Fast Video Demosaicking Solution for Mobile Phone Imaging Applications

Rastislav Lukac, *Member, IEEE*, and Konstantinos N. Plataniotis, *Senior Member, IEEE*

Abstract — *This paper introduces a cost-effective video demosaicking solution suitable for mobile phones with video capturing capabilities. The proposed solution is a spatio-temporal demosaicking scheme which operates over three subsequent frames of the acquired video sequence. To rapidly demosaick the raw, color filter array (CFA), spatiotemporal video data our solution uses multi-stage processing concepts and builds on the principles of refined spectral modeling and self-adaptation to varying spatiotemporal characteristics. The proposed solution accommodates cost-effective considerations inherent in mobile phone imaging and produces a visually pleasing demosaicked video sequence.* [1]

Index Terms — **Single-sensor imaging, digital camera image processing, video demosaicking, spatiotemoral video processing, Bayer pattern, color filter array interpolation, imaging-enabled consumer electronics.**

I. INTRODUCTION

Recent advances in hardware, digital image/signal processing and telecommunications have allowed for the miniaturization and incorporation of digital cameras in consumer electronic devices, such as mobile phones and wireless PDAs [1]. Due to the cost-effective considerations, such devices use only a single image sensor, usually a charge-coupled device (CCD) [2] or complementary metal oxide semiconductor (CMOS) [3] sensor. The monochromatic nature of the sensor necessitates to cover its surface by a color filter array (CFA) to capture all the three, Red-Green-Blue (RGB) primary colors at the same time [4],[5]. Since each sensor cell has its own spectrally selective filter (Fig. 1), the acquired image data constitute a mosaic-like gray-scale image (Fig. 2a). To recover the color information, the two missing color components are estimated from the adjacent pixels using a process called demosaicking [6]-[9].

Most single-sensor devices are used to capture digital still images. However, it is common today to record motion video or image sequences using digital still image or digital video cameras. Such a visual input represents a three-dimensional (3-D) image signal or a time sequence of two-dimensional (2-D) images (frames), and it usually exhibits significant correlation in both the spatial and temporal sense [10]. By omitting the essential temporal characteristics, spatial processing methods, which process separately the individual frames, produce an output video
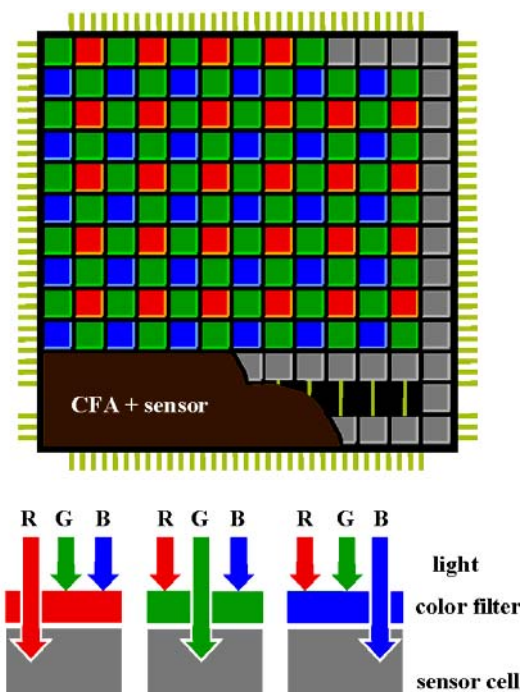
The authors are with The Edward S. Rogers Sr. Department of ECE, University of Toronto, Toronto, Canada.

Corresponding Author: Dr. Rastislav Lukac, Multimedia Laboratory, Room BA 4157, The Edward S. Rogers Sr. Department of ECE, University of Toronto, 10 King's College Road, Toronto, Ontario, M5S 3G4, Canada (e-mail: lukacr@ieee.org, web: http://www.dsp.utoronto.ca/~lukacr)

Fig. 1. Single-sensor imaging devices use a CFA, such as the considered here Bayer CFA pattern with the GRGR phase in the first row [5], placed at the top of the CCD or CMOS sensor.

sequence with motion artifacts [10]-[14]. To prevent this from happening, a well designed video processing technique should follow the representation of the signal, and utilize all the available information during operation. To that end, we introduce a unique spatiotemporal processing solution which can demosaick a sequence of CFA image frames. Following the findings in conventional video processing, and still image demosaicking, our solution uses the spatial, temporal, spectral and structural characteristics of the captured CFA video to produce a full-color, demosaicked, image sequence at the output. Please note that although extensive research has been devoted to the demosaicking of still images, there is no known work focusing on the video demosaicking for consumer electronics.

The rest of this paper is organized as follows. The proposed video demosaicking solution is introduced in Section II. Motivation and design characteristics are discussed in detail. In Section III, the proposed solution is tested using various color image sequences. Evaluations of performance, both objective and subjective, are provided. Finally, conclusions are drawn in Section IV.
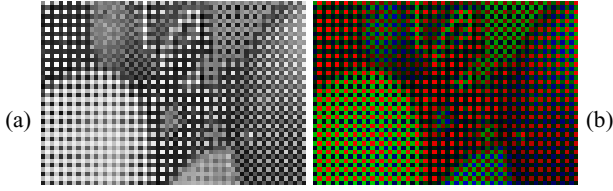
Fig. 2. CFA image (frame) obtained using the Bayer CFA pattern: (a) raw sensor data, (b) CFA data arranged as a color image.

Fig. 3. G component demosaicking step: (a) a localized area in the CFA image sequence, and (b)-(g) unidirectional filtering windows; (b) D_1, (c) D_2, (d) D_3, (e) D_4, (f) D_5, (g) D_6.

