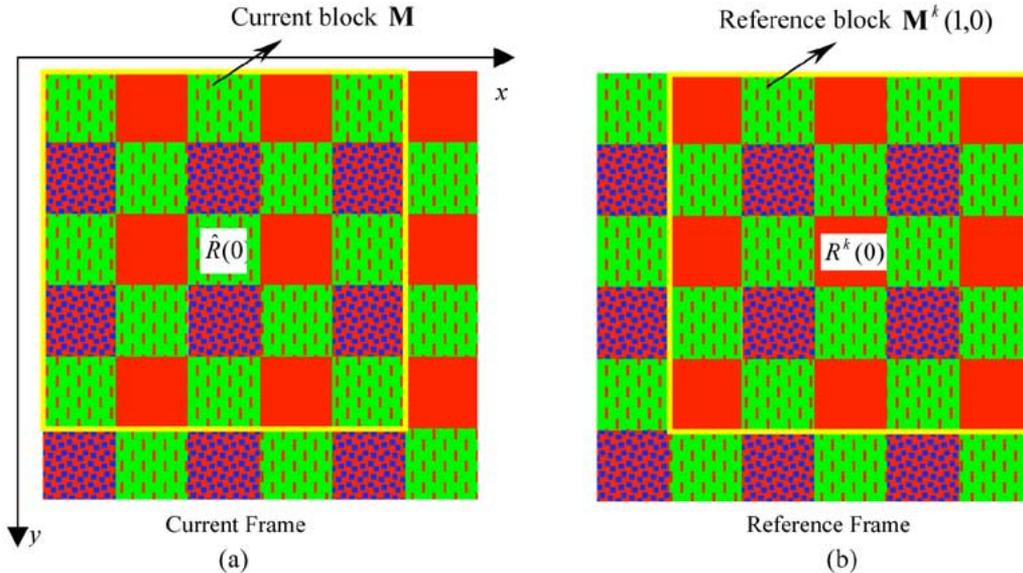


Fig. 8. Registration of red samples of the Bayer pattern in current and reference frames. Solid red squares are the red pixel positions of CFA; the red squares with blue dots are the blue pixel positions of CFA where the missing red samples have already been temporally recovered; the dashed red squares on green background are the green pixel positions of CFA where the missing red samples are interpolated by intra-frame demosaicking.

- 6) Temporally update the red/blue samples interpolated in step 5), as described in Section IV.

V. SYSTEM WORKFLOW AND COMPLEXITY

The workflow of current video capture devices is spatial color demosaicking followed by lossy video compression via either intra-frame (e.g., motion JPEG) or inter-frame coding (e.g., MPEG). In this system workflow, color demosaicking is a real-time process, and only a simple spatial demosaicking algorithm can be applied. Lossy compression of demosaicked video further aggravates the problem. MPEG compression, for instance, introduces artifacts of its own on high-frequency contents due to errors in motion vectors. Therefore, lossy compression of a spatially demosaicked video sequence denies the opportunity of achieving the best video fidelity allowed by the original mosaic data, even ample computation resources and time are permitted at a later time.

For high-end applications such as digital cinema, where the visual quality has paramount importance, spatio-temporal color demosaicking should be performed on original mosaic data. This can be achieved by a new system workflow. First, the captured raw mosaic data are compressed by a real-time mathematically lossless or near-lossless coder [29] and stored on camera. Then, in an offline process, the raw mosaic data are decompressed and processed by spatio-temporal demosaicking to reconstruct the full color video sequence. Finally, the high-quality demosaicked video sequence is compressed, possibly by MPEG, to meet the bandwidth requirement of the application.

A seeming drawback of the new workflow is the relatively low compression ratio obtainable by lossless coding of raw mosaic data. However, keep in mind that the raw mosaic image is only one-third of the demosaicked image in size. This effectively triples the compression ratio if measured in terms of the size of the demosaicked image. In fact, color demosaicking makes the task of compression more difficult. It increases the amount of

input data twofold. Ironically, a necessary step of compression is to decorrelate the color bands, which is essentially an attempt to reverse the color demosaicking process. Obviously, a direct compression on the raw mosaic data can avoid such problems.

