

Improvement of Color Video Demosaicking in Temporal Domain

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

Abstract—Color demosaicking is critical to the image quality of digital still and video cameras that use a single-sensor array. Limited by the mosaic sampling pattern of the color filter array (CFA), color artifacts may occur in a demosaicked image in areas of high-frequency and/or sharp color transition structures. However, a color digital video camera captures a sequence of mosaic images and the temporal dimension of the color signals provides a rich source of information about the scene via camera and object motions. This paper proposes an inter-frame demosaicking approach to take advantage of all three forms of pixel correlations: spatial, spectral, and temporal. By motion estimation and statistical data fusion between adjacent mosaic frames, the new approach can remove much of the color artifacts that survive intra-frame demosaicking and also improve tone reproduction accuracy. Empirical results show that the proposed inter-frame demosaicking approach consistently outperforms its intra-frame counterparts both in peak signal-to-noise measure and subjective visual quality.

Index Terms—Bayer color filter array, data fusion, digital video, subpixel motion estimation, temporal color demosaicking.

I. INTRODUCTION

MOST digital cameras capture a color image with a single sensor array that sub-samples color bands in a particular mosaic pattern, such as the Bayer color filter array (CFA) [3] shown in Fig. 1. At each pixel, only one of the three primary colors (red, green, and blue) is sampled. The full color image is reconstructed by interpolating the missing color samples. This process is called color demosaicking, which is critical to the quality of reconstructed color images. The problem of color demosaicking has been extensively studied in spatial and frequency domains for still digital cameras [1], [2], [4], [6], [7], [9], [11], [13]–[17], [22], [23], [26]. Early demosaicking techniques mostly work in the spatial domain, such as nearest neighbor replication, bilinear interpolation, and cubic B-spline interpolation [13]. They are easy to implement but susceptible to many artifacts such as blocking, blurring and zipper effect at edges. The problem can be mitigated by an appropriate use

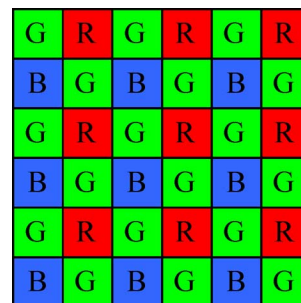


Fig. 1. Bayer pattern.

of strong spectral correlation that exists in most natural color images. Indeed, most modern demosaicking methods exploit the correlation between red, blue, and green channels [2], [4], [6], [7], [9], [11], [13]–[17], [22], [23], [26]. To distinguish the existing methodology from the one of temporal demosaicking, we classify all the above color demosaicking techniques, either purely spatial or spatio-spectral methods, into the class of intra-frame demosaicking.

Even the sophisticated spatio-spectral color demosaicking techniques, such as those recently published, can still fail to faithfully reproduce the color components in the presence of high-frequency and/or sharp color transition structures. In order to surpass the current state-of-the-art in terms of image quality, additional information and constraints of the original color signals have to be brought into the demosaicking process. For digital video cameras, the temporal dimension of a sequence of mosaic frames captures more and new data on the scene, which would be otherwise absent if only a single frame is sampled by the CFA. A natural enquiry of practical significance is how to best exploit the correlation between adjacent frames and improve the accuracy and robustness of intra-frame color demosaicking.

Somewhat surprisingly, there seems to be little research reported on temporal color demosaicking, despite its obvious potential. Recently, Wu *et al.* proposed a joint spatial-temporal color demosaicking technique [24], [25]. Their main idea is to match the CFA green sample blocks in adjacent frames in such a way that the missing red and blue samples in one frame can be inferred from available red and blue samples of a matched adjacent frame. This technique is effective if the motion between the frames is by certain integer offsets that happen to align an available blue/red sample in one frame with a missing blue/red sample in the other frame. Another weakness of the technique is that it selects only one best reference sample from the adjacent frames and then fuses it with a spatially interpolated value.

Manuscript received July 15, 2004; revised February 3, 2006. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada through an industrial research chair in digital cinema. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gabriel Marcu.

X. Wu is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1 Canada (e-mail: xwu@mail.ece.mcmaster.ca).

L. Zhang is with the Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: cslzhang@comp.polyu.edu.hk).

Color versions of Figs. 1, 3, 4, 6, 8, 9, 12, and 13 are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2006.877504

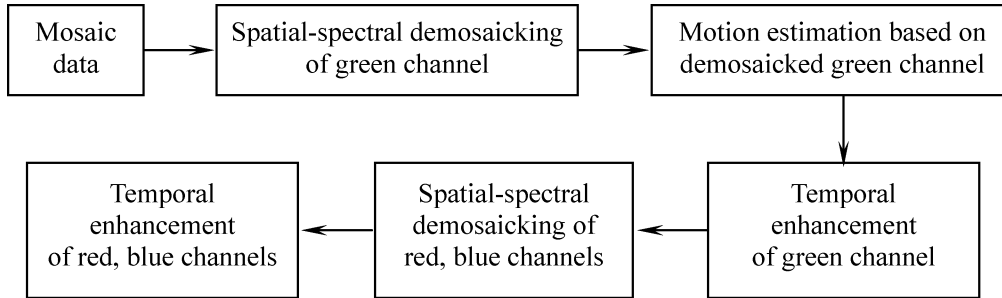


Fig. 2. Flowchart of the proposed temporal demosaicking scheme.

In this paper, we propose a new framework of temporal color demosaicking to overcome these limitations. The progress is made in making multiple estimates of a missing color sample temporally and spatially from several reference frames. These estimates are first derived by associating samples of different frames via subpixel motion vectors, and then optimally fused into a single estimate of the missing color value.

For clarity of presentation and without the loss of generality, our discussions and algorithm development are around the ubiquitous Bayer CFA (see Fig. 1), which is used in most digital video cameras. The temporal demosaicking technique to be developed can be readily generalized to other CFA patterns. Note that the sampling frequency of the green channel is twice that of the red or blue channel in the Bayer pattern. This is because the sensitivity of human visual system peaks at the green wavelength and the green channel contributes the most to the luminance of an image. Since the red and blue channels are more sparsely sampled, we naturally anchor their reconstruction on the green channel in temporal demosaicking.

Fig. 2 is a schematic description of the proposed spatial-temporal demosaicking framework. First, the green channels of all frames are demosaicked individually by intra-frame demosaicking. Because the sampling frequency of green channel is twice that of red or blue channel in the Bayer pattern, the motion estimation between adjacent frames for temporal color demosaicking is based on the reconstructed green channel sequence. This design is to feed the motion analysis with the best information. With the estimated motion vectors, adjacent frames are registered spatially. The reference green samples in adjacent frames are then fused with the intra-frame estimates of the missing green samples of the current frame to enhance the green channel. The temporally enhanced green channel is then used to reconstruct the red and blue channels, by interpolating the missing red and blue samples using both the intra-frame and inter-frame information. First, individual red and blue frames are interpolated spatially using the correlation with corresponding green frames that have by now been demosaicked temporally. Then, the resulting spatio-spectrally demosaicked red and blue frames are enhanced temporally, guided by the motion vectors, by being fused with adjacent red and blue frames.

This paper is structured as follows. Section II introduces a new gradient-based intra-frame demosaicking method for the

green channel by optimally weighting the horizontal and vertical interpolation results. The resulting green frames are used to compute the relative motions of adjacent frames in subpixel precision, which is the subject of Section III. After the frame registration, in Section IV, the reference frames are fused optimally with the current frame to obtain more robust estimates of the missing color samples. Section V presents the experimental results and Section VI concludes.

II. GRADIENT-BASED INTRA-FRAME DEMOSAICKING OF THE GREEN CHANNEL

Since human visual systems are sensitive to the edge structures in an image, all adaptive demosaicking methods strive to avoid interpolating across edges. To this end, the gradient is estimated at each pixel, and the color interpolation is carried out directionally based on the estimated gradient. Directional filtering is the most popular approach to color demosaicking. A well-known directional interpolation scheme is the second order Laplacian correction proposed by Hamilton and Adams [7]. They used the second-order gradients of blue and red samples and the first-order gradient of green samples to interpolate the green channel. The red and blue samples are interpolated similarly with the correction of the second order gradients of the green samples. In this section, we propose a new intra-frame demosaicking method. The goal is to provide a good base for the next step of temporal demosaicking at a reasonable computational cost.