II. PROPOSED VIDEO DEMOSAICKING SOLUTION

Let us consider a single-sensor device which uses the Bayer CFA (Fig. 1) to capture the raw (sensor, gray-scale) CFA data. Assuming that the device operates in a video-recording mode, the acquired data constitute a $K_1 \times K_2 \times K_3$ gray-scale image sequence $z : Z^3 \to Z$ of the K_3 frames $z(t) : Z^2 \to Z$ with the dimensions of $K_1 \times K_2$ image pixels $z_{(r,s,t)}$. Each CFA frame $z(t)$, for $t = 1, 2, ..., K_3$, represents a two-dimensional matrix of integer samples $z_{(r,s,t)}$, with $r = 1, 2, ..., K_1$ and $s = 1, 2, ..., K_2$ denoting the frame rows and columns, respectively. Based on the arrangements of color filters in the Bayer CFA, the gray-scale frame $z(t)$ shown in Fig. 2a can be transformed to a $K_1 \times K_2$ color (RGB) frame $\mathbf{x}(t) : Z^2 \to Z^3$ depicted in Fig. 2b, as follows [15],[16]:

$$\mathbf{x}_{(r,s,t)} = \begin{cases} [z_{(r,s,t)}, 0, 0] & \text{for } r \text{ odd and } s \text{ even,} \\ [0, 0, z_{(r,s,t)}] & \text{for } r \text{ even and } s \text{ odd,} \\ [0, z_{(r,s,t)}, 0] & \text{otherwise.} \end{cases} \quad (1)$$

where $\mathbf{x}_{(r,s,t)} = [x_{(r,s,t)1}, x_{(r,s,t)2}, x_{(r,s,t)3}]$ denote an RGB vector with $x_{(r,s,t)k}$ indicating the R $(k = 1)$, G $(k = 2)$ and B $(k = 3)$ component in a two-dimensional matrix $\mathbf{x}(t)$ of three-component samples $\mathbf{x}_{(r,s,t)}$, and $z_{(r,s,t)}$ denote a pixel in the CFA frame $z(t)$. By repeating the transformation in (1) over the temporal axis for $t = 1, 2, ..., K_3$, the process constitutes a $K_1 \times K_2 \times K_3$ color image sequence $\mathbf{x} : Z^3 \to Z^3$. Since each spatial locations (r,s) correspond to different RGB spectral bands, both image sequences z and $\mathbf{x}$ have a mosaic-like spatial structure.

Due to the double sampling frequency of the G components in the CFA frames achieved by the Bayer pattern, the proposed method starts the demosaicking process by processing of the G color plane in $\mathbf{x}(t)$. To eliminate color shifts and artifacts, each demosaicking operation reported in this paper employs the spectral model. This is used under the assumption that a typical natural image sequence exhibits significant spectral correlation among the RGB color planes of the individual frames [15],[16]. Please note that although recent research indicates that the sophisticated vector spectral model in [15] and the normalized ratio models in [16],[17] outperform the widely used conventional color-ratio [18] and color-difference [19] modeling approaches, the color-difference model is used throughout this paper due to the cost-effective considerations in mobile phone imaging.
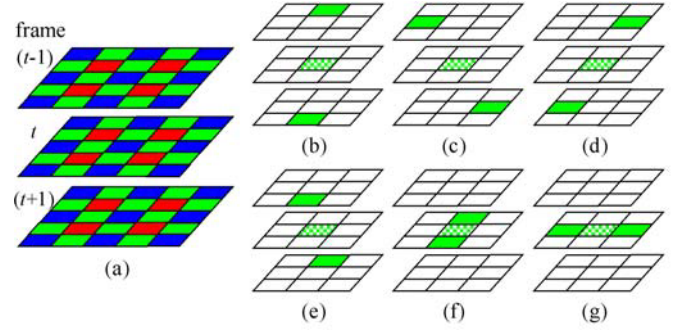
Following the spatiotemporal processing paradigm in [10]-[14], the proposed video demosaicking solution can be viewed as a spatiotemporal method which operates over three consecutive video frames $\mathbf{x}(t+i)$, with $i = -1$ indicating the past, $i = 0$ indicating the actual, and $i = 1$ denoting the future frame. In these three frames, our method forms the following spectral quantities $c_{(r-1,s,t+i)}, c_{(r,s-1,t+i)}, c_{(r,s+1,t+i)}, c_{(r+1,s,t+i)}$ which are defined as:

$$c_{(r-1,s,t+i)} = x_{(r-1,s,t+i)2} - (x_{(r,s,t+i)k} + x_{(r-2,s,t+i)k})/2$$
$$c_{(r,s-1,t+i)} = x_{(r,s-1,t+i)2} - (x_{(r,s,t+i)k} + x_{(r,s-2,t+i)k})/2$$
$$c_{(r,s+1,t+i)} = x_{(r,s+1,t+i)2} - (x_{(r,s,t+i)k} + x_{(r,s+2,t+i)k})/2 \quad (2)$$
$$c_{(r+1,s,t+i)} = x_{(r+1,s,t+i)2} - (x_{(r,s,t+i)k} + x_{(r+2,s,t+i)k})/2$$

In the equation above, the values $x_{(\cdot,\cdot,t+i)2}$ denote the original G CFA components located in the $(t+i)$th frame, while $(x_{(\cdot,\cdot,t+i)k} + x_{(\cdot,\cdot,t+i)k})/2$ denote the predicted R $(k = 1)$ or B $(k = 3)$ values in the same frame. If the demosaicked location (r,s,t) corresponds to an original R CFA component, then the parameter $k = 1$ is used in (2). Otherwise, the G component related demosaicking operations are executed in the original B CFA locations necessitating the use of parameter $k = 3$ in the spectral modeling step (2).

To follow the varying spatial and temporal characteristics in the image sequence $\mathbf{x}$, and at the same time to keep the computational complexity low, we use the set of unidirectional filtering windows [14],[20] centered at (r,s,t). As it can be seen in Figs. 3b-g, each of the six possible windows is centered at the demosaicked location (r,s,t) and contains two spectral quantities calculated in (2), which correspond to the window's marginal locations. Thus, each direction D_j, for $j = 1, 2, ..., 6$, can be represented as follows:

$$\bar{c}_1 = (c_{(r-1,s,t-1)} + c_{(r+1,s,t+1)})/2$$
$$\bar{c}_2 = (c_{(r,s-1,t-1)} + c_{(r,s+1,t+1)})/2$$
$$\bar{c}_3 = (c_{(r,s+1,t-1)} + c_{(r,s-1,t+1)})/2$$
$$\bar{c}_4 = (c_{(r+1,s,t-1)} + c_{(r-1,s,t+1)})/2 \quad (3)$$
$$\bar{c}_5 = (c_{(r-1,s,t)} + c_{(r+1,s,t)})/2$$
$$\bar{c}_6 = (c_{(r,s-1,t)} + c_{(r,s+1,t)})/2$$

