

Joint Spatial-Temporal Color Demosaicking

Xiaolin Wu* and Lei Zhang

Department of Electrical and Computer Engineering,
McMaster University,
Hamilton, Ontario, Canada L8S 4K1
{xwu, johnray}@mail.ece.mcmaster.ca

Abstract. Demosaicking of the color CCD data is a key to the image quality of digital still and video cameras. Limited by the Nyquist frequency of the color filter array (CFA), color artifacts often accompany high frequency contents in the reconstructed images. This paper presents a general approach of joint spatial-temporal color demosaicking that exploits all three forms of sample correlations: spatial, spectral, and temporal. By motion estimation and statistical data fusion between multiple estimates obtained from adjacent mosaic frames, the new approach can significantly outperform the existing spatial color demosaicking techniques both in objective measure and subjective visual quality.

1 Introduction

Digital photography has become a landmark of our information technology era. Unlike traditional film, the image sensors of most digital cameras capture a color image by sub-sampling 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 estimating the missing color samples based on spatial and spectral correlations. This process is called color demosaicking, which is critical to the quality of reconstructed color images.

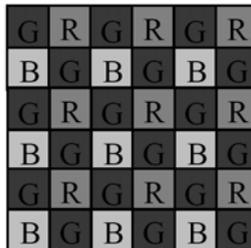


Fig. 1. The Bayer pattern

* This research is supported in part by the Natural Sciences and Engineering Research Council of Canada through an industrial research chair in digital cinema.

In the past decades color demosaicking has been extensively studied, but mostly in the spatial domain for still digital cameras. The earlier spatial demosaicking methods are nearest neighbor replication, bilinear and bicubic interpolations [15]. They are easy to implement but highly susceptible to color artifacts such as blocking, blurring and zipper effect at edges. Later demosaicking methods exploited the spectral correlation between color channels. The smooth hue transition (SHT) methods [6, 1] assume images having slowly varying hue. But SHT methods tend to cause large interpolation errors in the red and blue channels when green values abruptly change. 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 well-known second order Laplacian filter proposed by Hamilton and Adams [9, 2], 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 thresholding. The missing samples are estimated from the samples along the selected gradients. Recently, Zhang and Wu [28] proposed a linear minimum mean square-error estimation (LMMSE) 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 [17], 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 [18]. 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* [8]. 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. More recently reported demosaicking methods of increased sophistication include the method of adaptive homogeneity by Hiraoka and Parks [10], the primary-consistent soft-decision method of Wu and Zhang [27], the principal vector method by Kakarala and Baharav [11], the bilinear interpolation of color difference by Pei and Tam [19]. The above methods can be classified into the class of spatial color demosaicking techniques.

Nevertheless, around the edges where sharp changes in both chrominance and luminance happen, the spatial demosaicking techniques, including those recently-developed sophisticated ones, are error prone due to a lack of correlation in both spectral and spatial domains. 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 the 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 by this paper can be readily generalized to other CFA patterns. In the Bayer pattern the sampling frequency of green channel is twice that of red or blue channel. 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 [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 red/blue channel, it can then be used to guide the interpolation of 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 temporal demosaicking.

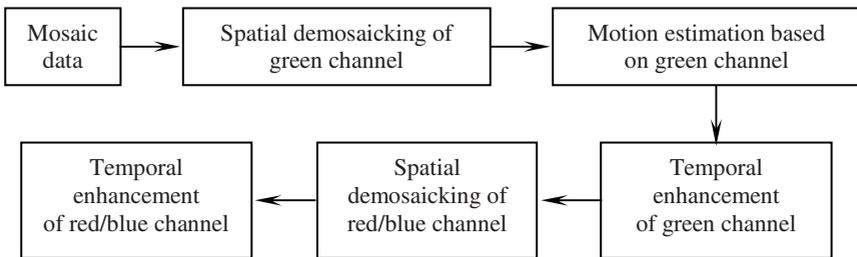


Fig. 2. Flow chart of the proposed temporal demosaicking scheme

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. The motion estimation between adjacent frames for temporal color demosaicking is based on the reconstructed green channel sequence in order to feed the motion analysis with sufficient 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 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 red and blue channels are first recovered spatially with the help of temporally demosaicked green channel, and then guided by the motion vectors, the reference red/blue samples are fused with the spatially estimated samples.