For ease of presentation and without loss of generality, we examine the case depicted by Fig. 3: a column and a row of alternating green and red samples intersect at a red sampling position where the missing green value needs to be estimated. The symmetric case of estimating the missing green values at the blue sampling positions of the Bayer pattern can be handled in the same way. Denote the red sample at the center of the window by R_0 . Its interlaced red and green neighbors in horizontal direction are labeled as R_i^h , $i \in \{-2, 2\}$, and G_i^h , $i \in \{-1, 1\}$, respectively; similarly, the red and green neighbors of R_0 in vertical direction are R_j^v , $j \in \{-2, 2\}$, and G_j^v , $j \in \{-1, 1\}$.

Most intra-frame demosaicking methods are based on an assumption that the difference between the green channel and the red/blue channel is a low-pass signal. Let $\Delta_0 = G_0 - R_0$ be the unknown difference between green and red channels at the

TABLE I
MEANS OF v_0^h AND v_0^v AND THEIR CORRELATION COEFFICIENTS FOR THE 16 TEST IMAGES IN FIG. 4

<i>Image</i>	1	2	3	4	5	6	7	8
m_h	0.0027	0.0013	0.0019	0.0004	-0.0002	-0.0005	0.0001	0.0018
m_v	-0.0007	-0.0009	-0.0014	0.0002	-0.0013	-0.0032	-0.0005	0.0030
c	0.0326	0.0927	-0.0539	0.0696	-0.0042	0.0281	0.0852	0.1592
<i>Image</i>	9	10	11	12	13	14	15	16
m_h	-0.0027	-0.0006	-0.0000	0.0000	0.0008	0.0002	0.0001	0.0003
m_v	0.0013	-0.0006	0.0009	0.0028	0.0003	-0.0004	-0.0015	0.0006
c	-0.0011	-0.0021	0.0176	0.0579	0.0597	0.1320	0.0575	0.0724

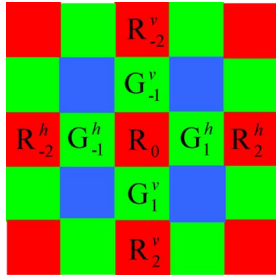


Fig. 3. Row and column of mosaic data that intersect at a red sampling position.

sample position of R_0 . The idea is to obtain an estimate of Δ_0 , denoted by $\hat{\Delta}_0$, and then recover the missing green sample by

$$G_0 \approx R_0 + \hat{\Delta}_0. \quad (2-1)$$

The reason for estimating the color difference signal $\Delta = G - R$ rather than the green signal G directly is that Δ is much smoother than G . Referring to Fig. 3, the horizontal and vertical differences between the green and red channels at R_0 can be estimated as

$$\Delta_0^h = \frac{1}{2} (G_{-1}^h + G_1^h) - \frac{1}{4} (2 \cdot R_0 + R_{-2}^h + R_2^h) \quad (2-2)$$

$$\Delta_0^v = \frac{1}{2} (G_{-1}^v + G_1^v) - \frac{1}{4} (2 \cdot R_0 + R_{-2}^v + R_2^v). \quad (2-3)$$

In [7], the authors set $\hat{\Delta}_0 = \Delta_0^h$ or $\hat{\Delta}_0 = \Delta_0^v$ depending on which of the horizontal and vertical gradients is smaller, but this binary decision discards a potentially useful estimate. Instead, we can fuse the two estimates to obtain a more robust estimate of Δ_0

$$\hat{\Delta}_0 = w_h \Delta_0^h + w_v \Delta_0^v \quad (2-4)$$

where $w_h + w_v = 1$.

Next, we discuss the determination of the weights w_h and w_v . Consider Δ_0^h and Δ_0^v as two independent measurements of the true color difference signal Δ_0

$$\Delta_0^h = \Delta_0 + v_0^h \text{ and } \Delta_0^v = \Delta_0 + v_0^v \quad (2-5)$$

where measurement noises v_0^h and v_0^v are the estimation errors of Δ_0^h and Δ_0^v . Denote by m_h and m_v the means of v_0^h and v_0^v and by c the correlation coefficient between v_0^h and v_0^v . Table I lists the values of m_h , m_v and c for the 16 test images in Fig. 4, indicating that v_0^h and v_0^v are zero mean and nearly uncorrelated. These properties allow us to derive the optimal weights

$$w_h = \frac{\sigma_v^2}{\sigma_h^2 + \sigma_v^2}, \quad w_v = \frac{\sigma_h^2}{\sigma_h^2 + \sigma_v^2} \quad (2-6)$$

where $\sigma_h^2 = Var(v_0^h)$ and $\sigma_v^2 = Var(v_0^v)$. Here, the optimality is in the sense of mean square error (MSE), i.e., w_h and w_v of (2-6) minimize the MSE of estimate $\hat{\Delta}_0$.

Empirically, we identify two main influence factors on the estimation errors of Δ_0^h and Δ_0^v . The first one is the amplitude of Δ_0 . Most natural scenes consist of predominantly pastoral (unsaturated) colors such that the color difference signal $\Delta = G - R$ (or $\Delta = G - B$) is not only smooth but also small in amplitude. The large amplitude of Δ_0 is typically associated with the discontinuity of the color difference signal at the position of R_0 , increasing the risk of large estimation errors. In other words, the amplitudes of Δ_0^h and/or Δ_0^v are proportional to the measurement noises v_0^h and v_0^v . The second factor affecting v_0^h and v_0^v is the presence of the high-frequency component in the luminance signal. To account for this, we measure the gradient of the red channel at R_0 by

$$d_0^h = R_0 - \frac{R_{-2}^h + R_2^h}{2} \text{ and } d_0^v = R_0 - \frac{R_{-2}^v + R_2^v}{2}. \quad (2-7)$$

Also, we observe that the level of v_0^h (v_0^v) is nearly linearly proportional to the magnitudes of Δ_0^h (Δ_0^v) and d_0^h (d_0^v). Let $\Lambda_h = |\Delta_0^h| + |d_0^h|$ and $\Lambda_v = |\Delta_0^v| + |d_0^v|$. Fig. 5 plots the curve of σ_h (σ_v) versus Λ_h (Λ_v) for the first image in Fig. 4. The curves for other images are similar. These curves suggest that σ_h (σ_v) is approximately proportional to Λ_h (Λ_v)

$$\sigma_h \approx \lambda_h \cdot \Lambda_h, \quad \sigma_v \approx \lambda_v \cdot \Lambda_v \quad (2-8)$$



Fig. 4. Test images.

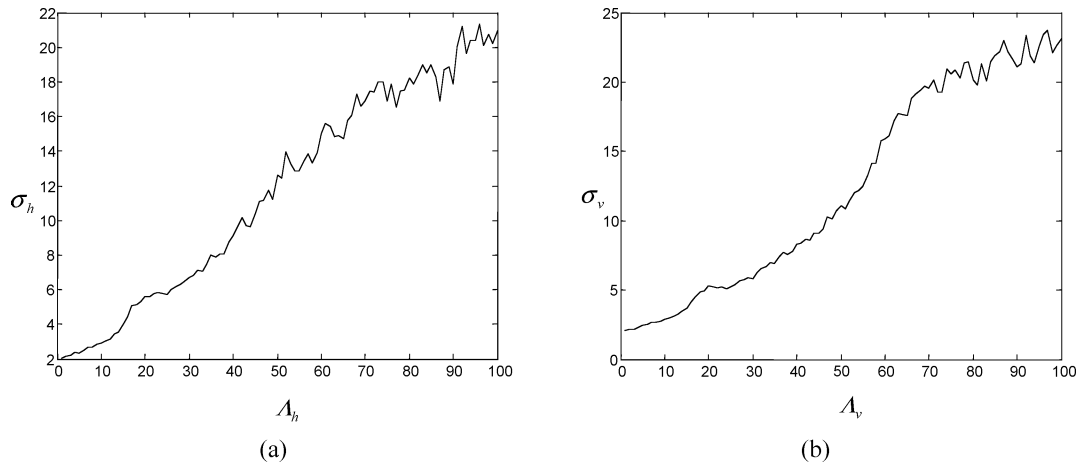


Fig. 5. Curves of σ_h versus λ_h and σ_v versus λ_v .

where λ_h and λ_v are constants, and $\lambda_h \approx \lambda_v$ (see Fig. 5). Substituting (2-8) into (2-6) yields

$$w_h = \frac{\Delta_v^2}{\Delta_h^2 + \Delta_v^2}, \quad w_v = \frac{\Delta_h^2}{\Delta_h^2 + \Delta_v^2}. \quad (2-9)$$

Obviously, if Δ_0^h and d_0^h have large magnitude, then w_h is small, reducing the influence of Δ_0^h on $\hat{\Delta}_0$, and vice versa. By fusing the two directional estimates Δ_0^h and Δ_0^v with optimal

weights w_h and w_v , the final estimate $\hat{\Delta}_0$ is more robust in reconstructing the missing green sample as $G_0 = R_0 + \hat{\Delta}_0$.