Fig. 4. Initial R (or B) component demosaicking step: (a) a localized area in the CFA image sequence, and (b)-(g) unidirectional filtering windows; (b) D_1, (c) D_2, (d) D_3, (e) D_4, (f) D_5, (g) D_6.
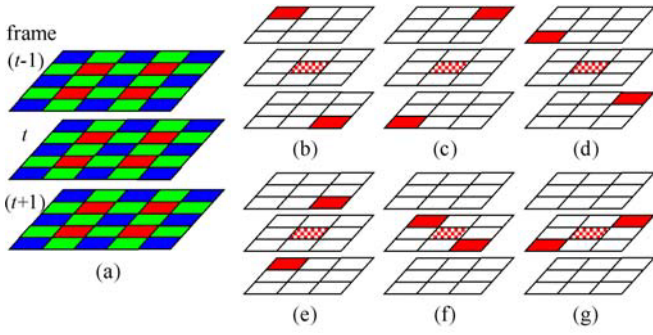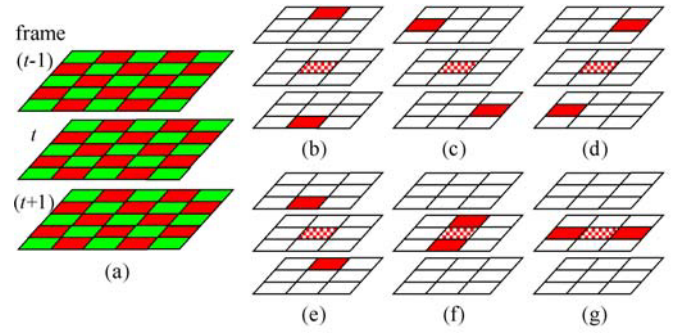

Fig. 5. Final R (or B) component demosaicking step: (a) a localized area in the CFA image sequence, and (b)-(g) unidirectional filtering windows; (b) D_1, (c) D_2, (d) D_3, (e) D_4, (f) D_5, (g) D_6.

The propriety of the above values should be evaluated with respect to structural (spatial) and motion (temporal) characteristics in the localized video area under consideration. The cost-effective calculations use the spectral quantities in (2) to enumerate the following adaptive coefficients:

$$
\begin{aligned}
w_1 &= 1/(1+ |\, c_{(r-1,s,t-1)} - c_{(r+1,s,t+1)}\, |) \\
w_2 &= 1/(1+ |\, c_{(r,s-1,t-1)} - c_{(r,s+1,t+1)}\, |) \\
w_3 &= 1/(1+ |\, c_{(r,s+1,t-1)} - c_{(r,s-1,t+1)}\, |) \\
w_4 &= 1/(1+ |\, c_{(r+1,s,t-1)} - c_{(r-1,s,t+1)}\, |) \\
w_5 &= 1/(1+ |\, c_{(r-1,s,t)} - c_{(r+1,s,t)}\, |) \\
w_6 &= 1/(1+ |\, c_{(r,s-1,t)} - c_{(r,s+1,t)}\, |)
\end{aligned}
\tag{4}
$$

Each weighting coefficient w_j, for $j = 1, 2, ..., 6$, is associated with the value $\overline{c}_j$ in (3). Large values of the absolute difference terms in (4) indicate that the corresponding components, which have been used to calculate $\overline{c}_j$, are located across spatiotemporal edges. Since (4) is defined using inverse gradients, the weights w_j are used to regulate the contribution of the associated quantities $\overline{c}_j$ in the demosaicking operations as follows:

$$
x_{(r,s,t)2} = x_{(r,s,t)k} + \sum_{j=1}^{6}\{w_j \overline{c}_j\} / \sum_{j=1}^{6} w_j
\tag{5}
$$

where $x_{(r,s,t)2}$ is the G component being demosaicked and $x_{(r,s,t)k}$ is the original R ($k = 1$) or B ($k = 3$) CFA component used to normalize the output of the weighted averaging operation. The coefficients w_j are used to emphasize spectral inputs $\overline{c}_j$ which are not positioned across an edge and to direct the demosaicking process along natural spatiotemporal edges in the image sequence. This ensures a sharply formed demosaicked output video.

Assuming that the G plane of the image frames $\mathbf{x}(t)$ has been fully populated via (5), the proposed video demosaicking scheme continues by populating the R and B color planes. Following the arrangement of the CFA components shown in Fig. 4a, the procedure first recovers the R (and B) components at B (and R) CFA locations. Since this demosaicking step

forms new arrangements, similar to the one shown in Fig. 5a, the procedure is completed by recovering of the R and B components at G CFA locations. The two demosaicking steps described above use the identical unbiased spectral estimator which is defined as follows:

$$
x_{(r,s,t)k} = x_{(r,s,t)2} + \sum_{j=1}^{6}\{w_j \overline{c}_j\} / \sum_{j=1}^{6} w_j
\tag{6}
$$

where $x_{(r,s,t)k}$ is the R ($k = 1$) or B ($k = 3$) component being demosaicked. In the scenario depicted in Fig. 4, the normalizing component $x_{(r,s,t)2}$ in (6) is the demosaicked G component obtained using (5), while for the situation depicted in Fig. 5 the normalizing component $x_{(r,s,t)2}$ is the original G CFA component.

