SPATIOTEMPORAL DEMOSAICKING USING MULTI-STAGE PROCESSING CONCEPTS

Rastislav Lukac and Konstantinos N. Plataniotis

Multimedia Laboratory - BA 4157, The Edward S. Rogers Sr. Department of ECE,
University of Toronto, 10 King's College Road, Toronto, Canada
lukacr@ieee.org, kostas@dsp.utoronto.ca

ABSTRACT

This paper presents a spatiotemporal demosaicking scheme suitable for digital video cameras. The proposed method uses multi-stage processing concepts to follow varying spatiotemporal characteristics of the captured video. The method also uses both the structural and spectral information during demosaicking in order to produce visually pleasing full-color videos.

1. INTRODUCTION

Consumer digital cameras use a single image sensor, usually charge-coupled device (CCD) or complementary metal oxide semiconductor (CMOS) sensor, to acquire the digital data representation of the true visual scene [1],[2]. To overcome a monochromatic nature of the sensor, the manufacturers place a color filter array (CFA) on top of the sensor. Since each sensor cell has its own spectrally selective filter, the acquired CFA data constitutes a mosaic-like gray-scale image [2]. The true color information is restored by estimating the missing color components at each spatial location of the acquired CFA image. The process, known as demosaicking [2]-[6], uses spectral interpolation to produce a full-color output [7],[8].

In this paper, the demosaicking concept used in digital still cameras is extended for demosaicking the captured video. Motion video represents a three-dimensional image signal or a time sequence of two-dimensional images [9]. Since such a visual input usually exhibits significant correlation in both the spatial and temporal sense, by omitting the essential temporal characteristics and processing separately the individual frames, the conventional demosaicking methods produce an output video sequence with motion artifacts [4].

To overcome the problem, the proposed video demosaicking method is a spatiotemporal processing solution which uses the spatial, temporal, spectral and structural characteristics of the captured CFA video to output a demosaicked full-color image sequence pleasing for viewing. The method can use either unidirectional or bidirectional processing windows [9] during demosaicking to overcome the mosaic-layout of CFA video data. By operating in individual directions, the method uses multi-stage processing concepts to perform cost-effective demosaicking operations. Thus, our method constitutes a practical real-time processing solution suitable for consumer single-sensor electronic devices, such as digital video cameras and video-enabled mobile phones and personal digital assistants.

2. PROPOSED METHOD

The video data acquired by the CFA-based sensor module constitutes a $K_1 \times K_2 \times K_3$ gray-scale image sequence $z : Z^3 \rightarrow Z$ with the pixels $z_{(r,s,t)}$. Each of K_3 frames $z(t) : Z^2 \rightarrow Z$ with dimensions of $K_1 \times K_2$ pixels has a mosaic-like structure due to CFA-driven acquisition of the visual scene. Occupying the spatial location (r, s), with the row coordinate $r = 1, 2, ..., K_1$ and the column coordinate $s = 1, 2, ..., K_2$, the value of $z_{(r,s,t)}$ acquired using the Bayer CFA [10] shown in *Fig.1* corresponds to the R component for (odd r, even s), the G component for (odd r, odd s) and (even r, even s), and the B component for (even r, odd s).

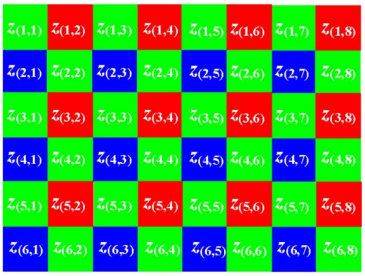

Fig. 1. Bayer CFA with the GRGR phase in the first row [10].

Following the spectral interpolation concept used during demosaicking [7],[8], each frame $z(t)$ is first transformed to a $K_1 \times K_2$ Red-Green-Blue (RGB) image $\mathbf{x}(t) : Z^2 \rightarrow Z^3$ with vectorial pixels $\mathbf{x}_{(r,s,t)} = [x_{(r,s,t)1}, x_{(r,s,t)2}, x_{(r,s,t)3}]$ where $x_{(r,s,t)k}$ denotes the intensity in the R ($k = 1$), G ($k = 2$), or B ($k = 3$) channel. This step produces vectors $\mathbf{x}_{(r,s,t)} = [z_{(r,s,t)}, 0, 0]$ for (odd r, even s), $\mathbf{x}_{(r,s,t)} = [0, 0, z_{(r,s,t)}]$ for (even r, odd s), and $\mathbf{x}_{(r,s,t)} = [0, z_{(r,s,t)}, 0]$ in all remaining spatial locations. Note that the values of the missing components $x_{(r,s,t)k}$ are equal to zeros to denote their portion to the coloration of $\mathbf{x}_{(r,s,t)}$. The missing components are estimated at each spatial location from the available neighboring components using image interpolation [2]-[6].