Admittedly, the proposed spatiotemporal demosaicking is computationally more expensive than spatial demosaicking techniques. The complexity of the proposed spatio-temporal demosaicking is dominated by the computation cost of motion estimation. However, this cost can be shared with the process of video compression that is to immediately follow spatio-temporal demosaicking in the new system workflow. In video compression, motion estimation is also an indispensable and computationally expensive step. Since the motion vectors computed for the purpose of temporal demosaicking are needed in video compression anyway, the overall system complexity remains roughly the same, regardless of whether spatial or spatiotemporal demosaicking is used.

One can also significantly reduce the complexity of spatio-temporal demosaicking by invoking it only when spatial demosaicking cannot produce good outputs. In smooth regions, which typically constitute the major portion of an image, the sampling frequency of the color mosaic is high enough to allow correct color demosaicking solely in the spatial domain. Only at localities of sharp edges and finely structured textures the CPU-intensive temporal color demosaicking will be activated.

VI. EXPERIMENTAL RESULTS

We present the experimental results on two video clips to evaluate the proposed temporal demosaicking algorithm in comparison with nine existing methods. The first video sequence is originally captured on film at a rate of 24 frames/second (fps) and then digitized by a high-resolution scanner. The frame spatial resolution is 1948×1280 , and Fig. 9(a) shows the scene of it. The original image has all three of the red, green, and blue color channels in full resolution, which provides the true sample



Fig. 9. (a) Scene in the first test clip. (b) Scene in the second test clip.



Fig. 10. (a) Original full-color image. Demosaicked images by the methods in (b) [9], (c) [5], (d) [11], (e) [18], and (f) [19]. Demosaicked images by the methods in (g) [17], (h) [20], (i) [10], (j) [28], (k) [26], and (l) the proposed method.

values to measure the demosaicking errors. The mosaic data are simulated by subsampling the true color image, according to the Bayer pattern. The second video sequence is captured directly by a single-sensor digital video camera at a rate of 25 fps. The spatial resolution is 640×480 , and Fig. 9(b) shows the scene of it (the mosaic image is displayed as a gray image).

Nine state-of-the-art spatial demosaicking algorithms and our earlier temporal method in [26] are included in our comparison study. The spatial methods are the second-order Laplacian filtering by Hamilton and Adams [9], the gradient-based method by Chang *et al.* [5], the principal vector method by Kakarala and Baharav [11], the bilinear interpolation of color difference by Pei and Tam [20], the normalized color-ratio modeling by Lukac and Plataniotis [18], the dif-

ference-plane-based color correlation correction by Lukac *et al.* [19], the demosaicked image postprocessing scheme by Lukac *et al.* [17] (in our experiments, the associated demosaicking process is [9]), the method of adaptive homogeneity by Hiraoka and Parks [10], and the directional filtering and fusion method by Zhang and Wu [28].

In the first movie clip, on the car, where some sharp color edges happen, spatial demosaicking produces severe color artifacts. Fig. 10(a) shows a 120×150 portion of the original frame in the test sequence. Fig. 10(b)–(j) are the demosaicked images by the spatial methods in [5], [9]–[11], [17]–[20], and [28]. There are highly visible color artifacts in these reconstructed images. The color edges, where spatial demosaicking algorithms fail, have discontinuities in both luminance and

TABLE I
PSNR RESULTS OF THE 11 DEMOSAICKING METHODS FOR THE FIRST VIDEO SEQUENCE

Methods		Method in [9]	Method in [5]	Method in [11]	Method in [18]	Method in [19]	Method in [17]	Method in [20]	Method in [10]	Method in [28]	Method in [26]	Proposed
PSNR(dB)	R	31.11	31.49	29.73	29.27	30.13	27.68	29.20	28.39	28.73	31.13	33.05
	G	34.15	33.51	34.16	32.52	33.46	33.00	33.58	33.52	33.68	33.48	35.01
	B	28.37	28.20	28.83	28.73	28.72	28.29	28.26	27.82	28.05	28.10	29.60