The weights w_j and the values $\overline{c}_j$, for $j = 1, 2, ..., 6$, are obtained during the initial R and B demosaicking step (Fig. 4) using the spectral quantities $c_{(\cdot,\cdot,\cdot)}$ defined as follows:

$$
\begin{aligned}
c_{(r-1,s-1,t+i)} &= x_{(r-1,s-1,t+i)k} - x_{(r-1,s-1,t+i)2} \\
c_{(r-1,s+1,t+i)} &= x_{(r-1,s+1,t+i)k} - x_{(r-1,s+1,t+i)2} \\
c_{(r+1,s-1,t+i)} &= x_{(r+1,s-1,t+i)k} - x_{(r+1,s-1,t+i)2} \\
c_{(r+1,s+1,t+i)} &= x_{(r+1,s+1,t+i)k} - x_{(r+1,s+1,t+i)2}
\end{aligned}
\tag{7}
$$

As before, the value of i is used to indicate the components in the past $(i = -1)$, actual $(i = 0)$ and future $(i = 1)$ frame $\mathbf{x}(t+i)$. As it is shown in Figs. 4b-g, by utilizing a set of unidirectional windows along the six possible directions D_j, for $j = 1, 2, ..., 6$, the spectral quantities $c_{(\cdot,\cdot,\cdot)}$ in (7) can be used in the calculation of the spectral inputs $\overline{c}_j$ in (6) as follows:

$$
\begin{aligned}
\overline{c}_1 &= (c_{(r-1,s-1,t-1)} + c_{(r+1,s+1,t+1)})/2 \\
\overline{c}_2 &= (c_{(r-1,s+1,t-1)} + c_{(r+1,s-1,t+1)})/2 \\
\overline{c}_3 &= (c_{(r+1,s-1,t-1)} + c_{(r-1,s+1,t+1)})/2 \\
\overline{c}_4 &= (c_{(r+1,s+1,t-1)} + c_{(r-1,s-1,t+1)})/2 \\
\overline{c}_5 &= (c_{(r-1,s-1,t)} + c_{(r+1,s+1,t)})/2 \\
\overline{c}_6 &= (c_{(r-1,s+1,t)} + c_{(r+1,s-1,t)})/2
\end{aligned}
\tag{8}
$$

Input:

$K_1 \times K_2 \times K_3$ CFA (gray-scale) video sequence z (Bayer CFA is used).

Video-demosaicking procedure:

1. Re-arrange z to $K_1 \times K_2 \times K_3$ color image sequence $\mathbf{x}$ using (1) with $t = 1, 2, ..., K_3$.
2. For $t = 1, 2$ and $t = K_3 - 1, K_3$ perform BI demosaicking.
3. For $t = 3, 4, ..., K_3 - 2$ perform (5) using (2)-(4) with $k = 1$ at R CFA and $k = 3$ at B CFA locations (r, s, t).
4. For $t = 3, 4, ..., K_3 - 2$ perform (6) using (7)-(9) at R and B CFA locations (r, s, t). Use $k = 1$ and $k = 3$ in populating of the R and B plane, respectively.
5. For $t = 3, 4, ..., K_3 - 2$ perform (6) using (10),(3) and (4) at G CFA locations (r, s, t). Use $k = 1$ and $k = 3$ in populating of the R and B plane, respectively.

Output:

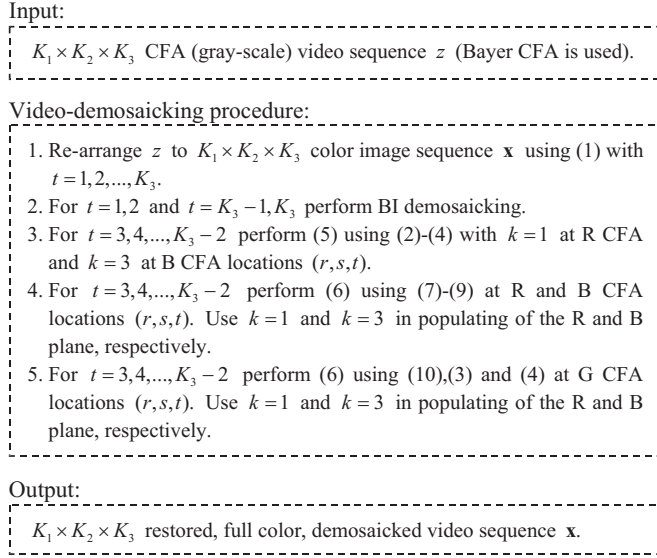$K_1 \times K_2 \times K_3$ restored, full color, demosaicked video sequence $\mathbf{x}$.

Fig. 6. Pseudo-code of the proposed algorithm suitable for the personal computer based implementation.

The corresponding weights w_j are obtained via

$$
\begin{aligned}
w_1 &= 1/(1 + |c_{(r-1,s-1,t-1)} - c_{(r+1,s+1,t+1)}|) \\
w_2 &= 1/(1 + |c_{(r-1,s+1,t-1)} - c_{(r+1,s-1,t+1)}|) \\
w_3 &= 1/(1 + |c_{(r+1,s-1,t-1)} - c_{(r-1,s+1,t+1)}|) \\
w_4 &= 1/(1 + |c_{(r+1,s+1,t-1)} - c_{(r-1,s-1,t+1)}|) \\
w_5 &= 1/(1 + |c_{(r-1,s-1,t)} - c_{(r+1,s+1,t)}|) \\
w_6 &= 1/(1 + |c_{(r-1,s+1,t)} - c_{(r+1,s-1,t)}|)
\end{aligned}
\tag{9}
$$

Assuming that the demosaicking step (6) with the inputs obtained through (7)-(9) has been completed for all frames $\mathbf{x}(t)$, the proposed solution concludes the demosaicking process by recovering the remaining R and B components. This is realized using the unbiased spectral estimator (6) operating over the spatial arrangements shown in Fig. 5. In this case, (6) uses the inputs obtained in (3) and (4) using the spectral quantities $c_{(\cdot,\cdot,\cdot)}$, which are defined as follows:

$$
\begin{aligned}
c_{(r-1,s,t+i)} &= x_{(r-1,s,t+i)k} - x_{(r-1,s,t+i)2} \\
c_{(r,s-1,t+i)} &= x_{(r,s-1,t+i)k} - x_{(r,s-1,t+i)2} \\
c_{(r,s+1,t+i)} &= x_{(r,s+1,t+i)k} - x_{(r,s+1,t+i)2} \\
c_{(r+1,s,t+i)} &= x_{(r+1,s,t+i)k} - x_{(r+1,s,t+i)2}
\end{aligned}
\tag{10}
$$

The proposed video demosaicking solution can be implemented in i) a companion personal computer (PC) which interfaces with a single-sensor device that stores the video in raw CFA format, or ii) a conventional digital camera which stores the demosaicked, full-color (RGB) video. Since each one of the above configurations may necessitate a specific demosaicking approach, two implementations are proposed.

Input:

$K_1 \times K_2 \times K_3$ CFA (gray-scale) video sequence z (Bayer CFA is used).

