TEMPORAL COLOR VIDEO DEMOSAICKING VIA MOTION ESTIMATION AND OPTIMAL DATA FUSING

Xiaolin Wu and Lei Zhang
Department of Electrical and Computer Engineering, McMaster University
Hamilton, Ontario, Canada, L8S 2M4
{xwu, johnray}@mail.ece.mcmaster.ca

ABSTRACT

Demosaicking of the color CCD sensor data is an important task in digital image/video acquisition. Due to the Nyquist frequency limit of color filter array (CFA), it is impossible to faithfully reconstruct some high frequency structures in the original scene if one is limited to the mosaic samples of a still frame. However, a color digital video camera captures a sequence of mosaic images and the temporal dimension provides a wealth of information about the scene via camera and object motions. With the help of adjacent CCD frames by motion estimation and statistical data fusion, we can significantly improve the demosaicking quality of the video sequence. This paper presents a general approach of temporal demosaicking. Our experimental results demonstrate clear advantages of the presented temporal color demosaicking approach over its intra-frame counterparts both in PSNR measure and subjective visual quality.

1. INTRODUCTION

Color demosaicking holds a key to the quality of color images reconstructed from digital CCD cameras and camcorders. The problem has been extensively studied in spatial domain for still digital cameras [2-7], which we call intra-frame color demosaicking. The intra-frame demosaicking techniques, including those recently-developed sophisticated ones [5-7], can still fail to faithfully reconstruct features whose sampling frequencies exceed the Nyquist limit. In order to overcome this difficulty, additional knowledge of the original color signals is 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 color filter array (CFA). The inter-frame correlation of adjacent frames is particularly beneficial to the demosaicking through estimating the camera and object motions and statistical data fusion.

Surprisingly, there seems to be very little research reported on the temporal color demosaicking, despite its great potential. Very recently, Wu and Zhang proposed a joint spatial and temporal color demosaicking technique [8]. Their main idea is to match the CFA green sample blocks in adjacent frames in such a way that missing red and blue samples in one frame can be inferred from available red and blue samples of a matched adjacent frame. However, this technique is effective only 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. In this paper we present a more general and principled temporal color demosaicking approach. First, all frames are demosaicked individually. Then motions between adjacent frames are estimated based on the reconstructed green images. The estimated motion vectors spatially register adjacent frames. The color samples of multiple frames are fused by an optimal weighting process to improve the quality of individual frames.

The paper is structured as follows. In Sec. 2, we discuss the roles of intra-frame demosaicking of green channel and motion estimation in temporal color demosaicking. Then in Secs. 3 and 4 we propose a new temporal demosaicking technique, first for the green and then for the red and blue channels. Experimental results are presented in Sec. 5.

2. MOTION ESTIMATION FOR DEMOSAICKING

In the following development color demosaicking is carried out with respect to the ubiquitous Bayer CFA pattern [1] (see Fig. 1), although the resulting techniques can be easily generalized to other mosaic patterns.

Accurate motion estimation of the true images is pivotal to temporal demosaicking. Motion analysis with mosaic images is difficult because none of the three color channels is fully sampled. Motion estimation techniques, such as those developed for video compression applications, assume the input video frames to be of a fully sampled regular grid. But they can still be used on color mosaic data. The existing motion estimation techniques circumvent the problem by working on frames

0-7803-8554-3/04/$20.00 ©2004 IEEE.

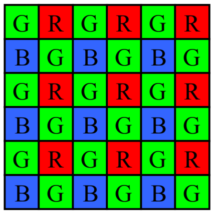


Figure 1. The Bayer Pattern.

that have been already reconstructed by an intra-frame demosaicking method.

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. To facilitate motion estimation we interpolate the missing green samples in the Bayer CFA using the well-known directional second-order Laplacian filter proposed by Hamilton and Adams [2]. They used the second order gradients of blue and red channels as the correction terms to interpolate the green channel. Any other interpolation methods can also be used [3-7]. We choose the directional second-order Laplacian filter as the spatial demosaicking method of each green frame for its simplicity and fast implementation.

In digital video consecutive frames are captured in a short time window. We can therefore assume that only translation motions take place in the scene in adjacent frames. This assumption is widely accepted in MPEG 2/4 video compression standards. In temporal demosaicking any given frame is associated with its neighboring frames both in the past and future. There is a rich body of research literature on the subject of motion estimation [9] in areas of video compression and computer vision. After all the green frames are interpolated, any of the existing motion estimation methods can be applied to the green sequence for the purpose of temporal color demosaicking. Motion compensation can be carried out in either integer or fractional pixel accuracy depending on the quality requirement and complexity constraint. The temporal color demosaicking approach to be presented in the next sections is independent of the motion estimation algorithm and the numerical precision of motion estimation.