This paper is structured as follows. Section II introduces a new gradient-based spatial demosaicking method for the green channel by optimally weighting the horizontal and vertical interpolation results. These resulting spatially demosaicked green frames are used to compute the relative motions of several 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.

2 Spatial Demosaicking of Green Channel

Most spatial demosaicking methods exploit the correlation between red, blue and green channels [2, 5, 8-11, 13, 16-20, 26-28]. Since human visual systems are sensitive to the edge structures in an image, it is important not to interpolate across edges. At each pixel the gradient is estimated, and the color interpolation is carried out directionally based on the estimated gradient. Directional filtering is the most popular approach to spatial demosaicking. A well-known directional interpolation scheme is the second order Laplacian correction proposed by Hamilton and Adams [9]. 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 spatial demosaicking method. The goal is to provide a good base for the next step of temporal demosaicking at a reasonable computational cost.

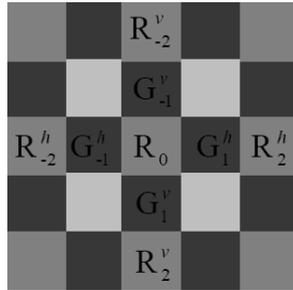


Fig. 3. A row and a column of mosaic data that intersect at a red sampling position

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 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

$$G_0 \approx R_0 + \hat{\Delta}_0 \tag{2-1}$$

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

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

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

One can select between the two estimates Δ_0^h and Δ_0^v , depending on the gradient. But the binary decision may lose the information in the discarded estimate. Instead, we fuse the two estimates to obtain a more robust estimate of Δ_0 :

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

where $w_h + w_v = 1$. Consider Δ_0^h and Δ_0^v as two independent measurements of the true color difference signal Δ_0 :

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

where 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 . We empirically observed that v_0^h and v_0^v are zero mean and nearly uncorrelated. These properties allow us to derive the optimal weights in the minimum mean square error sense:

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

where $\sigma_h^2 = \text{Var}(v_0^h)$ and $\sigma_v^2 = \text{Var}(v_0^v)$.

There are 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 evidently the presence of high frequency component of the luminance signal. To account for the second factor we measure the gradient of the red channel at R_0 by

$$d_0^h = R_0 - \frac{R_{-2}^h + R_2^h}{2} \text{ and } d_0^v = R_0 - \frac{R_{-2}^v + R_2^v}{2} \tag{2-7}$$

Let $A_h = |\Delta_0^h| + |d_0^h|$ and $A_v = |\Delta_0^v| + |d_0^v|$. Experimental data show that σ_h (σ_v) is approximately proportional to A_h (A_v), i.e., $\sigma_h \approx \lambda_h \cdot A_h$, $\sigma_v \approx \lambda_v \cdot A_v$ for some constant λ . Therefore,

$$w_h = \frac{A_v^2}{A_h^2 + A_v^2} \text{ and } w_v = \frac{A_h^2}{A_h^2 + A_v^2} \tag{2-8}$$

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 $\hat{\Delta}_0$; 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 $G_0 = R_0 + \hat{\Delta}_0$.

3 Motion Estimation and Re-Sampling

After spatially demosaicking individual green frames, we can improve the resulting green channel by exploiting the temporal correlation of the video signal. The green samples missed by the CFA subsampling process in one frame may be captured in neighboring frames. To use this information we register the frames by determining the relative motions between the current frame and the reference frames. Accurate motion estimation of the video frames is pivotal to temporal color demosaicking.