Video-demosaicking procedure:

1. For $t = 1$ and $t = 2$ re-arrange CFA frame $z(t)$ to $K_1 \times K_2$ color frame $\mathbf{x}(t)$ using (1) and perform BI demosaicking.
2. For $t = 3$ re-arrange CFA frame $z(t)$ to $K_1 \times K_2$ color frame $\mathbf{x}(t)$ using (1) and perform (5) using (2)-(4) with $k = 1$ at R CFA and $k = 3$ at B CFA locations (r, s, t).
3. For $t = 4$ re-arrange CFA frame $z(t)$ to $K_1 \times K_2$ color frame $\mathbf{x}(t)$ using (1) and perform (5) using (2)-(4) with $k = 1$ at R CFA and $k = 3$ at B CFA locations (r, s, t). Then perform (6) using (7)-(9) at R and B CFA locations $(r, s, t-1)$. Use $k = 1$ and $k = 3$ in populating of the R and B plane, respectively.
4. For $t = 5, 6, ..., K_3 - 2$ re-arrange CFA frame $z(t)$ to $K_1 \times K_2$ color frame $\mathbf{x}(t)$ using (1) and perform (5) using (2)-(4) with $k = 1$ at R CFA and $k = 3$ at B CFA locations (r, s, t). Then perform (6) using (7)-(9) at R and B CFA locations $(r, s, t-1)$ and perform (6) using (10),(3) and (4) at G CFA locations $(r, s, t-2)$. Use $k = 1$ and $k = 3$ in populating of the R and B plane, respectively.
5. For $t = K_3 - 1$ re-arrange CFA frame $z(t)$ to $K_1 \times K_2$ color frame $\mathbf{x}(t)$ using (1) and perform BI demosaicking. Then perform (6) using (7)-(9) at R and B CFA locations $(r, s, t-1)$ and perform (6) using (10),(3) and (4) at G CFA locations $(r, s, t-2)$. Use $k = 1$ and $k = 3$ in populating of the R and B plane, respectively.
6. For $t = K_3$ re-arrange CFA frame $z(t)$ to $K_1 \times K_2$ color frame $\mathbf{x}(t)$ using (1) and perform BI demosaicking. Then perform (6) using (10),(3) and (4) at G CFA locations $(r, s, t-2)$. Use $k = 1$ and $k = 3$ in populating of the R and B plane, respectively.

Output:

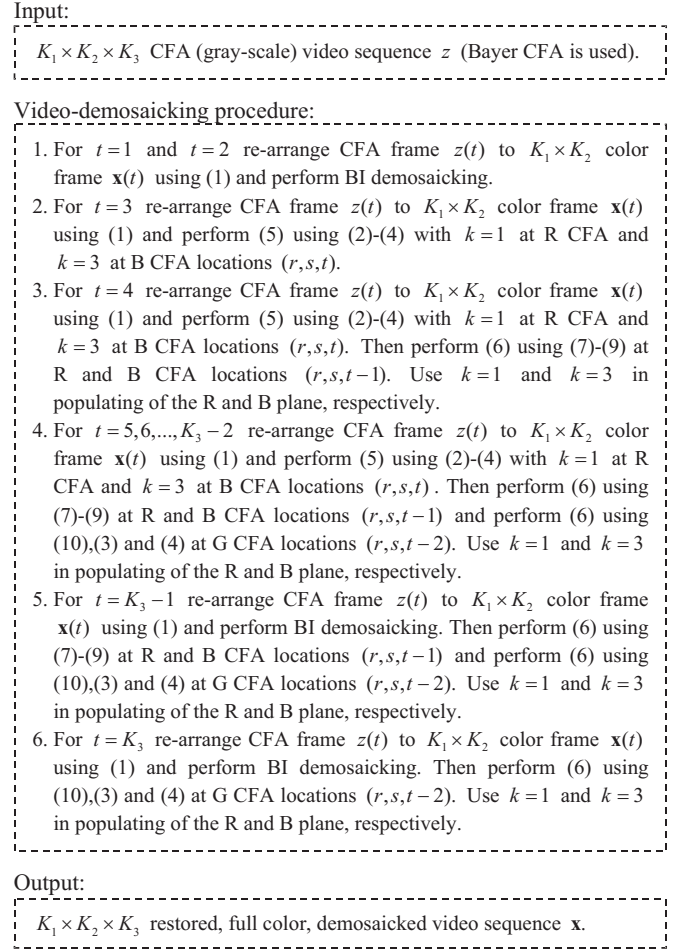$K_1 \times K_2 \times K_3$ restored, full color, demosaicked video sequence $\mathbf{x}$.

Fig. 7. Pseudo-code of the proposed algorithm suitable for the hardware implementation.

Figs. 6-7 list a pseudo-code representation of the proposed video demosaicking algorithm. Since our method is a spatiotemporal solution which operates in (5)-(6) over three consecutive frames, the two first and last CFA frames must be processed by a conventional (spatial only) demosaicking solution in order to eliminate the so-called border effect. For cost-effective considerations we will make use of the bilinear interpolation (BI) scheme [21], which is considered an industry standard with low computational cost.

In the case of PC-based implementations (Fig. 6), the frames' manipulation may not be an important factor. Therefore, the demosaicking step in (5) with (2)-(4) can be completed along the time axis t. Then, the procedure performs the processing step which is realized using (6) with (7)-(9) for all the possible values of t. Finally, the processing cycle completes using (6) with (10),(3),(4) for all t values.

However, implementations embedded in the camera or mobile phone hardware should take into consideration real-time implementation constraints. Therefore, the CFA frames coming from the sensor must be immediately demosaicked to produce the reconstructed, full-color frames for further real-

Fig. 8. Test color image sequences: (a) *Nature* **for** $t = 1,10,30,65,99$**; (b)** *Rafting* **for** $t = 1,25,50,75,99$**; (c)** *Bikes* **for** $t = 1,25,50,75,99$**; and (d)** *Coast* **for** $t = 1,20,50,75,99$**.**

time processing (e.g. MPEG compression), visualization or storing purposes. Please note that the demosaicked step (6) realized using (10),(3),(4) requests the presence of the components obtained in (6) with (7)-(9), as well as those obtained in (5) with (2)-(4). Similarly, the operations in (6) with (7)-(9) necessitate the use of the components calculated in (5) with (2)-(4). Since each one of the demosaicked steps (5) and (6) necessitates in the input the components from the three consecutive frames, the real-time processor should temporally access five consecutive frames $\mathbf{x}(t+i)$, for $i = -3, -2, ..., 1$, as it can be seen from a system level diagram depicted in Fig. 7. To delay the input and output frames, shift registers have to be incorporated into the architecture.