3. TEMPORAL GREEN DEMOSAICKING

Recall from the previous section that the infra-frame color demosaicking process has already generated initial estimates of all the missing green samples. These intra-frame green estimates of adjacent frames will be fused by temporal demosaicking as follows.

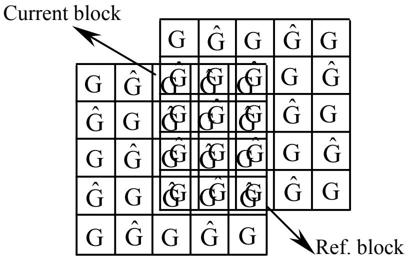


Figure 2. Current block and a reference block. G 's are the original CCD green samples and $\hat{G}$'s are the estimated green samples by intra-frame demosaicking.

Referring to Fig. 2, the original green samples are labeled by G and the estimated green values by the intra-frame demosaicking are labeled by $\hat{G}$. Let F_0 be a block of the current green frame and $\{F_i\}_{i=1,2,\ldots,K}$ be the matched blocks of F_0 in K reference frames via motion estimation. Denote by $\hat{G}_i$ the estimated green samples in block F_i, $i=0,1\ldots,K$, and they can be written as the noisy measurements of true samples G

$$\hat{G}_i = G + n_i, \quad i = 0,1\ldots,K \qquad (3\text{-}1)$$

where n_i are the corresponding measurement noises.

The objective of temporal color demosaicking is to fuse all the measurements $\hat{G}_i$ into a more robust estimate of G. To this end we adopt the weighted average strategy and let the fused estimate be

$$\overline{G} = \sum_{i=0}^{K} w_i \hat{G}_i \qquad (3\text{-}2)$$

where the sum of the affine weights w_i is 1. The criterion of determining w_i is to minimize the mean square error of the estimate $\overline{G}$, i.e., $E\left[(\overline{G}-G)^2\right] = \min$.

Denote by σ_i^2 the variance of noise n_i, $i=0,1\ldots,K$. Suppose that n_i and n_j, $i \neq j$, are uncorrelated. We can derive that the optimal weights $\mathbf{w} = col\{w_1, w_2, \ldots, w_K\}$ should be determined by

$$\mathbf{w} = \mathbf{S}^{-1}\mathbf{1} \qquad (3\text{-}3)$$

where **1** is the unit column vector and matrix **S** is

$$\mathbf{S} = \begin{bmatrix} 1+\sigma_1^2/\sigma_0^2 & 1 & \cdots & 1 \\ 1 & 1+\sigma_2^2/\sigma_0^2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1+\sigma_K^2/\sigma_0^2 \end{bmatrix} \qquad (3\text{-}4)$$

The remaining issue is how to estimate σ_i^2. Since n_i and n_j are uncorrelated, $i \neq j$, from (3-1) we have

$$d_{i,j} = E[(\hat{G}_i - \hat{G}_j)^2] = \sigma_i^2 + \sigma_j^2 \qquad (3\text{-}5)$$

We assume that the noise variances are constant for the associated samples in a block. Then the values of $d_{i,j}$ can be estimated adaptively in blocks F_i and F_j:

$$d_{i,j}=\frac{1}{M}\sum_{\hat{G}_i\in F_i\,,\hat{G}_j\in F_j}\left(\hat{G}_i-\hat{G}_j\right)^2 \tag{3-6}$$

where M is the number of missing green samples in the current frame F_0.

Consider $\mathbf{d}=col\{d_{0,1},...,d_{0,K},d_{1,2},...,d_{1,K},...,d_{K-1,K}\}$, the vector of all possible $K(K+1)/2$ pairs of $d_{i,j}$, and the $K(K+1)/2\times(K+1)$ matrix $\mathbf{H}$ such that

$$\mathbf{d}=\mathbf{H}\boldsymbol{\sigma} \tag{3-7}$$

where $\boldsymbol{\sigma}=col\{\sigma_0^2,...,\sigma_K^2\}$. Let $\mathbf{h}_{i,j}$ be the row of $\mathbf{H}$ such that $d_{i,j}=\mathbf{h}_{i,j}\boldsymbol{\sigma}$. Obviously only the i^{th} and j^{th} elements in $\mathbf{h}_{i,j}$ are 1 and all other elements are zeros. We can estimate $\boldsymbol{\sigma}$ by the least square estimation technique:

$$\boldsymbol{\sigma}=\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{d} \tag{3-8}$$

After $\mathbf{w}$ is obtained, w_0 can be computed by subtracting all $w_i, i=1,2,...,K$ from 1. Finally the fused estimate of G is computed by (3-2).