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 spatial demosaicking of green channel is performed prior to temporal demosaicking. Any of the existing motion estimation techniques can be used to estimate the motion vector in the green channel [7, 14, 21-24]. 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. 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 [14]. Specifically, we adopt the cross correlation based method proposed in [29] to compute the motion vector in subpixel precision.

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} . Let \mathbf{M} be a block of pixels in the current frame and $\mathbf{M}_{i,j}$ 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 \mathbf{M} from the current frame to the reference frame.

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 , which are the pixels in the reference frame, as shown in Fig. 4. 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

$$P_c = \sum_{i=1}^4 S_i \cdot P_i$$

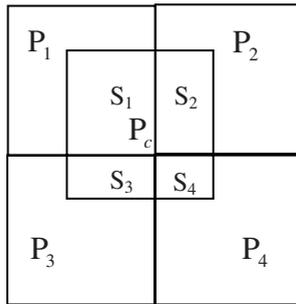


Fig. 4. Re-sampling of the reference sample

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} . 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 :

$$P_c = \sum_{i=1}^4 c_i \cdot S_i \cdot P_i \tag{3-1}$$

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 $\sum_{i=1}^4 c_i = 4$.

4 Temporal Demosaicking

Spatial demosaicking can fail if the discontinuity exists simultaneously in luminance and chrominance. In this case the artifacts cannot be removed by assuming high correlation between the color channels as most spatial demosaicking algorithms do. In contrast, temporal correlation of a mosaic color video signal 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 expression convenience, 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

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

where e_i are the interpolation errors of \hat{G}_i in the spatial demosaicking process, and they are nearly 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, 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 anyways.

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

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

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., $\{w_i\} = \arg \min_{\sum w_i=1} E[(\bar{G} - G)^2]$, 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

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

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

$$\mathbf{w} = \mathbf{S}^{-1} \mathbf{1} \tag{4-4}$$

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

$$\mathbf{S} = \begin{bmatrix} 1 + \sigma_1^2 / \sigma_0^2 & 1 & \dots & 1 \\ 1 & 1 + \sigma_2^2 / \sigma_0^2 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 + \sigma_K^2 / \sigma_0^2 \end{bmatrix} \tag{4-5}$$

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

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

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

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

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

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

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

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

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

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

$$\mathbf{d} = \mathbf{H}\boldsymbol{\sigma} \tag{4-10}$$

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:

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

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 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 spatial demosaicking of red/blue channel can be accomplished by any of the existing spatial methods. In this paper, we adopt the directional filtering strategy similar to Hamilton and Adams' method [9]. 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. 5, there are three cases depending on the positions of missing red samples. Figs. 5 (a) and (b) show the two cases of the missing red samples at the original green pixel positions. Fig. 5 (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. 5 (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. 5 (b), we can estimate the green-red difference signal in vertical direction similarly. In the case of Fig. 5 (c), we can estimate two green-red difference values in 45 degree and 135 degree 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 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.

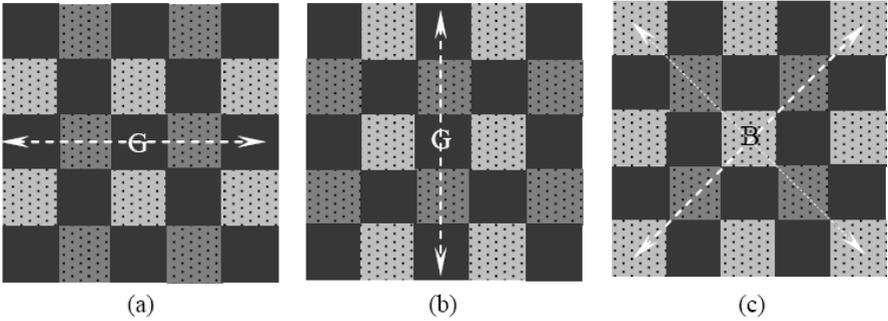


Fig. 5. The dark squares represent original green samples, and mid-gray squares original red samples and light-gray squares original blue samples. The dots on red and blue samples mean that the green samples have been estimated temporally. (a) Interpolation of a missing red sample at a green pixel whose 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

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 2D 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 \mathbf{M} and a reference block \mathbf{M}_c is computed, a pixel P_c in \mathbf{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.