III. MOTION ESTIMATION AND RE-SAMPLING

Upon spatially demosaicking the green channel of each frame, we take the next step of temporal demosaicking to enhance the green channel by exploiting the temporal correlation of the video signal. The promise of temporal color demosaicking lies in the fact that the color samples missed by

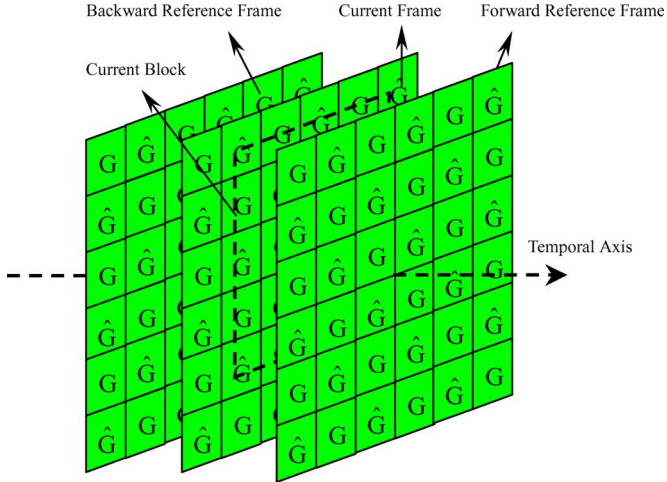


Fig. 6. Current green frame and its backward and forward neighboring frames.

the CFA subsampling process in one frame may be captured in neighboring frames. To realize this potential, we have to register the frames by determining the relative motion vectors between the current frame and the reference frames. Accurate motion estimation of the video frames is pivotal to temporal color demosaicking and many applications such as video coding, superresolution imaging, and computer vision [5], [8], [12], [18]–[21].

In the Bayer CFA, the green channel has twice as many samples as the red and blue channels. Furthermore, the green signal is a good approximation of the luminance signal. For these reasons, we estimate the motions in the green channel. This is also why an initial intra-frame demosaicking process is required to estimate the missing green samples prior to temporal demosaicking.

Any of the existing motion-estimation techniques can be used to estimate the motion vector in the green channel. A more accurate motion estimation method may lead to a better temporal demosaicking result, but of course at a higher computational cost. It should be stressed, however, that the temporal enhancement technique to be developed in the next section is independent of the motion estimation method. The main focus of this paper is on the methodology of temporal enhancement rather than motion estimation. For a good balance between estimation accuracy and low complexity, we choose the block-based motion-estimation technique, which is widely used in MPEG 2/4 and other video-coding standards [12]. Specifically, we adopt the cross-correlation-based method proposed in [27] to compute the motion vector in subpixel precision. For the sake of continuity, we briefly describe the method below.

As the convention of this paper, we denote the original green samples by G and the interpolated green samples through the intra-frame demosaicking by \hat{G} (referring to Fig. 6). Let M be a block of pixels in the current frame and $M_{i,j}$ a matched block in a reference frame with displacement (i, j) , where i and j are integers. Denote by (τ_x, τ_y) the real valued motion vector of M from the current frame to the reference frame.

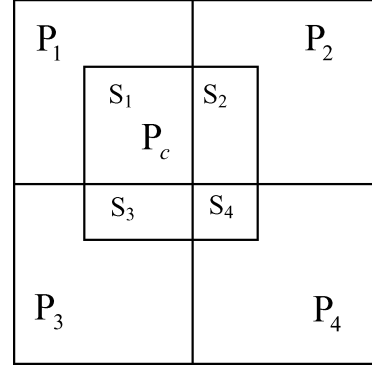


Fig. 7. Re-sampling of the reference block.

For convenience, let M and $M_{i,j}$ be the matrices whose elements are the sample values of blocks M and $M_{i,j}$. The cross-correlation function between M and $M_{i,j}$ is computed as

$$r(i, j) = \sum_{n=1}^{N_x} \sum_{m=1}^{N_y} M(n, m) M_{i,j}(n, m) \quad (3-1)$$

where the dimensions of M and $M_{i,j}$ are $N_x \times N_y$. If the motion vector (τ_x, τ_y) is restricted to integer precision, then it can be approximated by

$$(\tau_x, \tau_y) \approx (i_0, j_0) = \arg \max_{i, j \in Z} r(i, j). \quad (3-2)$$

However, since the true displacement of a scene in two frames is in general subpixel, we should allow the motion vector (τ_x, τ_y) to be real valued for higher precision.

Consider the motion in the horizontal and vertical directions separately. For a fixed y , we write the continuous cross-correlation function $r(x, y)$ as $r_x(x)$. In [27], we showed that the one-dimensional cross correlation function can be well approximated by a Gaussian function in the interval around its peak, for digital signals that are captured by sensors whose kernel function is Gaussian or box. Therefore, we write $r(x, y)$ in the horizontal direction as

$$r_x(x) = a \cdot e^{-b(x-c)^2} \quad (3-3)$$

and use the three most significant samples in horizontal direction $r(i_0, j_0)$, $r(i_0, j_0 - 1)$, and $r(i_0, j_0 + 1)$ to fit $r_x(x)$ and solve for the three parameters a , b , and c . Letting $r(i_0, j_0) = r_x(0)$, $r(i_0, j_0 - 1) = r_x(-1)$, and $r(i_0, j_0 + 1) = r_x(1)$, we have

$$\begin{cases} \ln r(i_0, j_0 - 1) = \ln a - b(1 + c)^2 \\ \ln r(i_0, j_0) = \ln a - bc^2 \\ \ln r(i_0, j_0 + 1) = \ln a - b(1 - c)^2. \end{cases} \quad (3-4)$$

Solving (3-4) yields

$$c = \frac{r(i_0, j_0 + 1) - r(i_0, j_0 - 1)}{4r(i_0, j_0) - 2r(i_0, j_0 - 1) - r(i_0, j_0 + 1)}. \quad (3-5)$$

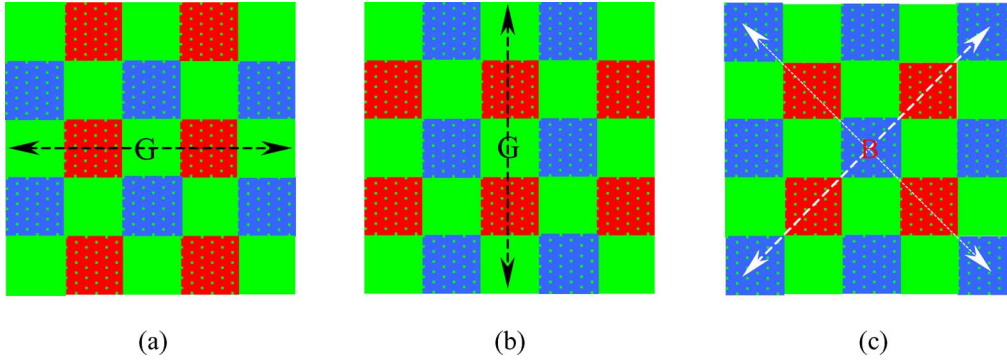


Fig. 8. (a) Interpolation of a missing red sample at a green pixel whose horizontal neighbors are red pixels. (b) Interpolation of a missing red sample at a green pixel whose vertical neighbors are red pixels. (c) Interpolation of a missing red sample at a blue pixel.

The motion value τ_x in the horizontal direction is determined to be the peak position of $r_x(x)$, i.e., τ_x is the sum of integral part j_0 and fractional part c solved in (3-5)

$$\tau_x = j_0 + c. \quad (3-6)$$

Similarly, we can compute τ_y , the motion in vertical direction, and then determine the motion vector (τ_x, τ_y) .