4. TEMPORAL RED/BLUE DEMOSAICKING

Estimating the missing R and B samples is more difficult because of the inferior sampling scheme of the Bayer CFA for the red and blue channels. The 2D sampling grid for R and B has a poor shape of square lattice which deviates greatly from the optimal hexagonal sample grid. This is why we demosaick the denser and better-shaped checker board green sample grid first and then use the demosaicked green channel to aid the demosaicking of the red and blue channels. In [8] a simple technique called motion-induced G-pattern matching (MIGM) was proven to be effective to estimate missing R and B samples of the current frame from reference frames. In this section we incorporate MIGM into the optimal data fusion framework developed in the previous section.

Denote by B_R (R_B) the missing B (R) value at the R (B) sample position in the Bayer CFA. The spatially closest available B (R) samples provided by CCD are at the four corners of B_R (R_B), representing the most difficult case of infra-frame demosaicking. Ideally, if the motion vector happens to have odd offset in both x and y directions, i.e., $\Delta x\in\{\pm1,\pm3,...\}$ and $\Delta y\in\{\pm1,\pm3,...\}$, then B_R (R_B) of the current frame can be matched to an existing B (R) sample of a reference frame. Assuming that the precision of motion estimation is one pixel, there is a probability of ¼ that both Δx and Δy are odd with respect to an adjacent frame. Therefore, to have a good chance of match we check four or more adjacent frames in motion estimation. These inter-frame estimates are fused with the corresponding intra-frame estimate of B_R (R_B) as described in the previous section. It should be stressed that the inter-frame green estimates $\overline{G}$ can improve the intra-frame estimate of B_R (R_B) as well, because all intra-frame demosaicking methods rely on correlation between color channels.

The remaining case is the estimation of missing B (R) value at the G sample position in the Bayer CFA, denoted by B_G (R_G). The closest B (R) samples in the current frame provided by CCD are either horizontal or vertical neighbors of B_G (R_G), which can be beneficial or counterproductive in infra-frame estimation of B_G (R_G) depending on the image gradient. If a motion vector is of form either $(\Delta x,\Delta y)=(2i,2j+1)$ or $(\Delta x,\Delta y)=(2i+1,2j)$, then either B_G is matched to an existing B sample or R_G to an existing R sample in the reference frame. Again, these temporal demosaicking results can be fused with the corresponding intra-frame estimates as discussed in Section 3.

5. EXPERIMENTS

The proposed temporal demosaicking algorithm was implemented and tested on a sequence of simulated CCD video frames, which was generated by down-sampling a high-resolution digitized film provided by IMAX Corp., according to the Bayer CFA. This gives us all the true color values so that we can validate the demosaicking results. We used the algorithms of [2] to demosaick each frame spatially, and employed the methods of [2] and [6] for comparison. Fig. 3 (a) is a portion of an original frame. Figs. 3 (b) ~ (d) present the demosaicked images by the algorithms of [2], [6] and the proposed method. Figs. 4 (a) ~ (d) are the zoom-in images of Figs. 3 (a) ~ (d). Note the color artifacts on the grill of the car in Fig. 3 (b). Although the method of [6] does not yield visible artifacts on the grill, it fails when colors changes sharply (referring to Fig. 4 (c)), a case where the method of [2] also fails. Figs. 3 (d) and 4 (d) show that the proposed temporal demosaicking method removes most of the artifacts. Table 1 presents the average PSNR values of six consecutive demosaicked frames by the three methods.

REFERENCE

[1] B. E. Bayer and Eastman Kodak Company, "Color Imaging Array," US patent 3 971 065, 1975.

[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] 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.

[4] R. Kimmel, "Demosaicing: Image reconstruction from CCD samples," *IEEE Trans. Image Processing*, vol. 8, pp. 1221-1228, 1999.

[5] 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.

[6] L. Zhang and X. Wu, "Color demosaicking via directional linear minimum mean square-error estimation," *IEEE Trans. Image Processing* (accepted).