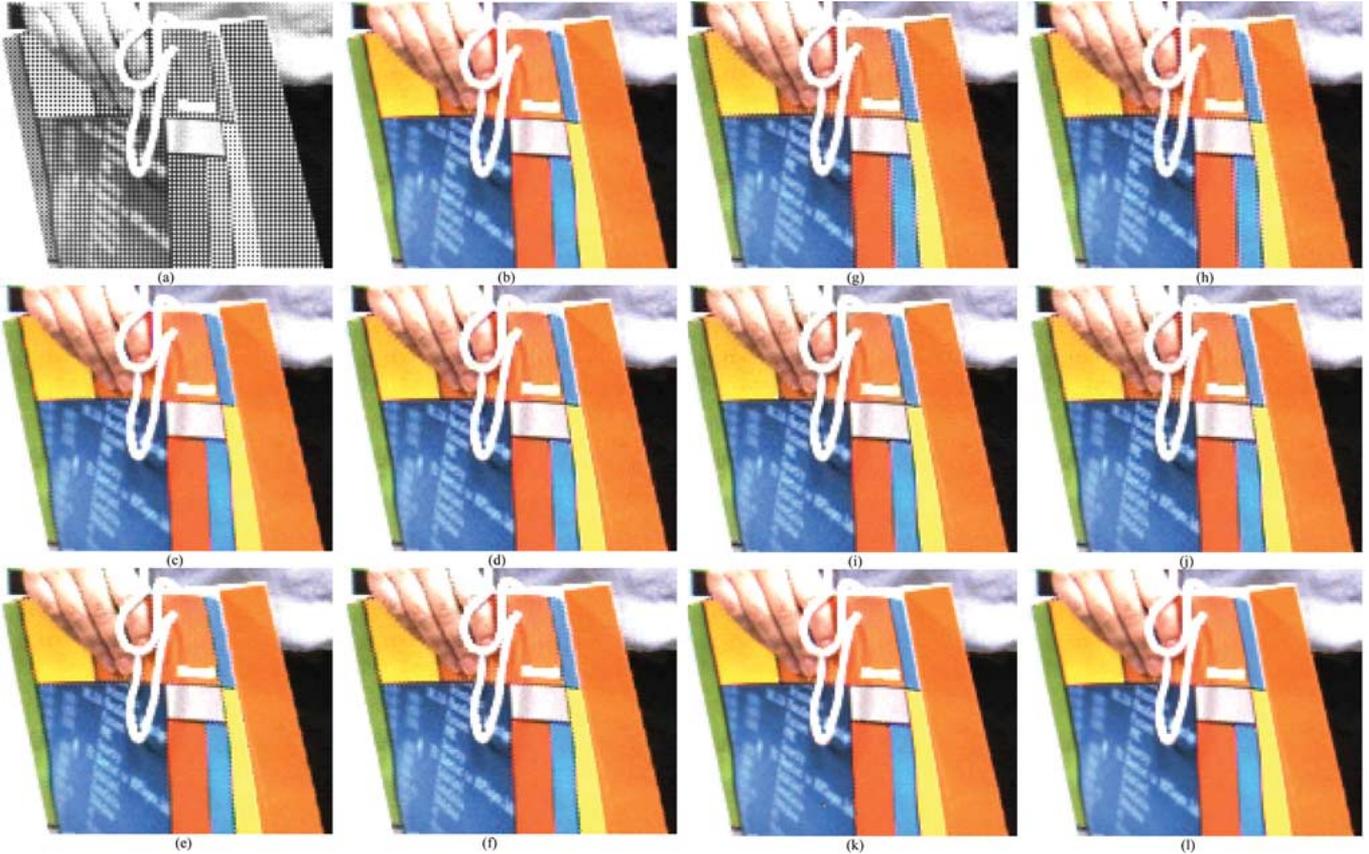


Fig. 11. (a) Original mosaic image. Demosaicked images by the methods in (b) [9], (c) [5], (d) [11], (e) [18], and (f) [19]. Demosaicked images by the methods in (g) [17], (h) [20], (i) [10], (j) [28], (k) [26], and (l) the proposed method.