5 Experimental Results

The proposed joint spatial-temporal color demosaicking algorithm was implemented and tested on simulated color mosaic and real mosaic video data. The first test video sequence was originally captured on film and then digitized by high-resolution scanner. 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. The second video sequence was captured directly by a digital video camera. In temporal demosaicking of a current frame, we used two immediately preceding frames and two immediately succeeding frames as reference frames. Figs. 6 (a) and (b) show the scenes of the two video clips.

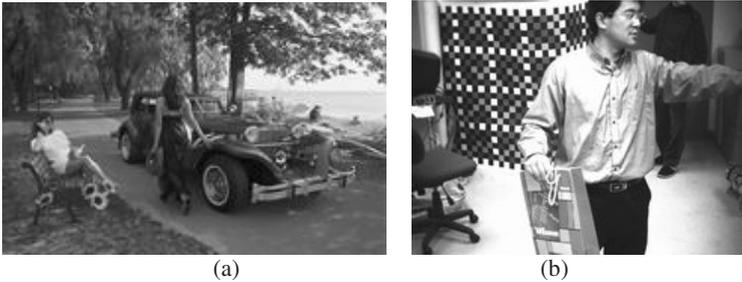


Fig. 6. (a) The scene in the first test clip. (b) The scene in the second test clip

Six recently developed spatial demosaicking algorithms were used in our comparison study: the method of second order Laplacian filtering by Hamilton and Adams [9], the gradients-based method by Chang *et al.* [5], the principal vector method by Kakarala and Baharav [11], the normalized color-ratio modeling by Lukac and Plataniotis [17], the method of adaptive homogeneity by Hirakawa and Parks [10], and the directional filtering and fusion method by Zhang and Wu [28].

The scene in the first movie clip is a car in a park. The car is still but the camera is rotating around it. 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. 7 (a) shows a 190×120 portion of the original frame in question.

Figs. 7 (b) ~ (h) show the demosaicked images by the methods in [9], [5], [11], [17], [10], [28] and the proposed method. Figs. 8 (a) ~ (h) present the close-ups of the demosaicked images by these methods. There are highly visible color artifacts in Figs. 7 (b) ~ (g), particularly on the grill of the car, where the true color signal frequency exceeds the sampling frequency of the Bayer CFA. The recent algorithm in [28] (see Fig. 7 (g)) has fewer color artifacts on the grill than other intra-frame demosaicking methods. But it still generates zipper effects along the boundary of the red and silver