III. EXPERIMENTAL RESULTS

To determine the performance of the proposed demosaicking solution, a number of test color image sequences have been utilized. Examples such as the $300 \times 200 \times 99$ test color videos shown in Fig. 8 are used to emulate the application scenario. These test videos, which vary in motion, color appearance, and in the complexity of the structural content (edges) have been captured using three-sensor devices and normalized to 8-bit per channel RGB representation. Tests were performed by sampling the original image sequences (Fig. 8) with a GRGR-phased Bayer CFA filter (Fig. 1) to obtain a CFA image sequence z with pixels $z_{(r,s,t)}$ following the standard practice previously reported in [5]-[9],[15],[16]:

$$z_{(r,s,t)} = \begin{cases} o_{(r,s,t)1} & \text{for } r \text{ odd and } s \text{ even,} \\ o_{(r,s,t)3} & \text{for } r \text{ even and } s \text{ odd,} \\ o_{(r,s,t)2} & \text{otherwise.} \end{cases} \quad (11)$$

where $o_{(r,s,t)k}$ denotes the R $(k = 1)$, G $(k = 2)$ and B $(k = 3)$ component of the original color vector $\mathbf{o}_{(r,s,t)} = [o_{(r,s,t)1}, o_{(r,s,t)2}, o_{(r,s,t)3}]$ with $r = 1, 2, ..., K_1$, $s = 1, 2, ..., K_2$, and $t = 1, 2, ..., K_3$. The demosaicked image sequence $\mathbf{x}$ is generated by applying the demosaicking solution onto the CFA image sequence z. Performance was measured by comparing the original full color image sequences to the demosaicked videos. Fig. 9 visualizes the evaluation procedure used in this
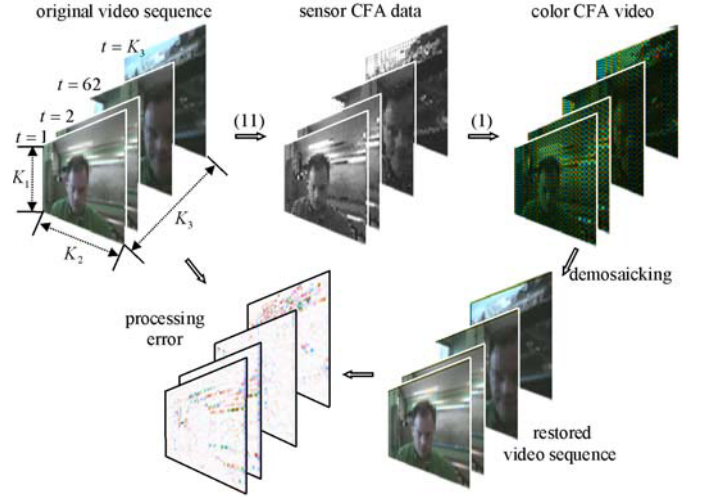

Fig. 9. Evaluation procedure.

experimentation. To facilitate the objective comparisons [1],[15], the RGB color space based mean absolute error (MAE) and mean square error (MSE) criteria, and the CIE-LUV color space based normalized color difference (NCD) criterion are used in this work. The MAE and MSE are defined as follows:

$$\text{MAE} = \frac{1}{k K_1 K_2 K_3} \sum_{r=1}^{K_1} \sum_{s=1}^{K_2} \sum_{t=1}^{K_3} \sum_{k=1}^{3} \left| o_{(r,s,t)k} - x_{(r,s,t)k} \right| \quad (12)$$

$$\text{MSE} = \frac{1}{k K_1 K_2 K_3} \sum_{r=1}^{K_1} \sum_{s=1}^{K_2} \sum_{t=1}^{K_3} \sum_{k=1}^{3} \left(o_{(r,s,t)k} - x_{(r,s,t)k} \right)^2 \quad (13)$$

where $\mathbf{o}_{(r,s,t)} = [o_{(r,s,t)1}, o_{(r,s,t)2}, o_{(r,s,t)3}]$ is the original RGB pixel, $\mathbf{x}_{(r,s,t)} = [x_{(r,s,t)1}, x_{(r,s,t)2}, x_{(r,s,t)3}]$ is the restored RGB pixel with (r, s, t) denoting the spatiotemporal position in a $K_1 \times K_2 \times K_3$ image sequence and k characterizing the color channel.

The perceptual similarity between the original and the demosaicked video sequence is quantified using the normalized color difference (NCD) criterion:

$$\text{NCD} = \frac{\sum_{r=1}^{K_1} \sum_{s=1}^{K_2} \sum_{t=1}^{K_3} \sqrt{\sum_{k=1}^{3} \left(\overline{o}_{(r,s,t)k} - \overline{x}_{(r,s,t)k} \right)^2}}{\sum_{r=1}^{K_1} \sum_{s=1}^{K_2} \sum_{t=1}^{K_3} \sqrt{\sum_{k=1}^{3} \left(\overline{o}_{(r,s,t)k} \right)^2}} \quad (14)$$

where $\overline{\mathbf{o}}_{(r,s,t)} = [\overline{o}_{(r,s,t)1}, \overline{o}_{(r,s,t)2}, \overline{o}_{(r,s,t)3}]$ and $\overline{\mathbf{x}}_{(r,s,t)} = [\overline{x}_{(r,s,t)1}, \overline{x}_{(r,s,t)2}, \overline{x}_{(r,s,t)3}]$ are the vectors representing the RGB vectors $\mathbf{o}_{(r,s,t)}$ and $\mathbf{x}_{(r,s,t)}$, respectively, in the CIE LUV color space.