Since (τ_x, τ_y) is real valued in subpixel precision, the corresponding reference block matched to \mathbf{M} , denoted by \mathbf{M}_c , should be re-sampled from the reference frame. In the literature the value of a pixel is commonly modeled as the integral of the light over a unit square. Let P_c be a pixel in \mathbf{M}_c and suppose the pixel square of P_c overlaps with those of P_1, P_2, P_3 , and P_4 , as shown in Fig. 7, which are the pixels in the reference frame. P_c is to be reproduced from P_1, P_2, P_3 , and P_4 . Denote the areas of the overlaps as S_1, S_2, S_3 , and S_4 , which can be computed from the fractional part of the real valued coordinate (τ_x, τ_y) . Then the value of pixel P_c can be calculated as the sum of the intensities over S_1, S_2, S_3 , and S_4 : $P_c = \sum_{i=1}^4 S_i \cdot P_i$.

Due to the structure of the sampling grid of the green channel, two of the four squares P_1, P_2, P_3 , and P_4 are the original green samples G and the other two are the interpolated green samples \hat{G} (see Fig. 6). To factor in higher confidence on G than on \hat{G} , we put different confidence factors c_i on P_1, P_2, P_3 , and P_4 when computing P_c

$$P_c = \sum_{i=1}^4 c_i \cdot S_i \cdot P_i \quad (3-7)$$

where weight $c_i = 1.2$ if P_i is an original green sample and $c_i = 0.8$ if P_i is an interpolated green sample. The sum of weights should be $\sum_{i=1}^4 c_i = 4$.

IV. TEMPORAL IMPROVEMENT OF DEMOSAICKING

Intra-frame spatial demosaicking has an inherent limitation. It is impossible to reconstruct faithfully the color signal at/near the edge/texture structures whose frequency exceeds the Nyquist sampling limit. To aggravate the problem, the human visual system is very sensitive to the color artifacts across edges. The demosaicking errors become visually highly objectionable if the

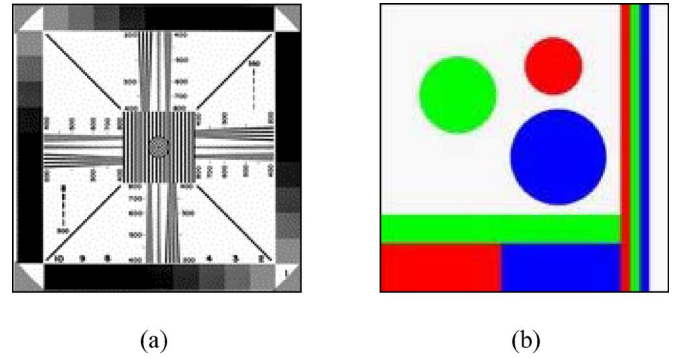


Fig. 9. (a) Resolution chart image. (b) Synthetic image that contains pure red, green, and blue color objects.

discontinuity exists simultaneously in luminance and chrominance. In this case, the artifacts cannot be removed by assuming a high correlation between the color channels as most intra-frame demosaicking algorithms do. In contrast, temporal correlation of a mosaic color video signal, which commonly exists, provides badly needed information to resolve such difficult cases for color demosaicking.

A. Temporal Update of Green Channel

With the motion estimation and re-sampling algorithms described in Section III, we can get a reference block of the current block \mathbf{M}_0 in each reference frame. Suppose that K reference frames are used, and denote by $\{\mathbf{M}_i\}_{i=1,2,\dots,K}$ the re-sampled reference blocks. The spatially demosaicked sample in \mathbf{M}_0 is to be fused with the matched samples in \mathbf{M}_i . For convenience of expression, we denote the spatially interpolated green sample in \mathbf{M}_0 by \hat{G}_0 , the unknown true green sample corresponding to \hat{G}_0 by G , and the associated reference samples in \mathbf{M}_i by \hat{G}_i . Naturally, we can write \hat{G}_0 and \hat{G}_i as the measurements of true sample G

$$\hat{G}_i = G + e_i, \quad i = 0, 1, \dots, K \quad (4-1)$$

where e_i are the interpolation errors of \hat{G}_i in the spatial demosaicking process, and they are uncorrelated with G . Since \hat{G}_i are independent observations of G in different frames, we assume that errors $e_i, i = 0, 1, \dots, K$, are mutually nearly uncorrelated.

Indeed, the correlation coefficient between e_i and e_j , $i \neq j$, is only 0.11 on average for the test sequences used in Section V. Furthermore, this correlation is much weaker in areas of edges and textures, where temporal demosaicking offers clear advantage over spatial demosaicking, than smooth areas. If e_i 's are significantly correlated, then the observations \hat{G}_i are very similar (e.g., if there is no acquisition noise and no motion, \hat{G}_i will be identical to each other). In this case, the reference frames offer very little new information and temporal demosaicking cannot improve spatial demosaicking.

In order to fuse all the measurements \hat{G}_i into a more robust estimate of G , we consider the weighted estimate

$$\bar{G} = \sum_{i=0}^K w_i \hat{G}_i \quad (4-2)$$

where weights $\sum_{i=0}^K w_i = 1$. The criterion of determining w_i is to minimize the MSE of \bar{G} , i.e., $\{w_i\} = \arg \max_{\sum_i w_i=1} E[(\bar{G} - G)^2]$, where E is the expectation operator.

The weights w_i may be determined off-line using an appropriate training set. The weights optimized for the training set can then be used in (4-2) to obtain the fused estimate \bar{G} . However, if the training dataset is not available, or/and if the best color demosaicking performance is desired, on-line adaptive estimation can be made as described below. Let

$$\Omega = E[(\bar{G} - G)^2] = E \left[\left(\sum_{i=1}^K w_i e_i + \left(1 - \sum_{i=1}^K w_i \right) e_0 \right)^2 \right]. \quad (4-3)$$

Denote by σ_i^2 the variance of error e_i . Differentiating Ω with respect to e_i , $i = 1, 2, \dots, K$, and setting the partial derivatives to zero. With $E[e_i e_j]_{i \neq j} \approx 0$, we have

$$\frac{\partial \Omega}{\partial w_i} = w_i \sigma_i^2 - \left(1 - \sum_{j=1}^K w_j \right) \sigma_0^2 = 0$$

from which we obtain the optimal weight vector $\mathbf{w} = \text{col}\{w_1, w_2, \dots, w_K\}$ for the estimates in the K reference frames

$$\mathbf{w} = \mathbf{S}^{-1} \mathbf{1} \quad (4-4)$$

where $\mathbf{1}$ is a column vector whose elements are all ones and the $K \times K$ matrix \mathbf{S} is

$$\mathbf{S} = \begin{bmatrix} 1 + \sigma_1^2/\sigma_0^2 & 1 & \cdots & 1 \\ 1 & 1 + \sigma_2^2/\sigma_0^2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 + \sigma_K^2/\sigma_0^2 \end{bmatrix}. \quad (4-5)$$