2.1. Demosaicking Using Unidirectional Windows

Due to the double number of G filters in the Bayer CFA, the proposed method first demosaicks the G color plane in $\mathbf{x}(t)$. To reduce, if not eliminate, the visual impairments such as color shifts and artifacts usually introduced during demosaicking [11], the method uses the essential spectral characteristics of the three consecutive video frames $\mathbf{x}(t + i)$ as follows:

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

where $i = -1$, 0 and -1 indicates the components from the past, actual and future frame, respectively. The values $x_{(\cdot,\cdot,t+i)2}$ denote the

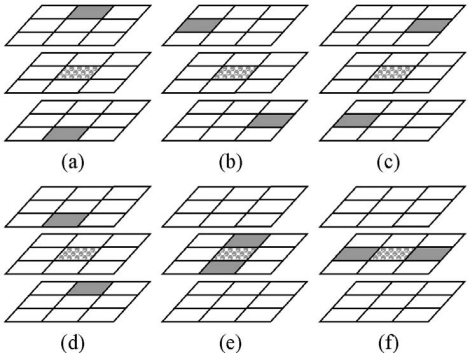

Fig. 2. Unidirectional processing windows: *Configuration 1*.

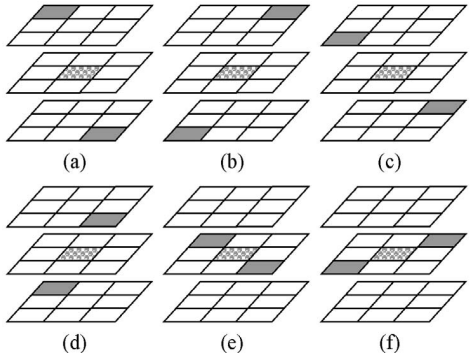

Fig. 3. Unidirectional processing windows: *Configuration 2*.

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 $k=1$ is used in (1). Otherwise, (r,s,t) corresponds to a B CFA location necessitating the use of (1) with $k=3$.

Using unidirectional processing windows (*Fig.2*) centered at the location (r,s,t) under consideration, the spectral quantities obtained in (1) are used to represent each of six possible directions as follows:

$$\begin{array}{l} \bar{c}_1 = (c_{(r-1,s,t-1)} + c_{(r+1,s,t+1)})/2 \\ \bar{c}_2 = (c_{(r,s-1,t-1)} + c_{(r,s+1,t+1)})/2 \\ \bar{c}_3 = (c_{(r,s+1,t-1)} + c_{(r,s-1,t+1)})/2 \\ \bar{c}_4 = (c_{(r+1,s,t-1)} + c_{(r-1,s,t+1)})/2 \\ \bar{c}_5 = (c_{(r-1,s,t)} + c_{(r+1,s,t)})/2 \\ \bar{c}_6 = (c_{(r,s-1,t)} + c_{(r,s+1,t)})/2 \end{array} \tag{2}$$

The suitability of the above values is evaluated with respect to the structural and temporal characteristics of the captured video using the adaptive weights:

$$\begin{array}{l} w_1 = 1/(1+|c_{(r-1,s,t-1)} - c_{(r+1,s,t+1)}|) \\ w_2 = 1/(1+|c_{(r,s-1,t-1)} - c_{(r,s+1,t+1)}|) \\ w_3 = 1/(1+|c_{(r,s+1,t-1)} - c_{(r,s-1,t+1)}|) \\ w_4 = 1/(1+|c_{(r+1,s,t-1)} - c_{(r-1,s,t+1)}|) \\ w_5 = 1/(1+|c_{(r-1,s,t)} - c_{(r+1,s,t)}|) \\ w_6 = 1/(1+|c_{(r,s-1,t)} - c_{(r,s+1,t)}|) \end{array} \tag{3}$$

where w_j is associated with $\bar{c}_j$, for $j=1,2,...,6$.