In the experiments reported here, the proposed spatiotemporal video demosaicking solution was compared against the conventional BI demosaicking scheme [21] which is commonly considered as the cost-effective industry standard for single-sensor imaging applications. Please note that in the context of the video demosaicking, the BI scheme constitutes a spatial only processing solution.

(a)

(b)

(c)

(d)

Fig. 10. Cropped areas of: (left) the original frames, (middle) BI demosaicked outputs, and (right) the proposed demosaicked outputs. The results correspond to: (a) *Nature* **for** $t = 5$**, (b)** *Nature* **for** $t = 80$**, (c)** *Rafting* **for** $t = 20$**, (d)** *Rafting* **for** $t = 90$**.**
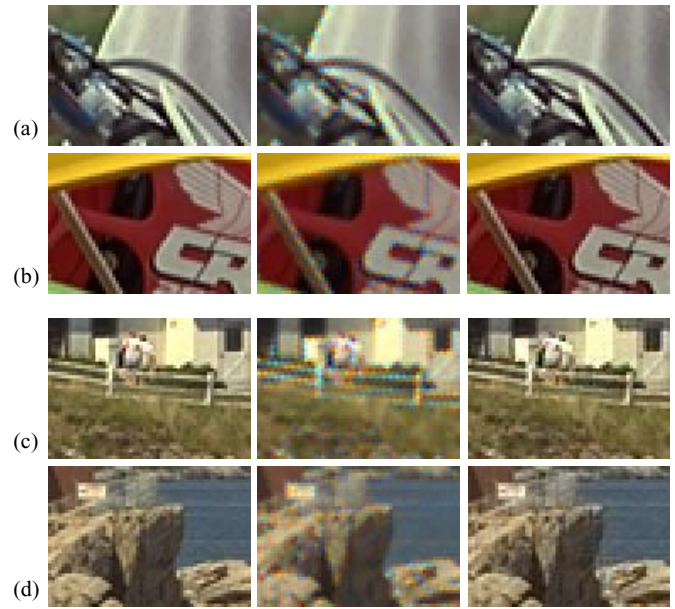

(a)

(b)

(c)

(d)

Fig. 11. Cropped areas of: (left) the original frames, (middle) BI demosaicked outputs, and (right) the proposed demosaicked outputs. The results correspond to: (a) *Bikes* **for** $t = 10$**, (b)** *Bikes* **for** $t = 90$**, (c)** *Coast* **for** $t = 30$**, (d)** *Coast* **for** $t = 95$**.**

TABLE I
OBJECTIVE EVALUATION OF THE VIDEO-DEMOSAICKING PERFORMANCE

Image sequence	Method	MAE	MSE	NCD
Nature	BI	7.920	248.3	0.1233
	Proposed	2.985	30.6	0.0543
Rafting	BI	5.734	143.6	0.0795
	Proposed	4.111	94.1	0.0637
Bikes	BI	6.263	163.1	0.1016
	Proposed	1.919	13.9	0.0400
Coast	BI	5.897	165.6	0.0902
	Proposed	2.587	24.8	0.0431

TABLE II
COMPUTATIONAL COMPLEXITY OF THE PROPOSED FRAMEWORK

Estimator / Criterion	ADs	SBs	MUs	DIs	AVs
(5) with (2)-(4)	35	18	6	25	6
(6) with (7)-(9) or (10),(3),(4)	23	18	6	13	6

Demosaicking results reported in Table I indicate that the proposed method is robust and significantly outperforms the industry standard with respect to all objective criteria used in this paper. Simple inspection of the numerical results and the corresponding video sequences (Fig. 8) shows the influence of extreme motion in the Rafting sequence on performance of the proposed spatiotemporal method. Therefore, additional improvements in performance are expected, as a part of our future work, to be obtained through the introduction of the motion compensation module into our framework.

Figs. 10-11 depict enlarged parts of the test video sequences cropped in areas with significant structural contents. The results allow for the visual comparison of the original images and demosaicked results. As it can be seen, the proposed spatiotemporal solution produces significantly higher visual quality compared to the results obtained using the (spatial only) BI industry standard. The BI scheme blurs structural information, introduces visually annoying zipper effects, and produces demosaicked video frames with color shifts and artifacts. However, this is not the case when the proposed spatiotemporal solution is used. Our solution excellently preserves edges and details, and produces the video frames with true coloration and the highest fidelity to the original video frames.

Apart from the actual performance of the proposed solution, its computational complexity is analyzed here in terms of normalized operations, such as additions (ADs), subtractions (SBs), multiplications (MUs), and absolute values (AVs). As it can be seen in Table II, by forming the spectral quantities through the predicted R or B components in (2) a G demosaicking step (5) requires to implement twelve extra ADs and DIs compared to R or B demosaicking steps using (6). The recorded values, as well as execution time measured using a conventional PC with a standard operating system and programming environment indicate that our solution is efficient. When implemented in software, the execution of the proposed video demosaicking tool on an Intel Pentium IV 2.40 GHz CPU, 512 MB RAM box with Windows XP operating system and MS Visual C^{++} 5.0 programming environment, took (on average) 0.291 seconds to demosaick a 300×200 CFA frame of video sequences.

In the summary, the following conclusions can be drawn: i) the utilization of spatiotemporal, spectral and structural characteristics is essential in video demosaicking, ii) the proposed video demosaicking solution produces visually pleasing, demosaicked, full-color videos, and iii) our solution is sufficiently robust and computationally efficient.

IV. CONCLUSION

A unique video demosaicking solution, suitable for single sensor (digital) still image and video cameras, was introduced. The proposed spatiotemporal approach utilizes unidirectional filtering windows and multistage processing concepts to follow the varying spatiotemporal characteristics of CFA image sequences. Through the utilization of adaptive weights and refined spatiotemporal spectral estimation, our solution recovers the full-color information of the captured video without introducing spectral, structural and motion artifacts. This results in visually pleasing color image sequences and performance which is significantly better, both objectively and subjectively, than that of the industry standard at a reasonable cost. The excellent design and performance characteristics make our solution ideal for video-capturing enabled, single-sensor consumer electronics.

REFERENCES

[1] R. Lukac, K. Martin, and K. N. Plataniotis, "Digital camera zooming based on unified CFA image processing steps," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp.15-24, February 2004.