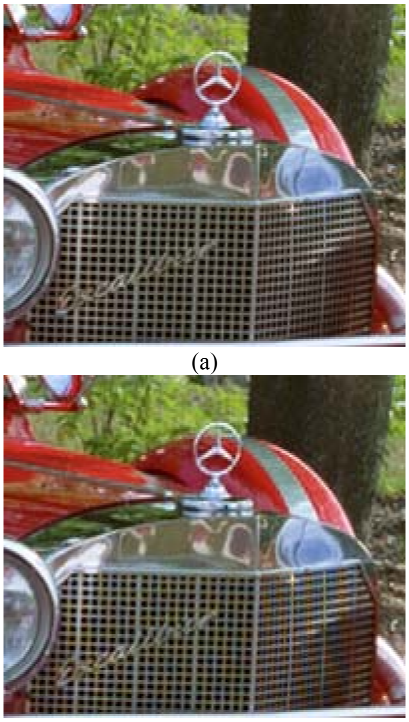
[7] X. Wu and N. Zhang, "Primary-consistent soft-decision color demosaicking for digital cameras", *IEEE Trans. Image Processing*, 2004 (to appear).

[8] X. Wu and N. Zhang, "Joint temporal and spatial color demosaicking", *Proc. of Electronic Imaging Conf.* SPIE, 2004.

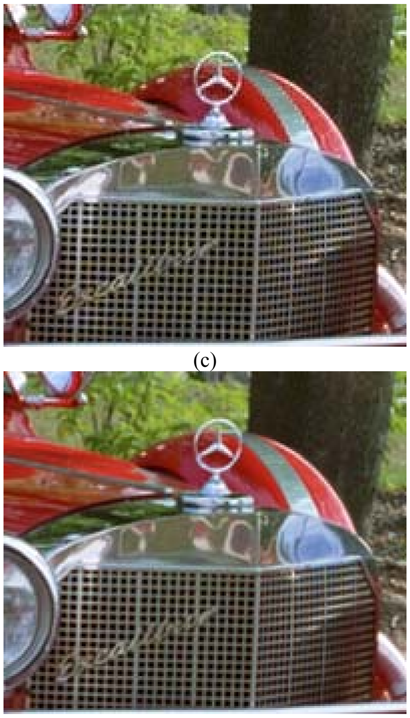
[9] Kuhn, P., *Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation*, Kluwer Academic Publishers, Boston, 1999.

Table 1. Average PSNR (dB) values of six consecutive demosaicked frames by intra and inter-frame demosaicking.

Channel	R	G	B
Method of [2]	31.47	33.63	27.86
Method of [6]	30.80	35.45	29.24
Proposed method	33.84	36.22	30.41

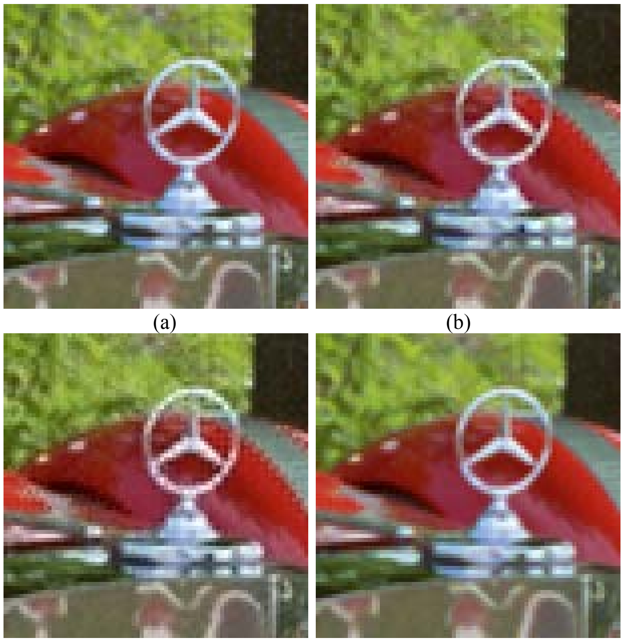

(a)

(b)


(c)

(d)

Figure 3. (a) Original image; (b) ~ (d) are the demosaicked images by the methods of [2], [6] and the proposed method respectively.


(a) (b)

(c) (d)

Figure 4. Zoom-in images of (a) the original image and (b) ~ (d) the demosaicked images by the methods of [2], [6] and the proposed method.