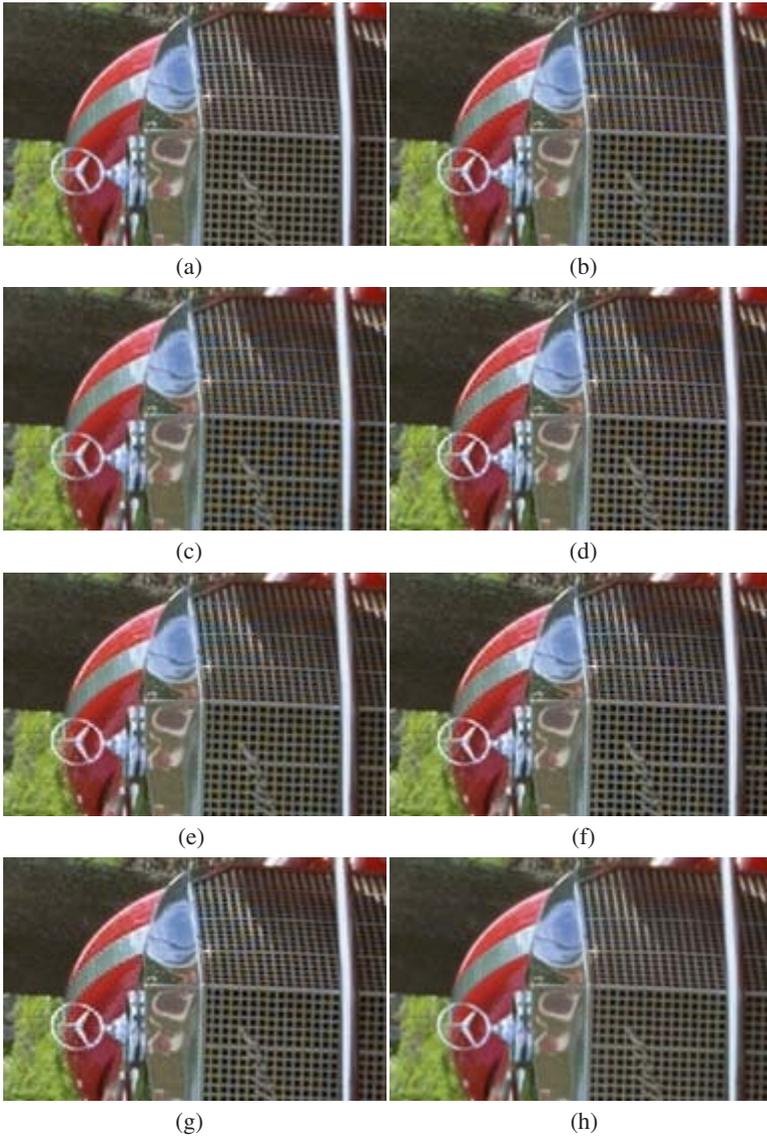


Fig. 7. (a) Original image; demosaicked images by the methods in (b) [9]; (c) [5]; (d) [11]; (e) [17]; (f) [10]; (g) [28]; and (h) the proposed temporal scheme

colors and on the emblem of the car (see Fig. 8 (g)), as other spatial demosaicking methods (see Fig. 8 (b) ~ (f)). The color edges on which all spatial demosaicking algorithms fail have discontinuities in both luminance and chrominance. This invalidates the assumption underlying many of these methods that the color difference signal (chrominance) is smooth. For sharp edges of highly saturated colors where