[2] P. L. P. Dillon, D. M. Lewis, and F. G. Kaspar, "Color imaging system using a single CCD area array," *IEEE Journal of Solid-State Circuits*, vol. 13, no. 1, pp. 28-33, February 1978.

[3] T. Lule, S. Benthien, H. Keller, F. Mutze, P. Rieve, K. Seibel, M. Sommer, and M. Bohm, "Sensitivity of CMOS based imagers and scaling perspectives," *IEEE Transactions on Electron Devices*, vol. 47, no. 11, pp. 2110-2122, November 2000.

[4] K. Parulski and K. E. Spaulding, "Color image processing for digital cameras," in *Digital Color Imaging Handbook*, (eds.) G. Sharma, CRC Press, Boca Raton, FL., pp.728-757, 2002.

[5] B. E. Bayer, "Color imaging array," *U.S. Patent 3 971 065*, July 1976.

[6] W. Lu and Y. P. Tang, "Color filter array demosaicking: new method and performance measures," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1194-1210, October 2003.

[7] 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.

[8] X. Wu and N. Zhang, "Primary-consistant soft-decision color demosaicking for digital cameras," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1263-1274, September 2004.

[9] B. K. Gunturk, J. Glotzbach, Y. Altunbasak, R. W. Schaffer, and R. M. Murserau, "Demosaicking: color filter array interpolation," *IEEE Signal Processing Magazine*, vol. 22, no. 1, pp. 44-54, January 2005.

[10] F. J. Sanchez-Marin, Y. Srinivas, K. N. Jabri, and D. L. Wilson, "Quantitative image quality analysis of a nonlinear spatio-temporal filter," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 288-295, February 2001.

[11] K. J. Boo and N. K. Bose, "A motion-compensated spatio-temporal filter for image sequences with signal-dependent noise," *IEEE Trans. Circuits and Systems Video Technology*, vol. 8, no. 3, pp. 287-298, June 1998.

[12] R. Lukac and S. Marchevsky, "LUM smoother with smooth control for noisy image sequences," *EURASIP Journal of Applied Signal Processing*, vol. 2001, no. 2, pp. 110-120, June 2001.

[13] J. S. Kim and H. W. Park, "Adaptive 3-D median filtering for restoration of an image sequence corrupted by impulse noise," *Signal Processing: Image Communication*, vol. 16, no. 7, pp. 657-668, April 2001.

[14] G. R. Arce, "Multistage order statistic filters for image sequence processing," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1146-1163, May 1991.

[15] R. Lukac, and K. N. Plataniotis, "A vector spectral model for digital camera image processing," *IEEE Transactions on Circuits and Systems for Video Technology*, submitted.

[16] R. Lukac, and K. N. Plataniotis, "Normalized color-ratio modelling for CFA interpolation," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 2, pp. 737-745, May 2004.

[17] R. Lukac, K. Martin, and K. N. Plataniotis, "Demosaicked image post-processing using local color ratios," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 14, no. 6, pp. 914-920, June 2004.

[18] D. R. Cok, "Signal processing method and apparatus for producing interpolated chrominance values in a sampled color image signal," *US Patent* 4 642 678, February 1987.

[19] J. Adams, "Design of practical color filter array interpolation algorithms for digital cameras," *Proc. SPIE*, vol. 3028, pp. 117-125, Feb. 1997.

[20] R. Lukac, B. Smolka, K. Martin, K. N. Plataniotis, and A. N. Venetsanopulos, "Vector filtering for color imaging," *IEEE Signal Processing Magazine*; Special Issue on Color Image Processing, vol. 22, no. 1, pp. 74-86, January 2005.

[21] T. Sakamoto, C. Nakanishi, and T. Hase, Software pixel interpolation for digital still cameras suitable for a 32-bit MCU," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 4, pp. 1342-1352, November 1998.

Rastislav Lukac received the M.S. (Ing.) and Ph.D. degrees in Telecommunications from the Technical University of Kosice, Slovak Republic in 1998 and 2001, respectively. From February 2001 to August 2002 he was an Assistant Professor at the Department of Electronics and Multimedia Communications at the Technical University of Kosice. Since August 2002 he is a Researcher in Slovak Image Processing Center in Dobsina, Slovak Republic. From January 2003 to March 2003 he was a Postdoctoral Fellow at the Artificial Intelligence & Information Analysis Lab at the Aristotle University of Thessaloniki, Greece. Since May 2003 he has been a Post-doctoral Fellow with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering at the University of Toronto in Toronto, Canada. His research interests include digital camera image processing, microarray image processing, multimedia security, and nonlinear filtering and analysis techniques for color image & video processing.

Dr. Lukac is a Member of the IEEE Circuits and Systems, IEEE Consumer Electronics, and IEEE Signal Processing Societies. He serves as a Technical Reviewer for various scientific journals and he participates as a Member of numerous International Conference Committees. In 2003 he was awarded the NATO/NSERC Science Award.

Konstantinos N. Plataniotis received the B. Engineering degree in Computer Engineering from the Department of Computer Engineering and Informatics, University of Patras, Patras, Greece in 1988 and the M.S and Ph.D degrees in Electrical Engineering from the Florida Institute of Technology (Florida Tech), Melbourne, Florida in 1992 and 1994 respectively. From August 1997 to June 1999 he was an Assistant Professor with the School of Computer Science at Ryerson University. He is currently an Associate Professor at the Edward S. Rogers Sr. Department of Electrical & Computer Engineering where he researches and teaches adaptive systems and multimedia signal processing. He co-authored, with A.N. Venetsanopoulos, a book on "Color Image Processing & Applications", Springer Verlag, May 2000, he is a contributor to seven books, and he has published more than 300 papers in refereed journals and conference proceedings in the areas of multimedia signal processing, image processing, adaptive systems, communications systems and stochastic estimation.

Dr. Plataniotis is a Senior Member of IEEE, a past member of the IEEE Technical Committee on Neural Networks for Signal Processing. He was the Technical Co-Chair of the Canadian Conference on Electrical and Computer Engineering (CCECE) 2001, and CCECE 2004. He is the Vice-chair for the 2006 IEEE Intelligent Transportation Systems Conference, and the Technical Program Co-Chair of the 2006 IEEE International Conference in Multimedia and Expo (ICME 2006).