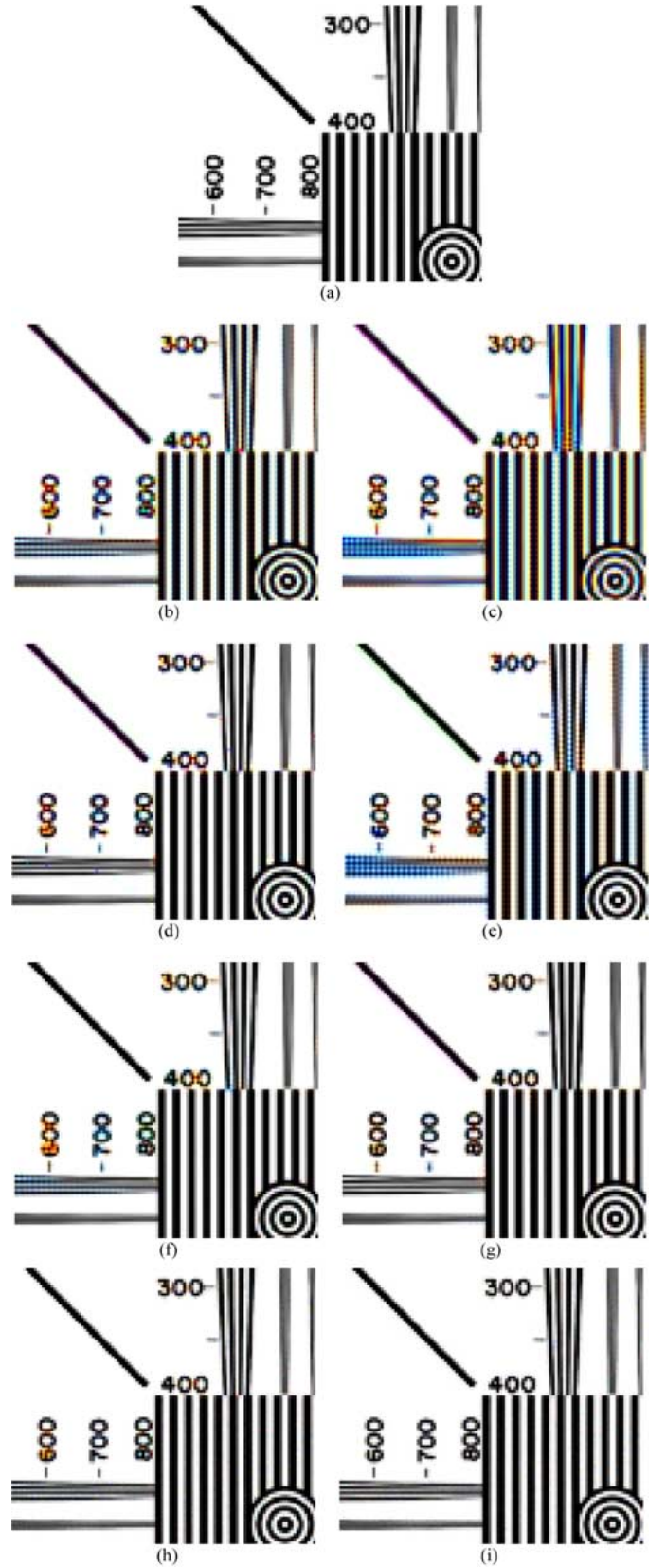


Fig. 10. (a) Original image; demosaicked images by the methods in (b) [7]; (c) [4]; (d) [9]; (e) [16]; (f) [15]; (g) [26]; (h) [24]; and (i) the proposed temporal scheme.

Solving (4-4) for w_i , the fused estimate of G is then computed by (4-2).

TABLE II
PSNR (dB) RESULTS ON THE TWO SYNTHETIC SEQUENCES BY DIFFERENT METHODS

<i>Demosaicking Methods</i>	Simulated sequence 1			Simulated sequence 2		
	red	green	blue	Red	green	blue
Method in [7]	23.98	30.54	23.95	23.99	36.83	23.43
Method in [4]	17.95	27.95	17.99	8.06	8.00	7.90
Method in [9]	29.82	31.78	29.94	22.66	33.91	22.40
Method in [16]	24.71	28.69	24.76	22.43	28.58	23.99
Method in [15]	30.17	33.49	30.21	24.26	37.00	23.63
Method in [26]	34.84	36.41	34.53	22.55	35.45	21.96
Method in [24]	32.85	37.58	33.13	32.87	43.58	31.78
Proposed temporal method	35.76	40.38	35.77	33.69	46.20	32.60

B. Estimation of the Error Variances

To implement the above algorithm, the error variances σ_i^2 need to be estimated. From $\hat{G}_i = G + e_i$ and $E[e_i e_j]_{i \neq j} \approx 0$, we have

$$d_{i,j} = E \left[(\hat{G}_i - \hat{G}_j)^2 \right] = \sigma_i^2 + \sigma_j^2, \quad i, j = 0, 1, \dots, K \text{ and } i \neq j. \quad (4-6)$$

The values of $d_{i,j}$ can be estimated adaptively from blocks \mathbf{M}_i and \mathbf{M}_j , $i \neq j$

$$d_{i,j} = \frac{1}{L} \sum_{\hat{G}_i \in \mathbf{M}_i, \hat{G}_j \in \mathbf{M}_j} (\hat{G}_i - \hat{G}_j)^2 \quad (4-7)$$

where L is the total number of missing green samples in blocks \mathbf{M}_i .

If σ_0^2 , i.e., the variance of e_0 in the current block \mathbf{M}_0 , is a known prior, then the values of σ_i^2 for other i 's can be calculated by (4-6) and (4-7). Otherwise, all the values of σ_i^2 can be estimated as follows. Let

$$\boldsymbol{\sigma} = \text{col} \{ \sigma_0^2, \dots, \sigma_K^2 \} \quad (4-8)$$

be a $K + 1$ -dimensional vector that encompass all the σ_i^2 , and let

$$\mathbf{d} = \text{col} \{ d_{0,1}, \dots, d_{0,K}, d_{1,2}, \dots, d_{1,K}, \dots, d_{K-1,K} \} \quad (4-9)$$

be a $K(K + 1)/2$ -dimensional vector that encompass all the $d_{i,j}$, then there exists a $K(K + 1)/2 \times (K + 1)$ matrix \mathbf{H} such that

$$\mathbf{d} = \mathbf{H}\boldsymbol{\sigma}. \quad (4-10)$$

Denote $\mathbf{h}_{i,j}$ as the row in \mathbf{H} such that $d_{i,j} = \mathbf{h}_{i,j}\boldsymbol{\sigma}$. Clearly, only the i^{th} and j^{th} elements in $\mathbf{h}_{i,j}$ are 1 and all other elements are zeros. We estimate $\boldsymbol{\sigma}$ by the least-square estimation technique

$$\boldsymbol{\sigma} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{d}. \quad (4-11)$$

C. Joint Spatial-Temporal Interpolation of Red/Blue Channels

After the green estimates are improved by the temporal demosaicking process described in Sections IV-A and IV-B, they can in turn guide the demosaicking of the red and blue channels. Similarly to the demosaicking of the green channel, the missing red and blue samples are recovered in two steps. First, we spatially interpolate them with the help of the temporally demosaicked green channel and then temporally improve the interpolation results aided by motion vectors.

The intra-frame demosaicking of red/blue channel can be accomplished by any of the existing methods. In this paper, we adopt the directional filtering strategy similar to Hamilton and Adams' method [7]. Since the interpolation of blue channel is symmetrical to that of red channel, we only describe the process of interpolating the red channel.

Referring to Fig. 8, there are three cases depending on the positions of missing red samples. Fig. 8(a) and (b) shows the two cases of the missing red samples at the original green pixel positions. Fig. 8(c) shows the case of a missing red sample at the original blue pixel position. We stress the fact that the missing green samples at the red/blue positions have already been estimated. In the case of Fig. 8(a), we can estimate the green-red difference signal by using the true red samples and temporally estimated green samples in horizontal direction, and in the case of Fig. 8(b), we can estimate the green-red difference signal in vertical direction similarly. In the case of Fig. 8(c), we can estimate two green-red difference values in 45° and 135° directions. These two values are fused to one result as what we did in Section II. The missing red sample is estimated by subtracting the estimated green-red difference $\hat{\Delta}$ from the original green value, $\hat{R} = G - \hat{\Delta}$, in cases (a) and (b), or from the estimated green value, $\hat{R} = \hat{G} - \hat{\Delta}$, in case (c). Since the above spatial