Large gradients in (3) indicate that the corresponding inputs are located across spatiotemporal edges. Using the data-adaptive concepts [2], the coefficients w_j are used to emphasize spectral inputs $\bar{c}_j$ which are not positioned across an edge and to direct the demosaicking process along natural spatiotemporal edges in the image sequence in order produce a sharply formed output video. Thus, the G component $x_{(r,s,t)2}$ is demosaicked as follows:

$$x_{(r,s,t)2} = x_{(r,s,t)k} + \sum_{j=1}^{N} \{w_j \bar{c}_j\} / \sum_{j=1}^{N} w_j \tag{4}$$

where $x_{(r,s,t)k}$ is the original R ($k=1$) or B ($k=3$) CFA component used to normalize the output of the weighted averaging operation over N color-difference values, i.e. $N=6$.

Since the demosaicking step in (4) fully populates the G plane of $\mathbf{x}(t)$, the similar unbiased spectral estimator can be used in demosaicking the R ($k=1$) and B ($k=3$) plane as follows:

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

Following the arrangement of the original R or B components shown in *Fig.3*, the procedure first recovers the R (and B) components at B (and R) CFA locations. This demosaicking step forms new arrangements depicted in *Fig.2*, however, it does not fully populate R and B planes. Therefore, the procedure is completed by recovering of the R and B components at G CFA locations.

Following the arrangements shown in *Fig.3*, the initial R and B demosaicking step uses the spectral quantities obtained as follows:

$$\begin{array}{l} c_{(r-1,s-1,t+i)} = x_{(r-1,s-1,t+i)k} - x_{(r-1,s-1,t+i)2} \\ c_{(r-1,s+1,t+i)} = x_{(r-1,s+1,t+i)k} - x_{(r-1,s+1,t+i)2} \\ c_{(r+1,s-1,t+i)} = x_{(r+1,s-1,t+i)k} - x_{(r+1,s-1,t+i)2} \\ c_{(r+1,s+1,t+i)} = x_{(r+1,s+1,t+i)k} - x_{(r+1,s+1,t+i)2} \end{array} \tag{6}$$

The above values are used to calculate the representatives

$$\begin{array}{l} \bar{c}_1 = (c_{(r-1,s-1,t-1)} + c_{(r+1,s+1,t+1)})/2 \\ \bar{c}_2 = (c_{(r-1,s+1,t-1)} + c_{(r+1,s-1,t+1)})/2 \\ \bar{c}_3 = (c_{(r+1,s-1,t-1)} + c_{(r-1,s+1,t+1)})/2 \\ \bar{c}_4 = (c_{(r+1,s+1,t-1)} + c_{(r-1,s-1,t+1)})/2 \\ \bar{c}_5 = (c_{(r-1,s-1,t)} + c_{(r+1,s+1,t)})/2 \\ \bar{c}_6 = (c_{(r-1,s+1,t)} + c_{(r+1,s-1,t)})/2 \end{array} \tag{7}$$

of the inputs spawned by unidirectional processing windows. The same spectral quantities in (6) are used to determine the adaptive weights as follows:

$$\begin{array}{l} w_1 = 1/(1+|c_{(r-1,s-1,t-1)} - c_{(r+1,s+1,t+1)}|) \\ w_2 = 1/(1+|c_{(r-1,s+1,t-1)} - c_{(r+1,s-1,t+1)}|) \\ w_3 = 1/(1+|c_{(r+1,s-1,t-1)} - c_{(r-1,s+1,t+1)}|) \\ w_4 = 1/(1+|c_{(r+1,s+1,t-1)} - c_{(r-1,s-1,t+1)}|) \\ w_5 = 1/(1+|c_{(r-1,s-1,t)} - c_{(r+1,s+1,t)}|) \\ w_6 = 1/(1+|c_{(r-1,s+1,t)} - c_{(r+1,s-1,t)}|) \end{array} \tag{8}$$

Inspection of (6)-(8) reveals that similarly to (4) the unbiased spectral interpolator in (5) performs spatiotemporal processing over three consecutive frames and uses six inputs, i.e. $N=6$.