chrominance. Fig. 10(k) is the result by our earlier temporal demosaicking method in [26]. This method gives better visual quality than the spatial demosaicking methods, but it has spot artifacts due to a lack of data fusion of current and reference samples. Fig. 10(l) is the demosaicked image by the proposed temporal demosaicking method. Clearly, it is better than all other images in terms of visual quality. Most of the color artifacts are eliminated, and many sharp edge structures that are badly distorted in intra-frame demosaicking are well reconstructed by the temporal demosaicking procedure. The peak signal-to-noise ratio (PSNR) results of the three color channels by these demosaicking methods are listed in Table I. The proposed method outperforms the existing methods by 2 ~ 4 dB, depending on the color bands.

In the second clip, the camera is still, but the man is moving with a colorful bag in hand. Fig. 11(a) shows a 120×150 portion of the mosaic image, where sharp edges exist. Fig. 11(b)–(j)

are the results by the spatial demosaicking methods. We can see many artifacts associated with sharp edges. Fig. 11(k) is the result of the method in [26], and Fig. 11(l) is the demosaicked image by the proposed temporal demosaicking method. From our observation, the proposed method has the best visual quality among the competing methods.

VII. CONCLUSION

We have proposed a joint temporal-spatial color demosaicking approach that uses sample correlations in spatial, spectral, and temporal domains to recover the missing color samples of raw mosaic CCD data. A spatially interpolated sample of the current frame is matched to an original pixel in a reference frame via motion analysis, and these two estimates of the missing color component are fused to achieve a more robust estimate. The experimental results showed that the proposed

approach outperformed the existing spatial demosaicking methods by an appreciable margin in both PSNR and visual quality.

ACKNOWLEDGMENT

The authors thank IMAX Corporation, Mississauga, Canada, and Microsoft Research Asia, Beijing, China, for providing them the test sequences. They are also indebted to Dr. Lukac, Dr. Kakarala, and Dr. Hirakawa for sharing with them their demosaicking programs.