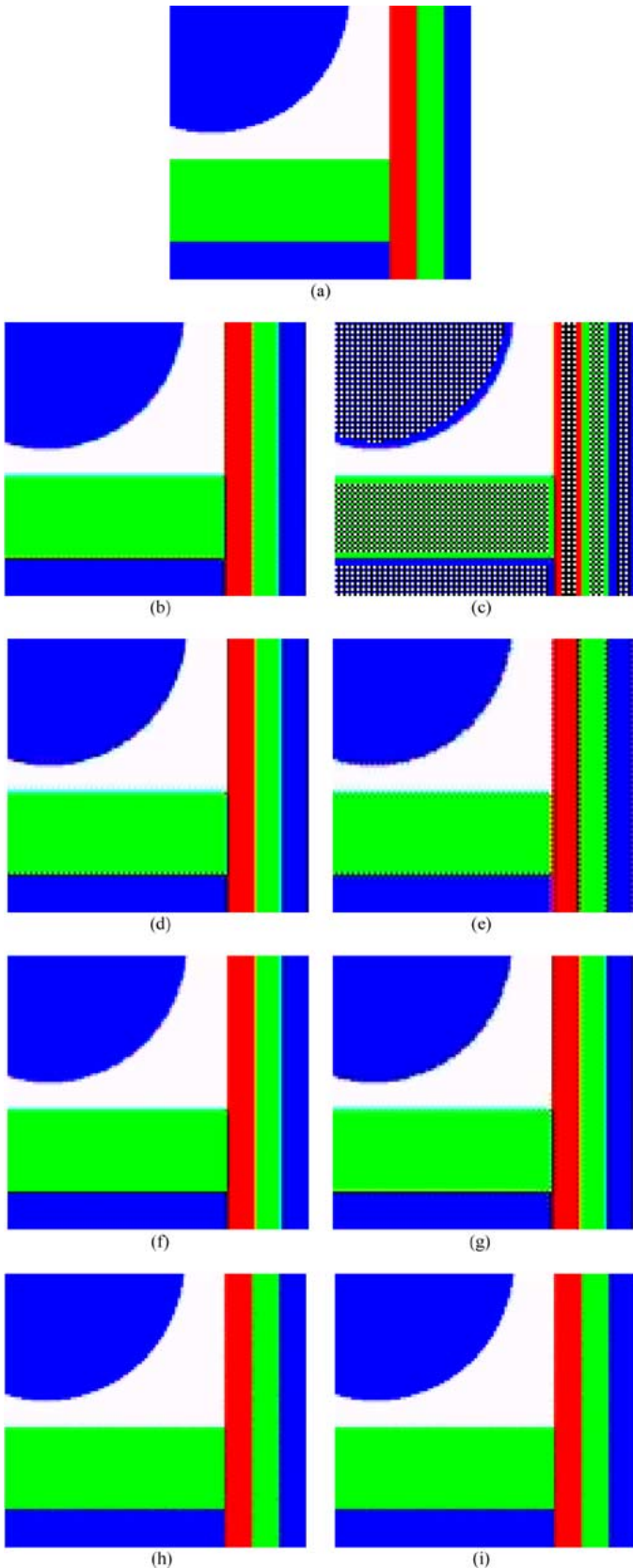


Fig. 11. (a) Original image; demosaicked images by the methods in (b) [7]; (c) [4]; (d) [9]; (e) [16]; (f) [15]; (g) [26]; (h) [24]; and (i) the proposed temporal scheme.

demosaicking process exploits the spectral correlation between red/blue and green, and it operates on temporally demosaicked

green channel, the spatial interpolation of red and blue channels indirectly benefits from the temporal redundancy in adjacent frames.

Similarly to the green channel, the spatially interpolated red and blue channels can be further improved via motion estimation and data fusion. However, the motion vectors are still computed in the green channel, because the motion estimation accuracy in the green channel is much higher than in the red and blue channels. Indeed, the sample frequency of red/blue channel is only half of that of the green channel, and the two-dimensional (2-D) sampling grid of red/blue channel employs inefficient square lattice as opposed to the diamond lattice for the green channel.

The temporal enhancement process of red/blue channel is similar to that of green channel. The main difference is in the confidence factor determination in the re-sampling step. Take the red channel for example, after the motion vector (τ_x, τ_y) between a current block M and a reference block M_c is computed, a pixel P_c in M_c needs to be re-sampled from the four neighboring pixels $P_1, P_2, P_3,$ and P_4 in the reference frame. In the sampling grid of the red channel, only one of the four pixels is an original red sample R and the other three are interpolated ones \hat{R} . The confidence factors in the re-sampling process $P_c = \sum_{i=1}^4 c_i \cdot S_i \cdot P_i$ are $c_i = 1.6$ if P_i is an original red sample and $c_i = 0.8$ if P_i is an interpolated red sample. This is to guarantee that the sum of all c_i to be 4.

V. EXPERIMENTAL RESULTS

The proposed joint spatial-temporal color demosaicking algorithm was implemented and tested on two synthetic sequences and two real video clips. To evaluate the performance of the proposed algorithm in comparison with other intra-frame demosaicking algorithms and our earlier temporal demosaicking algorithm [24], [25], we present in this section both the demosaicked images and PSNR results. Besides the temporal method in [24], the six state-of-the-art intra-frame demosaicking algorithms used in our comparison are: the method of second-order Laplacian filtering by Hamilton and Adams [7], the method of variable number of gradients by Chang *et al.* [4], the principal vector method by Kakarala and Baharav [9], the bilinear interpolation of color difference by Pei and Tam [16], the normalized color-ratio modeling by Lukac and Plataniotis [15], and the directional filtering and fusion method by Zhang and Wu [26].

A. Test on Synthetic Mosaic Video Sequences

1) *Resolution Chart Video*: To validate the proposed demosaicking algorithm, we used two synthetic mosaic test sequences. The first sequence is generated from a gray-scale resolution chart [Fig. 9(a)], which is widely used in evaluating superresolution algorithms. Denote by I the original resolution chart image. We simulated five mosaic video frames with relative motions as follows.

To simulate a video camera, I is first smoothed by convoluting with a low-pass filter $f_s : I_s = I * f_s$, where $f_s = [1, 1, 1; 1, 8, 1; 1, 1, 1]/16$ in this experiment. Then five



Fig. 12. Scene in the (a) first test clip and (b) second test clip.

TABLE III
PSNR (dB) RESULTS ON THE TWO SYNTHETIC SEQUENCES BY DIFFERENT METHODS

<i>Demosaicking Methods</i>	Test sequence 1			Test sequence 2		
	red	green	blue	Red	green	blue
Method in [7]	31.53	33.61	27.87	33.16	38.30	36.16
Method in [4]	31.46	33.41	27.52	32.61	37.90	33.79
Method in [9]	30.66	34.08	29.24	31.76	38.42	35.72
Method in [16]	30.83	34.37	29.14	31.59	36.29	35.15
Method in [15]	30.11	33.09	28.99	31.68	36.71	35.93
Method in [26]	30.48	35.25	29.74	31.23	38.24	35.62
Method in [24]	32.24	35.15	30.10	34.15	38.11	36.06
Proposed temporal method	34.42	36.10	30.57	36.62	38.87	37.85

frames are created by down-sampling I_s . The current frame to be enhanced is

$$F_0(n, m, \bullet) = I_s(2n, 2m, \bullet)$$

where “ \bullet ” is the index for color channels. The other four reference frames are

$$\begin{aligned} F_1(n, m, \cdot) &= I_s(2n, 2m + 2, \bullet) + v_1 \\ F_2(n, m, \bullet) &= I_s(2n + 2, 2m, \bullet) + v_2 \\ F_3(n, m, \bullet) &= I_s(2n + 2, 2m + 2, \bullet) + v_3 \\ F_4(n, m, \bullet) &= I_s(2n - 1, 2m + 1, \bullet) + v_4 \end{aligned}$$

where v_1, v_2, v_3 , and v_4 are Gaussian white noises and they are mutually uncorrelated. In our simulation, $v_i \sim N(0, 8)$, $i = 1, 2, 3, 4$. The relative motions between F_i , $i = 1, 2, 3, 4$ and F_0 are $(0, 1)$, $(1, 0)$, $(1, 1)$, and $(-0.5, 0.5)$, respectively. Finally, each frame F_i is down-sampled again according to the Bayer pattern.

The demosaicked results of different methods for the resolution chart video are presented in Fig. 10, where (a) is the original, (b)–(g) show the demosaicked images by the intra-frame methods in [4], [7], [9], [15], [16], and [26]. These methods produce highly visible color artifacts. Fig. 10(h) and (i) shows the output images of a previous temporal demosaicking method [24] and the proposed method. One can see that only the new method

is free of color artifacts on the resolution test sequence. The PSNR results are listed in Table II, which are in agreement with the visual quality evaluations. The proposed joint spatial-temporal method has the highest PSNR.

2) *Saturated Color Video*: The simulation results on the resolution chart demonstrate the effect of the joint spatial-temporal demosaicking approach on the high-frequency luminance contents. Now we turn to the other more difficult form of discontinuity for color demosaicking: abrupt chrominance changes. Fig. 9(b) represents such an artificial worst-case scenario: a scene consisting of sharp objects of highly saturated colors (pure red, green, and blue) in a white background. We generated a mosaic video sequence from Fig. 9(b) in the same way as described in the previous subsection. The demosaicked images of this test sequence are presented in Fig. 11. All the intra-frame methods generate highly visible color artifacts around the object boundaries. In contrast, the temporal demosaicking approach performs very well in reproducing the sharp color edges. The PSNR results are also listed in Table II.

The color artifacts caused by chrominance discontinuities (e.g., lack of spectral correlation in saturated colors) are more difficult to remove than those caused by luminance discontinuities within a frame. In this case, the common assumption made by almost all intra-frame demosaicking algorithms on the smoothness of the color difference signal no longer holds. As we see in this experiment and the following experiments, this problem can be mitigated by exploiting the temporal correlation.