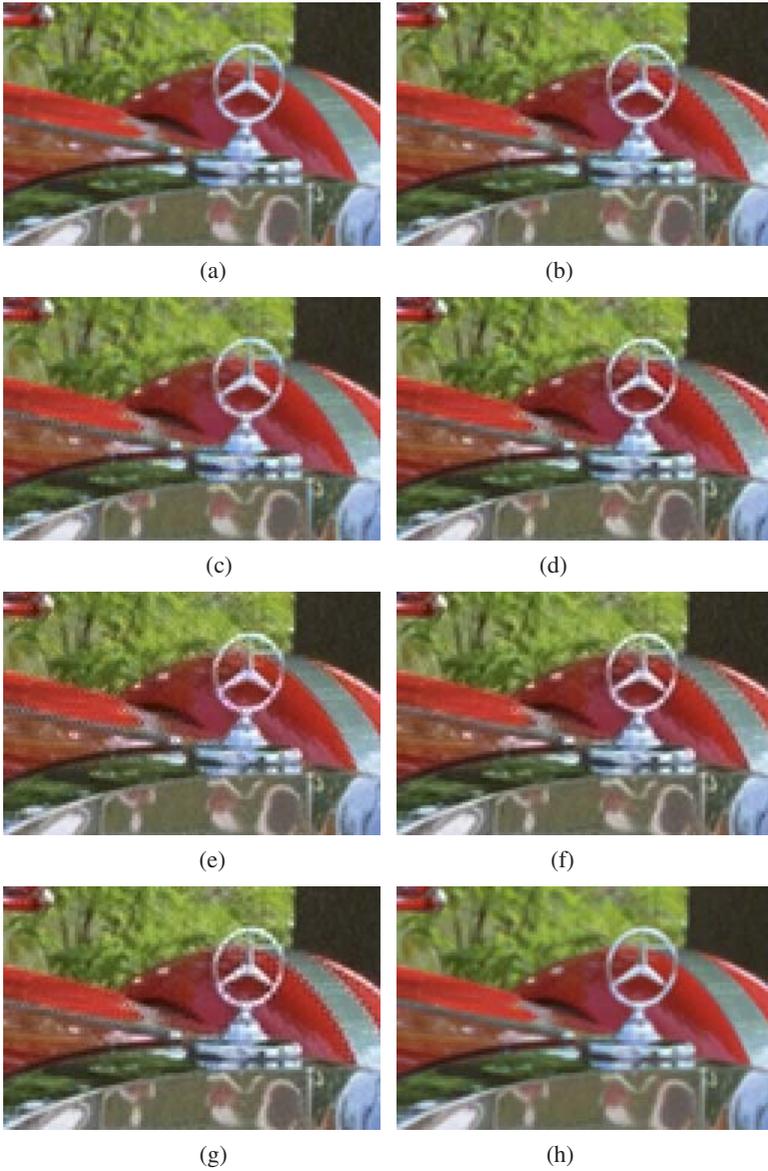


Fig. 8. Zoom-in images of the demosaicked results. (a) Original image; demosaicked images by the methods in (b) [9]; (c) [5]; (d) [11]; (e) [17]; (f) [10]; (g) [28]; and (h) the proposed spatiotemporal method

spectral correlation is weak, spatial demosaicking does not have sufficient information to reconstruct the color signal. This is where the temporal correlation can come to the rescue.

Table 1. The PSNR results of the 7 demosaicking methods for the first video sequence

Methods		Method in [9]	Method in [5]	Method in [11]	Method in [17]	Method in [10]	Method in [28]	Proposed
PSNR(dB)	R	30.58	30.85	29.45	29.75	28.14	29.59	33.86
	G	32.33	32.61	32.92	32.87	32.68	35.32	35.54
	B	26.03	26.25	28.54	28.70	28.66	30.24	30.43