Once the the demosaicking step (5) with the inputs obtained through (6)-(8) has been completed, the proposed solution concludes the demosaicking process by demosaicking the remaining R and B

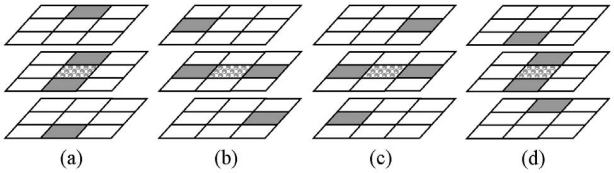

Fig. 4. Bidirectional processing windows: *Configuration 1*.

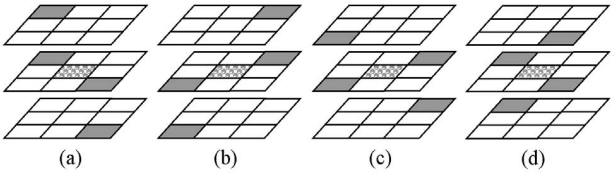

Fig. 5. Bidirectional processing windows: *Configuration 2*.

components using the arrangements shown in *Fig.2*. This demosaicking step uses (5) with (2) and (3) based on the following spectral quantities:

$$\begin{aligned} c_{(r-1,s,t+i)} &= x_{(r-1,s,t+i)k} - x_{(r-1,s,t+i)2} \\ c_{(r,s-1,t+i)} &= x_{(r,s-1,t+i)k} - x_{(r,s-1,t+i)2} \\ c_{(r,s+1,t+i)} &= x_{(r,s+1,t+i)k} - x_{(r,s+1,t+i)2} \\ c_{(r+1,s,t+i)} &= x_{(r+1,s,t+i)k} - x_{(r+1,s,t+i)2} \end{aligned} \quad (9)$$

The demosaicking step in (4) with (1)-(3) can be completed along the time axis t. Then, the procedure performs (5) with (6)-(8) for all the possible values of t. The video demosaicking process completes using (5) with (2), (3) and (9) for all values t. Note that such an implementation approach may be suitable for demosaicking the CFA video using a personal computer. An alternative variant, which is suitable for real-time processing [4], should temporally access five consecutive frames in order to use the original CFA values along with the fully-demosaicked G plane and the partially demosaicked R and B planes. Therefore shift registers have to be incorporated into the architecture to delay input and output frames.

2.2. Demosaicking Using Bidirectional Windows

Another proposed solution uses bidirectional processing windows (*Figs.4-5*) to enhance the performance of video-demosaicking. Following the configuration shown in *Fig.4*, the solution uses spectral quantities obtained in (1) to determine the interpolator's inputs

$$\begin{aligned} \bar{c}_1 &= \left(c_{(r-1,s,t-1)}+c_{(r+1,s,t+1)}+c_{(r-1,s,t)}+c_{(r+1,s,t)}\right)/4 \\ \bar{c}_2 &= \left(c_{(r,s-1,t-1)}+c_{(r,s+1,t+1)}+c_{(r,s-1,t)}+c_{(r,s+1,t)}\right)/4 \\ \bar{c}_3 &= \left(c_{(r,s+1,t-1)}+c_{(r,s-1,t+1)}+c_{(r,s-1,t)}+c_{(r,s+1,t)}\right)/4 \\ \bar{c}_4 &= \left(c_{(r+1,s,t-1)}+c_{(r-1,s,t+1)}+c_{(r-1,s,t)}+c_{(r+1,s,t)}\right)/4 \end{aligned} \quad (10)$$

as well as the corresponding weights obtained as follows:

$$\begin{aligned} w_1 &= 1/\left(1+\left|c_{(r-1,s,t-1)}-c_{(r+1,s,t+1)}\right|+\left|c_{(r-1,s,t)}-c_{(r+1,s,t)}\right|\right) \\ w_2 &= 1/\left(1+\left|c_{(r,s-1,t-1)}-c_{(r,s+1,t+1)}\right|+\left|c_{(r,s-1,t)}-c_{(r,s+1,t)}\right|\right) \\ w_3 &= 1/\left(1+\left|c_{(r,s+1,t-1)}-c_{(r,s-1,t+1)}\right|+\left|c_{(r,s-1,t)}-c_{(r,s+1,t)}\right|\right) \\ w_4 &= 1/\left(1+\left|c_{(r+1,s,t-1)}-c_{(r-1,s,t+1)}\right|+\left|c_{(r-1,s,t)}-c_{(r+1,s,t)}\right|\right) \end{aligned} \quad (11)$$

Using the values obtained in (1), (10) and (11), the missing G component $x_{(r,s,t)2}$ is produced through (4) with $N = 4$.