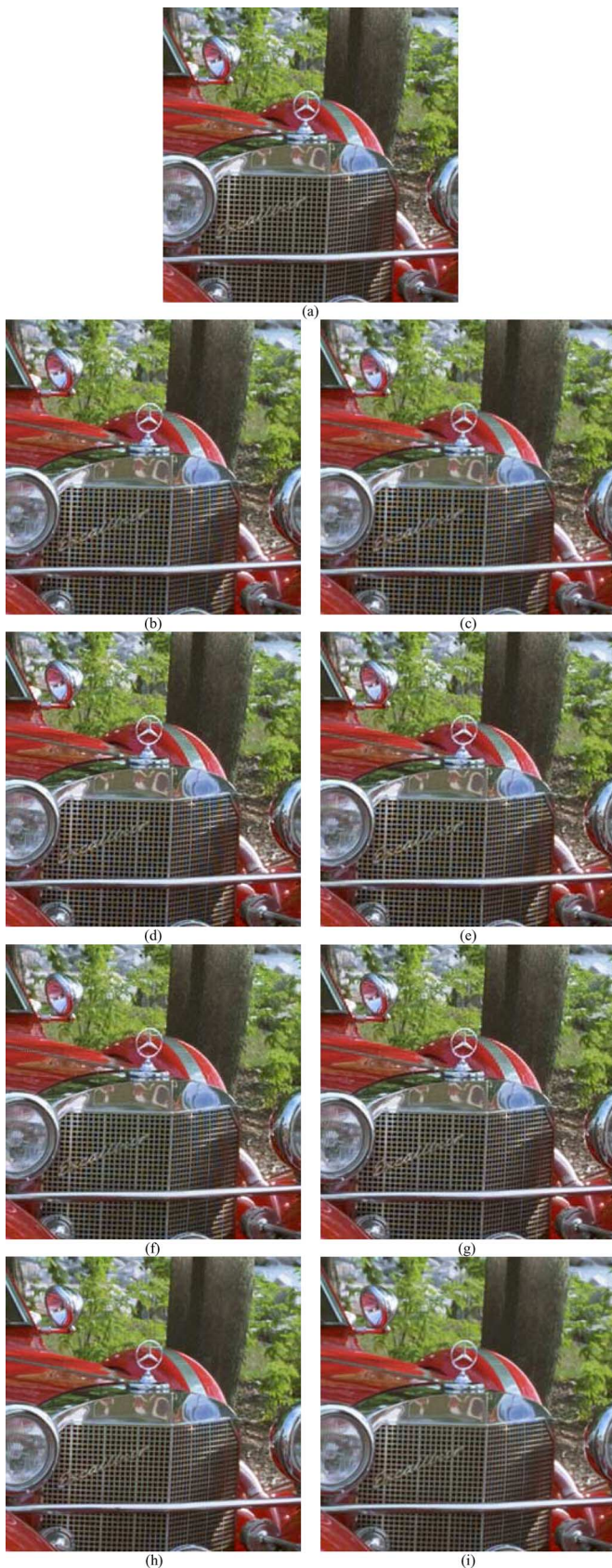


Fig. 13. (a) Original image; demosaicked images by the methods in (b) [7]; (c) [4]; (d) [9]; (e) [16]; (f) [15]; (g) [26]; (h) [24]; and (i) the proposed temporal scheme.

B. Experiments on Real Video Clips

A similar comparison study was also carried out on two real video clips. The movie sequences were originally captured on film and then digitized by a high-resolution scanner.¹ All the three color channels were known and we simulated the mosaic data by subsampling the red, green, blue channels according to the Bayer pattern. In temporal demosaicking of a current frame, we used two immediately preceding frames and two immediately succeeding frames as reference frames.

The scene in the first movie clip is a car in a park. The car is still but the camera is rotating around it. Fig. 12(a) shows the scene in a frame. The video is captured at a rate of 24-frames/second. The spatial resolution of each frame is 1948×1280 and the bit depth is 8 bits per color channel. In this clip, most of the smooth background objects such as the road, the lake, and trees can be reconstructed free of visible artifacts by spatial demosaicking techniques. However, on the car, where some sharp edges accompany abrupt color changes, the spatial demosaicking cannot faithfully recover the missing color components. Fig. 13(a) shows a 256×256 portion of the original frame in question.

Fig. 13(b)–(i) shows the demosaicked images by the eight methods. Fig. 14(a)–(i) presents the close-ups of the demosaicked images by these methods for easier visual inspection. There are highly visible color artifacts in Fig. 13(b)–(g), particularly on the grill of the car, where the true color signal frequency exceeds the sampling frequency of the Bayer CFA. The algorithm in [26] [see Fig. 13(g)] has fewer color artifacts on the grill than other intra-frame demosaicking methods, but it still generates significant zipper effects along the boundary of the red and silver colors and on the emblem of the car [see Fig. 14(g)], as other spatial demosaicking methods [see Fig. 14(b)–(f)]. The color edges that failed all intra-frame demosaicking methods have discontinuities in both luminance and chrominance. For sharp edges of highly saturated colors of weak spectral correlation, no sufficient information exists within a frame to reconstruct the color signal. This is the situation where the temporal correlation can come to the rescue.

Fig. 13(h) and Fig. 14(h) are the results by the temporal demosaicking method in [24]. It offers better visual quality than intra-frame demosaicking, but some color artifacts still exist and the improvement in PSNR is quite small (referring to Table III). Fig. 13(i) and Fig. 14(i) are the demosaicked images by the proposed temporal demosaicking method. The proposed method has clear advantages over all others in terms of visual quality. Most of the color artifacts are eliminated and many sharp edge structures that are missing or distorted in intra-frame demosaicking are well reconstructed by the joint temporal-spatial demosaicking. The PSNR results (we demosaicked eight consecutive frames and computed the average PSNR) of the three color channels by these demosaicking methods are listed in Table III.

¹Our thanks to the IMAX Corporation, Mississauga, ON, Canada, for providing the test data.

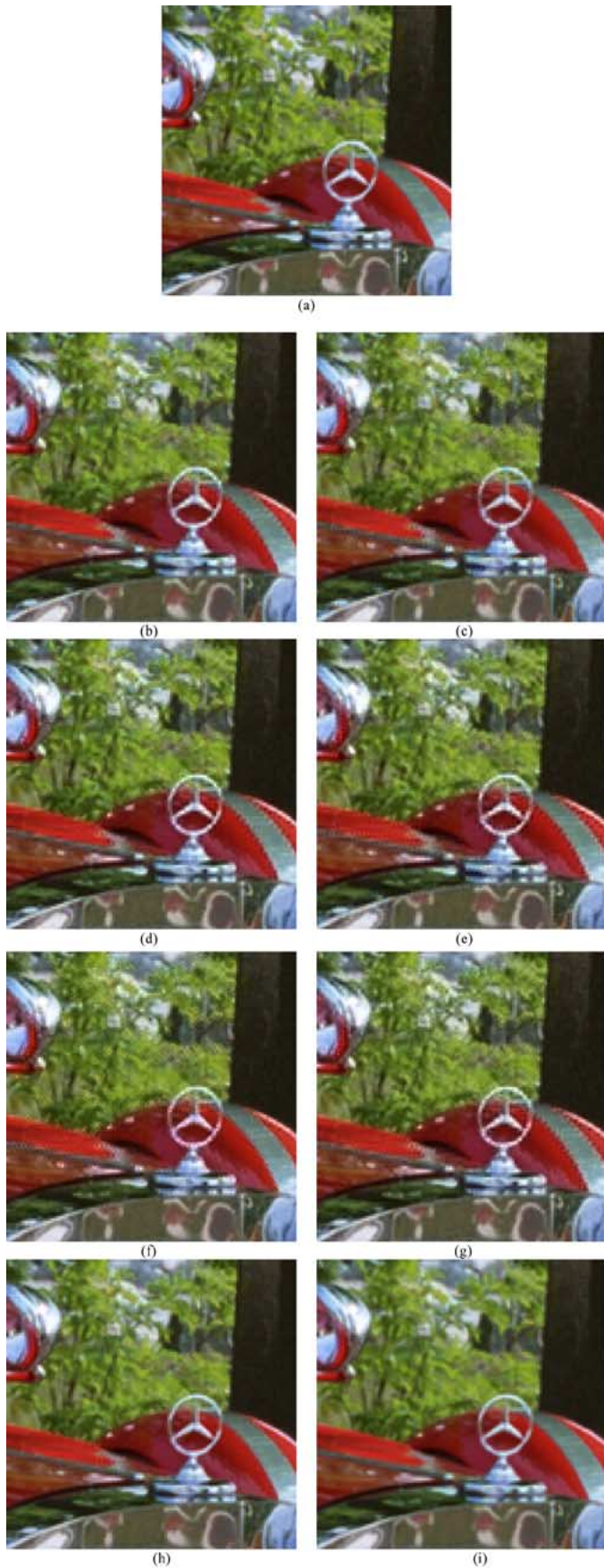


Fig. 14. Zoom-in images of the demosaicked results. (a) Original image; demosaicked images by the methods in (b) [7]; (c) [4]; (d) [9]; (e) [16]; (f) [15]; (g) [26]; (h) [24]; and (i) the proposed temporal scheme.