REFERENCES

- [1] J. E. Adams, "Intersections between color plane interpolation and other image processing functions in electronic photography," *Proc. SPIE*, vol. 2416, pp. 144–151, Mar. 1995.
- [2] —, "Design of practical color filter array interpolation algorithms for digital cameras," *Proc. SPIE*, vol. 3028, pp. 117–125, Apr. 1997.
- [3] B. E. Bayer and Eastman Kodak Company, "Color Imaging Array," U.S. Patent 3 971 065, Jul. 20, 1976.
- [4] R. Bedford and G. Wyszecki, "Wavelength discrimination for point sources," *J. Opt. Soc. Amer.*, vol. 48, p. 129-ff, 1958.
- [5] E. Chang, S. Cheung, and D. Y. Pan, "Color filter array recovery using a threshold-based variable number of gradients," *Proc. SPIE*, vol. 3650, pp. 36–43, Mar. 1999.
- [6] D. R. Cok and Eastman Kodak Company, "Signal Processing Method and Apparatus for Producing Interpolated Chrominance Values in A Sampled Color Image Signal," U.S. Patent 4 642 678, Feb. 10, 1987.
- [7] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 997–1013, Sep. 2002.
- [8] B. K. Gunturk *et al.*, "Demosaicking: color filter array interpolation," *IEEE Signal Process. Mag.*, vol. 22, no. 1, pp. 44–54, Jan. 2005.
- [9] J. F. Hamilton, Jr. and J. E. Adams, "Adaptive color plane interpolation in single sensor color electronic camera," U.S. Patent 5 629 734, May 13, 1997.
- [10] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicking algorithm," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 360–369, Mar. 2005.
- [11] R. Kakarala and Z. Baharav, "Adaptive demosaicking with the principal vector method," *IEEE Trans. Consum. Electron.*, vol. 48, no. 4, pp. 932–937, Nov. 2002.
- [12] N. Kehtarnavaz, H.-J. Oh, and Y. Yoo, "Color filter array interpolation using color correlation and directional derivatives," *J. Electron. Imag.*, vol. 12, pp. 621–632, Oct. 2003.
- [13] R. Kimmel, "Demosaicking: Image reconstruction from CCD samples," *IEEE Trans. Image Process.*, vol. 8, pp. 1221–1228, Sep. 1999.
- [14] P. Kuhn, *Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation*. Norwell, MA: Kluwer, 1999.
- [15] P. Longère, X. Zhang, P. B. Delahunt, and D. H. Brainard, "Perceptual assessment of demosaicking algorithm performance," *Proc. IEEE*, vol. 90, no. 1, pp. 123–132, Jan. 2002.
- [16] W. Lu and Y.-P. Tan, "Color filter array demosaicking: New method and performance measures," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1194–1210, Oct. 2003.
- [17] R. Lukac, K. Martin, and K. N. Plataniotis, "Demosaicked image post-processing using local color ratios," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 6, pp. 914–920, Jun. 2004.
- [18] R. Lukac and K. N. Plataniotis, "Normalized color-ratio modeling for CFA interpolation," *IEEE Trans. Consum. Electron.*, vol. 50, no. 2, pp. 737–745, May 2004.
- [19] R. Lukac, K. N. Plataniotis, D. Hatzinakos, and M. Aleksic, "A novel cost effective demosaicking approach," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 256–261, Feb. 2004.
- [20] S. C. Pei and I. K. Tam, "Effective color interpolation in CCD color filter arrays using signal correlation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 6, pp. 503–513, Jun. 2003.
- [21] R. Ramanath and W. E. Snyder, "Adaptive demosaicking," *J. Electron. Imag.*, vol. 12, no. 4, pp. 633–642, 2003.
- [22] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, Jun. 1996.
- [23] C. Stiller and J. Konrad, "Estimating motion in image sequence," *IEEE Signal Process. Mag.*, no. 7, pp. 70–91, Jul. 1999.
- [24] H. J. Trussel and R. E. Hartwing, "Mathematics for demosaicking," *IEEE Trans. Image Process.*, vol. 11, pp. 485–492, Apr. 2002.
- [25] X. Wu, W. K. Choi, and P. Bao, "Color restoration from digital camera data by pattern matching," *Proc. SPIE*, vol. 3018, pp. 12–17, Apr. 1997.
- [26] X. Wu and N. Zhang, "Joint temporal and spatial color demosaicking," *Proc. SPIE*, vol. 5017, pp. 307–313, May 2003.
- [27] —, "Primary-consistent soft-decision color demosaicking for digital cameras," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1263–1274, Sep. 2004.
- [28] L. Zhang and X. Wu, "Color demosaicking via directional linear minimum mean square-error estimation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2167–2178, Dec. 2005.
- [29] N. Zhang and X. Wu, "Lossless compression of color mosaic images," in *Proc. Int. Conf. Image Process.*, Singapore, Nov. 2004, pp. 517–520.



Xiaolin Wu (M'89–SM'96) received the B.Sc. degree from Wuhan University, Wuhan, China, in 1982, and the Ph.D. degree from the University of Calgary, Calgary, AB, Canada, in 1988.

He is currently a Professor with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, and a Research Professor of Computer Science with Polytechnic University, Brooklyn, NY, and he holds the NSERC-DALSA research chair in Digital Cinema. His research interests include multimedia coding and communications, image processing, signal quantization and compression, and joint source-channel coding. He has published over 100 research papers, and holds two patents in these fields.

Dr. Wu's awards include the 2003 Nokia Visiting Fellowship, the 2000 Monsteds Fellowship, and the 1998 UWO Distinguished Research Professorship.



Lei Zhang (M'05) was born in 1974 in China. He received the B.S. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, China, and the M.S. and Ph.D. degrees in electrical and computer engineering from Northwestern Polytechnical University, Xi'an, China, in 1998 and 2001, respectively. From 2001 to 2002, he was a Research Associate in the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. Currently, he is a Postdoctoral Fellow in the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. His research interests include digital signal and image processing, wavelet transform, pattern recognition, biometrics, and optimal estimation theory.