After the G color plane is fully populated, the proposed solution continues by demosaicking R and B color planes using the bidirectional windows shown in *Fig.5* and the spectral interpolator defined in (5) with $N = 4$. To determine both the inputs and weights in (5), the values obtained in (6) are used as follows:

$$\begin{aligned} \bar{c}_1 &= \left(c_{(r-1,s-1,t-1)}+c_{(r+1,s+1,t+1)}+c_{(r-1,s-1,t)}+c_{(r+1,s+1,t)}\right)/4 \\ \bar{c}_2 &= \left(c_{(r-1,s+1,t-1)}+c_{(r+1,s-1,t+1)}+c_{(r-1,s+1,t)}+c_{(r+1,s-1,t)}\right)/4 \\ \bar{c}_3 &= \left(c_{(r+1,s-1,t-1)}+c_{(r-1,s+1,t+1)}+c_{(r-1,s+1,t)}+c_{(r+1,s-1,t)}\right)/4 \\ \bar{c}_4 &= \left(c_{(r+1,s+1,t-1)}+c_{(r-1,s-1,t+1)}+c_{(r-1,s-1,t)}+c_{(r+1,s+1,t)}\right)/4 \end{aligned} \quad (12)$$

Fig. 6. $300 \times 200 \times 99$ test color video sequences: (a) Nature, (b) Bikes, (c) Coast.

$$\begin{aligned} w_1 &= 1/\left(1+\left|c_{(r-1,s-1,t-1)}-c_{(r+1,s+1,t+1)}\right|+\left|c_{(r-1,s-1,t)}-c_{(r+1,s+1,t)}\right|\right) \\ w_2 &= 1/\left(1+\left|c_{(r-1,s+1,t-1)}-c_{(r+1,s-1,t+1)}\right|+\left|c_{(r-1,s+1,t)}-c_{(r+1,s-1,t)}\right|\right) \\ w_3 &= 1/\left(1+\left|c_{(r+1,s-1,t-1)}-c_{(r-1,s+1,t+1)}\right|+\left|c_{(r-1,s+1,t)}-c_{(r+1,s-1,t)}\right|\right) \\ w_4 &= 1/\left(1+\left|c_{(r+1,s+1,t-1)}-c_{(r-1,s-1,t+1)}\right|+\left|c_{(r-1,s-1,t)}-c_{(r+1,s+1,t)}\right|\right) \end{aligned} \quad (13)$$

Since the processing step in (5) with (6), (12) and (13) cannot demosaick all the missing R ($k = 1$) and B ($k = 3$) components $x_{(r,s,t)k}$, the proposed method completes the demosaicking process by demosaicking the remaining R and B locations. Following the arrangements shown in *Fig.4*, the method uses (5) with (9), (12) and (13) to fully populate the R and B planes. Similarly to the previous proposed solution, the video-demosaicking process based on bidirectional processing windows can be implemented in two different ways, however, both implementation approaches produce the same demosaicked full-color video.

3. EXPERIMENTAL RESULTS

To determine the performance of the proposed video demosaicking method, test image sequences shown in *Fig.6* have been utilized. These test videos, which vary in motion, color appearance, and in the complexity of the structural content have been processed using the Bayer CFA shown in *Fig.1* to obtain a CFA image sequence z with pixels $z_{(r,s,t)}$. The demosaicked image sequence $\mathbf{x}$ is generated by applying the video-demosaicking solution to the CFA image sequence z. To facilitate the objective comparisons, 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 to evaluate the difference (or closeness) of the original (test) full-color image sequences to the demosaicked videos. The reader can find the definitions of these criteria in our previous works [2],[4],[7].

Table 1 and *Fig.7* summarize the achieved results. As it can be seen, the proposed solutions are robust, excellently preserve edges, details, and restore the color information without noticeable side effects. Although both proposed solutions produce high-qualities demosaicked videos, the use of bidirectional processing windows results in the best performance among the tested solutions.