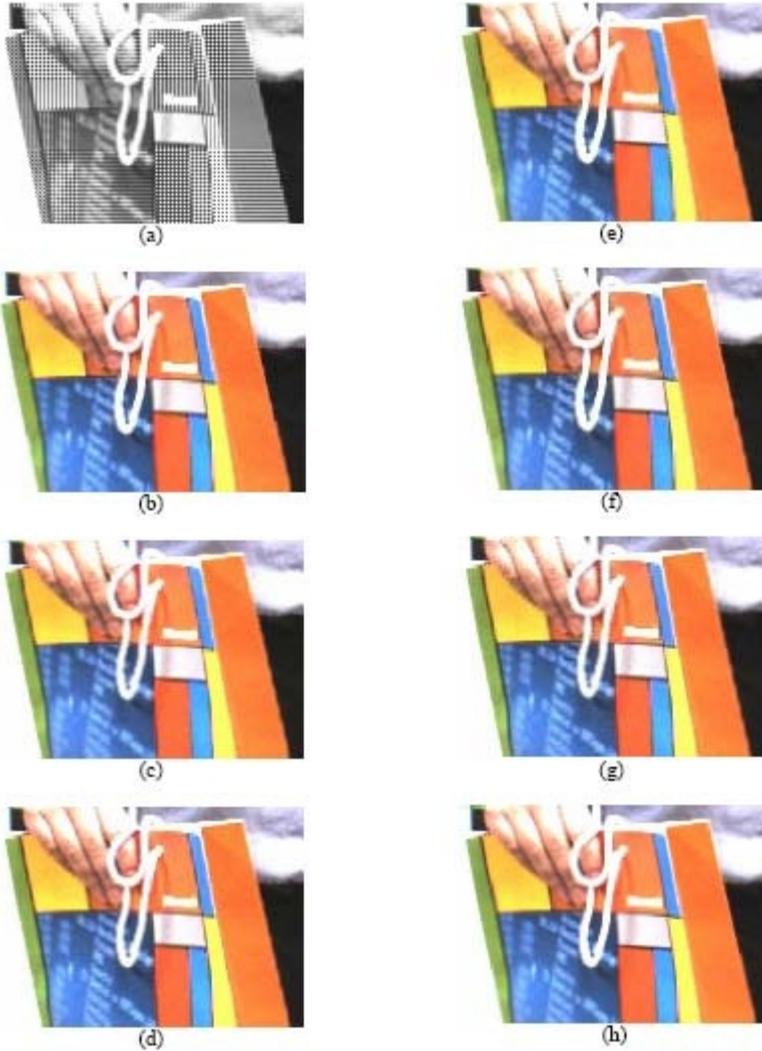


Fig. 9. (a) Original mosaic image; demosaicked images by the methods in (b) [9]; (c) [5]; (d) [11]; (e) [17]; (f) [10]; (g) [28]; and (h) the proposed spatiotemporal method

Fig. 7 (h) and Fig. 8 (h) are the demosaicked images by the proposed spatiotemporal demosaicking method. The proposed method has clear advantages over all others in terms of visual quality. Most of the color artifacts are eliminated. It nicely reconstructs many sharp edge structures that are missing or distorted by spatial demosaicking. The PSNR results of the three color channels by these demosaicking methods are listed in Table 1. The proposed method achieves significantly higher PSNR than others as well.

In the second clip, the camera is still but the man is moving with a colorful bag in hand. It was captured directly by a digital video camera at a rate of 25-frames/second. The spatial resolution is 640×480. Fig. 9 (a) shows a 120×150 portion of the mosaic image, where sharp edges of saturated colors exist. Figs. 9 (b) ~ (g) are the results by the spatial demosaicking methods. We can see many artifacts associated with sharp edges. Fig. 9 (h) is the result by the proposed temporal demosaicking method. The proposed method has the best visual quality among the competing methods.

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. Such significant 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 in the spatial domain only. This problem can be greatly alleviated by temporal demosaicking.

6 Conclusion

We proposed a general spatiotemporal color demosaicking approach that utilizes spatial, spectral, and temporal correlations to recover the missing color samples of raw mosaic data. 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 a spatial demosaicking method and then temporally enhance it with the help of adjacent frames. The experimental results showed that the proposed approach outperforms the existing spatial color demosaicking techniques by a significant margin, and can remove much of the color artifacts of the latter.

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 temporal demosaicking judiciously to regions of high frequency contents, where intra-frame demosaicking is unreliable. In smooth regions of an image, which typically constitute the major portion of a frame, the sampling frequency of color mosaic is high enough to allow correct color demosaicking solely in spatial domain.