Fig. 15. (a) Original image; demosaicked images by the methods in (b) [7]; (c) [4]; (d) [9]; (e) [16]; (f) [15]; (g) [26]; (h) [24]; and (i) the proposed temporal scheme.

The proposed method achieves significantly higher PSNR than others as well.

The other test sequence, shown in Fig. 12(b), is a skating girl. In this scene, both the object (the girl) and the camera are moving. The motion is a combination of rotation and translation. The video is captured at 24-frames/second and 8 bits per color channel. The original resolution of the sequence is 1948×1280 and we down-sampled it to 974×640 in the experiment. Fig. 15(a) shows a zoom-in crop (128×128) of the scene. As in the previous example, by comparing the other demosaicking techniques in Fig. 15(b)–(h) with the proposed temporal method in Fig. 15(i), we see that most of the color artifacts and zipper effects are suppressed by the proposed method. The PSNR results are also listed in Table III. The proposed method had limited PSNR improvement on the green channel because the green signal of this sequence is so smooth that the spatial demosaicking by itself suffices for good reconstruction.

Finally, we want to bring the reader's attention to the significant PSNR increases in the reconstructed red and blue channels by the proposed demosaicking method. This means that besides reducing color artifacts the proposed method also reproduces the color tones more precisely than other methods. The big improvements in reproduction precision of the red and blue channels should not come as a surprise, considering that the Bayer CFA has much lower sampling frequency and inferior sampling grid pattern for the red and blue channels. The design bias of the Bayer CFA against the red and blue channels in favor of the green channel makes the faithful reproduction of the red and blue signals more difficult if color demosaicking is carried out on a frame-by-frame basis. Temporal demosaicking is capable of much higher tone accuracy in color reproduction.

VI. CONCLUSION

We proposed an inter-frame color demosaicking approach that utilizes all three forms of pixel correlations: spatial, spectral, and temporal. The green channel is first reconstructed and it acts as an anchor to help recovering the red and blue channels. In reconstructing each one of the three channels, we first interpolate it using an intra-frame demosaicking method and then temporally enhance it with the help of adjacent frames. The experimental results showed that the proposed approach appreciably improved intra-frame spatial color demosaicking techniques, removing much of the color artifacts of the latter and ensuring higher tone accuracy. Temporal color demosaicking requires a fairly large buffer to hold multiple reference frames, and involves quite extensive computations compared with the intra-frame demosaicking. We can reduce the complexity by invoking the proposed temporal color demosaicking algorithm only when intra-frame demosaicking can not produce good outputs. Only at localities of sharp edges and finely structured textures the CPU-intensive temporal color demosaicking will be activated. In smooth regions of an image, which typically constitute the major portion of a scene, the sampling frequency of the color mosaic is high enough to allow correct color demosaicking solely in the spatial domain.

ACKNOWLEDGMENT

The authors would like to thank the IMAX Corporation, Mississauga, ON, Canada, for providing the test data, and Dr. Kakarala and Dr. Lukac for sharing 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, 1995.
- [2] —, "Design of practical color filter array interpolation algorithms for digital cameras," *Proc. SPIE*, vol. 3028, pp. 117–125, 1997.
- [3] B. E. Bayer, Eastman Kodak Company, Color Imaging Array, U.S. patent 3 971 065 1975.
- [4] 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, 1999.
- [5] F. Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: A review and a new contribution," *Proc. IEEE*, vol. 83, no. 6, pp. 858–876, Jun. 1995.
- [6] 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.
- [7] J. F. Hamilton, Jr. and J. E. Adams, Adaptive Color Plane Interpolation in Single Sensor Color Electronic Camera, U.S. Patent 5 629 734 1997.
- [8] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 525–533, Feb. 1993.
- [9] R. Kakarala and Z. Baharav, "Adaptive demosaicing with the principal vector method," *IEEE Trans. Consum. Electron.*, vol. 48, no. 11, pp. 932–937, Nov. 2002.
- [10] S. M. Kay, *Fundamentals of Statistical Signal Processing*, 1st ed. New York: Pearson Education, Mar. 1993, vol. I, Estimation Theory.
- [11] R. Kimmel, "Demosaicing: Image reconstruction from CCD samples," *IEEE Trans. Image Process.*, vol. 8, no. 9, pp. 1221–1228, Sep. 1999.
- [12] P. Kuhn, *Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation*. Boston, MA: Kluwer, 1999.
- [13] P. Longère, X. Zhang, P. B. Delahunt, and D. H. Brainard, "Perceptual assessment of demosaicing algorithm performance," *Proc. IEEE*, vol. 90, no. 1, pp. 123–132, Jan. 2002.
- [14] W. Lu and Y.-P. Tan, "Color filter array demosaicing: New method and performance measures," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1194–1210, Oct. 2003.
- [15] R. Lukac and K. N. Plataniotis, "Normalized color-ratio modelling for CFA interpolation," *IEEE Trans. Consum. Electron.*, vol. 50, no. 5, pp. 737–745, May 2004.
- [16] 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.
- [17] R. Ramanath and W. E. Snyder, "Adaptive demosaicking," *J. Electron. Imag.*, vol. 12, no. 4, pp. 633–642, 2003.
- [18] R. R. Schultz, L. Meng, and R. L. Stevenson, "Subpixel motion estimation for super-resolution image sequence enhancement," *J. Vis. Commun. Image Represent.*, vol. 9, pp. 38–50, Mar. 1998.
- [19] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, pp. 996–1011, Jun. 1996.
- [20] C. Stiller and J. Konrad, "Estimating motion in image sequence," *IEEE Signal Process. Mag.*, no. 7, pp. 70–91, Jul. 1999.
- [21] B. C. Tom and A. K. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 278–287, Feb. 2001.
- [22] H. J. Trussel and R. E. Hartwing, "Mathematics for demosaicking," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 485–492, Apr. 2002.
- [23] X. Wu, W. K. Choi, and P. Bao, "Color restoration from digital camera data by pattern matching," in *Proc. SPIE*, Apr. 1997, vol. 3018, pp. 12–17.
- [24] X. Wu and L. Zhang, "Temporal color video demosaicking via motion estimation and data fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 2, pp. 231–240, Feb. 2006.
- [25] X. Wu and N. Zhang, "Joint temporal and spatial color demosaicking," in *Proc. Electronic Imaging Conf. SPIE*, May 2003, vol. 5017, pp. 307–313.

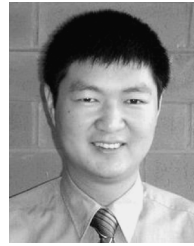
- [26] L. Zhang and X. Wu, "Color demosaicking via directional linear minimum mean square-error estimation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2167–2178, Dec. 2005.
- [27] —, "On cross correlation function based discrete time delay estimation," in *Proc. ICASSP*, Philadelphia, PA, Mar. 2005, vol. 4, pp. 981–984.



Xiaolin Wu (SM'96) received the B.Sc. from Wuhan University, Wuhan, China, in 1982 and the Ph.D. 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, where 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 160 research papers and holds two patents in these fields.

Dr. Wu received the 2003 Nokia Visiting Fellowship, the 2000 Monsteds Fellowship, and the 1998 UWO Distinguished Research Professorship.



Lei Zhang (M'04) received the B.S. from Shenyang Institute of Aeronautical Engineering, Shenyang, China, in 1995 and the M.S. and Ph.D. in electrical 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. From 2003 to 2006, he was a Postdoctoral Fellow in the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. Since 2006,

he has been an Assistant Professor in the Department of Computing, The Hong Kong Polytechnic University. His research interests include image and video processing, biometrics, pattern recognition, multisensor data fusion, and optimal estimation theory.