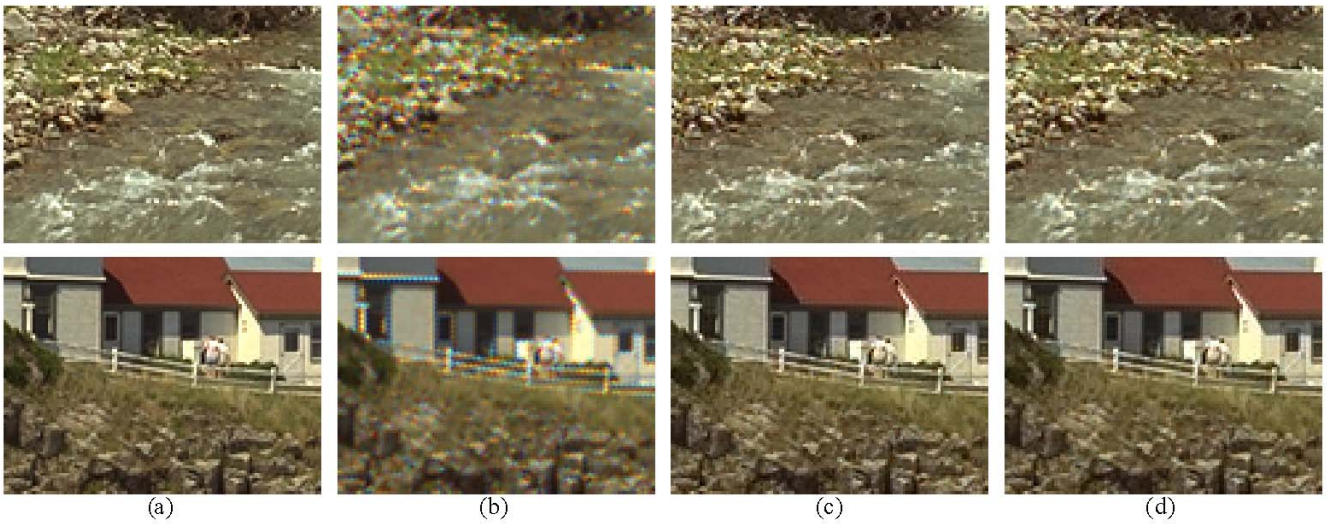

Fig. 7. Achieved results corresponding to the test video sequence Nature (top) and Coast (bottom): (a) original frame, (b) bilinear spatial demosaicking, (c) unidirectional spatiotemporal demosaicking, (d) bidirectional spatiotemporal demosaicking.

Table 1. Performance comparison of the various video demosaicking methods.

Video Sequence	Nature (*Fig.6a*)			Bikes (*Fig.6b*)			Coast (*Fig.6c*)		
CFA / Method	MAE	MSE	NCD	MAE	MSE	NCD	MAE	MSE	NCD
Bilinear	10.238	355.1	0.1448	6.498	174.5	0.1103	6.640	177.5	0.0994
Unidirectional	3.369	37.4	0.0545	1.986	14.7	0.0439	2.660	24.6	0.0439
Bidirectional	3.035	28.7	0.0503	1.912	13.4	0.0424	2.519	21.4	0.0421

4. CONCLUSION

A spatiotemporal demosaicking method suitable for restoring the color information in video captured using single-sensor digital video cameras was introduced. The proposed method uses multistage processing concepts defined over the samples inside unidirectional or bidirectional supporting windows along with an adaptive spectral interpolator to recover the full-color information of the captured video. Thus, the method avoids the visual impairments, such as various spectral, structural and motion artifacts, and produces visually pleasing demosaicked full-color image sequences.

5. REFERENCES

[1] 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.

[2] R. Lukac and K. N. Plataniotis, "Data-adaptive filters for demosaicking: A framework," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 2, pp. 560-570, May 2005.

[3] 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.

[4] R. Lukac and K. N. Plataniotis, "Fast video demosaicking solution for mobile phone imaging applications. *IEEE Transactions on Consumer Electronics*, vol. 51, no. 2, pp. 675-681, May 2005.

[5] 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.

[6] D. D. Muresan and T. W. Parks, "Demosaicing using optimal recovery," *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp. 267-278, February 2005.

[7] R. Lukac, K. N. Plataniotis, and D. Hatzinakos, "Color image zooming on the Bayer pattern," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 15, no. 11, November 2005.

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

[9] 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.

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

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