Acknowledgement

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

References

- [1] J. E. Adams, "Intersections between color plane interpolation and other image processing functions in electronic photography," *Proceedings of SPIE*, vol. 2416, pp. 144-151, 1995.
- [2] J. E. Adams, "Design of practical color filter array interpolation algorithms for digital cameras," *Proceedings of SPIE*, vol. 3028, pp. 117-125, 1997.
- [3] B. E. Bayer and Eastman Kodak Company, "Color Imaging Array," US patent 3 971 065, 1975.
- [4] R. Bedford and G. Wysecki, "Wavelength discrimination for point sources," *Journal of the Optical Society of America*, vol. 48, pp. 129-ff, 1958.
- [5] E. Chang, S. Cheung and D. Y. Pan, "Color filter array recovery using a threshold-based variable number of gradients," *Proceedings of SPIE*, vol. 3650, pp. 36-43, 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," US patent 4 642 678, 1987.
- [7] F. Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: a review and a new contribution," *Proc. of IEEE*, vol. 83, pp. 858-876, June 1995.
- [8] B. K. Gunturk, Y. Altunbasak and R. M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Trans. Image Processing*, vol. 11, pp. 997-1013, 2002.
- [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, 1997.
- [10] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicing algorithm", *IEEE Trans. on Image Processing*, vol. 14, pp. 360-369, Mar. 2005.
- [11] R. Kakarala and Z. Baharav, "Adaptive demosaicing with the principal vector method," *IEEE Trans. Consumer Electronics*, vol. 48, pp. 932-937, Nov. 2002.
- [12] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Pearson Education; 1st edition, Mar. 1993.
- [13] R. Kimmel, "Demosaicing: Image reconstruction from CCD samples," *IEEE Trans. Image Processing*, vol. 8, pp. 1221-1228, 1999.
- [14] P. Kuhn, *Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation*, Kluwer Academic Publishers, Boston, 1999.
- [15] P. Longère, Xuemei Zhang, P. B. Delahunt and Davaid H. Brainard, "Perceptual assessment of demosaicing algorithm performance," *Proc. of IEEE*, vol. 90, pp. 123-132, 2002.
- [16] Wenmiao Lu and Yap-peng Tan, "Color filter array demosaicking: new method and performance measures," *IEEE Trans. Image Processing*, vol. 12, pp. 1194-1210, Oct. 2003.
- [17] R. Lukac and K.N. Plataniotis, "Normalized color-ratio modelling for CFA interpolation," *IEEE Trans. Consumer Electronics*, vol. 50, pp.737- 745, May 2004.

- [18] R. Lukac, K.N. Plataniotis, D. Hatzinakos, and M. Aleksic, "A novel cost effective demosaicing approach," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 256-261, February 2004.
- [19] S. C. Pei and I. K. Tam, "Effective color interpolation in CCD color filter arrays using signal correlation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, pp. 503-513, June 2003.
- [20] R. Ramanath and W. E. Snyder, "Adaptive demosaicking," *Journal of Electronic Imaging*, vol. 12, No. 4, pp. 633-642, 2003.
- [21] R. R. Schultz, Li Meng and R. L. Stevenson, "Subpixel motion estimation for super-resolution image sequence enhancement," *Journal of Visual Communication and Image Representation*, vol. 9, pp. 38-50, March 1998.
- [22] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, pp. 996-1011, June 1996.
- [23] C. Stiller and J. Konrad, "Estimating motion in image sequence," *IEEE Signal Processing Magazine*, No. 7, pp. 70-91, July, 1999.
- [24] B. C. Tom and A. K. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Trans. Image Processing*, vol. 10, pp. 278-287, Feb. 2001.
- [25] H. J. Trussel and R. E. Hartwing, "Mathematics for demosaicking," *IEEE Trans. Image Processing*, vol. 11, pp. 485-492, 2002.
- [26] X. Wu, W. K. Choi and Paul Bao, "Color restoration from digital camera data by pattern matching," *Proceedings of SPIE*, vol. 3018, pp. 12-17, 1997.
- [27] X. Wu and N. Zhang, "Primary-consistent soft-decision color demosaicking for digital cameras", *IEEE Trans. Image Processing*, vol. 13, pp. 1263-1274, Sept. 2004.
- [28] L. Zhang and X. Wu, "Color demosaicking via directional linear minimum mean square-error estimation," to appear in *IEEE Trans. Image Processing*.
- [29] L. Zhang and X. Wu, "On cross correlation function based discrete time delay estimation," *ICASSP 2005*, Philadelphia, PA, USA.