Efficient Rolling Shutter Distortion Removal using Hierarchical Block-based Motion Estimation

Donggeun Lee[1] and Kang-Sun Choi[2]

[1] School of Electrical, Electronics, and Commun. Eng., KOREATECH / Cheonan, Rep. of Korea ssinger1@koreatech.ac.kr
[2] Interdisciplinary Program in Creative Engineering, KOREATECH / Cheonan, Rep. of Korea ks.choi@koreatech.ac.kr

***** Corresponding Author: Kang-Sun Choi

Received December 20, 2013; Revised February 27, 2013; Accepted May 12, 2014; Published August 31, 2014

* *Extended from a Conference: Preliminary results of this paper were presented at the IEEE ISCE 2014. This present paper has been accepted by the editorial board through the regular reviewing process that confirms the original contribution.*

***Abstract*:** This paper reports an efficient algorithm for accurate rolling shutter distortion removal. A hierarchical global motion estimation approach for a group of blocks reduces the level of computation by three orders of magnitude. In addition, the motion of each scanline is determined accurately by averaging two candidates obtained through cubic spline interpolation. The experimental results show that the proposed method produces accurate motion information with significant computation reduction and corrects the rolling shutter distortion effectively.

***Keywords*:** Rolling shutter distortion, Hierarchical motion estimation, Motion interpolation

1. Introduction

The demand for image sensors is growing rapidly due to the increasing popularity of digital video cameras in surveillance, personal communication and consumer electronics. Although the charge-coupled device (CCD) has been the dominant technology for image sensors, the complementary metal-oxide semiconductor (CMOS) technology has become a popular alternative since the early 1990s because it was quickly realized that adding an amplifier to each pixel increases the sensor speed significantly and improves its signal-to-noise ratio (SNR), thereby overcoming the shortcomings of the passive pixel sensor. The massively parallel conversion and digital readout provide very high speed readout, enabling new applications, such as wider dynamic range imaging. As a result, CMOS sensors can directly integrate with image processing providing low cost, low power solutions for various applications.

On the other hand, due to the inherent sequential-readout nature of CMOS sensor array, each row (scan-line) of the sensor array is exposed for slightly different times, as shown in Fig. 1. Each scanline is assumed to be sampled at a specific instant in the exposure time. The interval between the sampling instants of two successive scanlines is defined as the unit time. Such a phenomenon is the so-called *rolling shutter* that induces geometric distortion to images if the video camera or the object moves during image acquisition. Fig. 2 gives an example of the rolling shutter effect, where all the buildings appear slanted to the

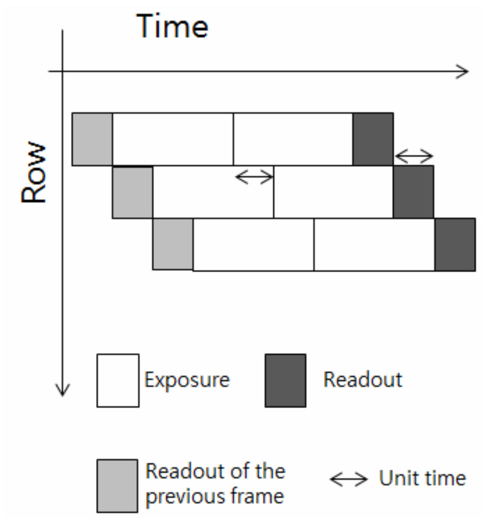


Fig. 1. Timing diagram of the CMOS sensor array. Each scanline was sampled at a specific instant in the exposure time. The interval between the sampling instants of two successive scanlines is defined as the unit time.

Fig. 2. Illustration of the rolling shutter effect. All the buildings appear slanted as the camera pans across the scene.

right as the camera pans to the left across the scene. Such undesirable effects can seriously degrade the visual quality of the image.

The extent of the rolling shutter effect depends on the exposure time and readout time. When the imaging chip has a standalone frame buffer, the readout process simply transfers the pixel data to the frame buffer. This operation can be performed in a few clock cycles, allowing the total readout time per frame to be less than a few thousand cycles (in the order of milliseconds for a typical imaging chip) [1, 2]. The readout time in this case is much smaller than the exposure time. Therefore, there is little rolling shutter effect unless the object moves extremely fast (more than one pixel in a single readout operation).

On the other hand, in the case that there is no on-chip frame buffer, which is a common design choice to save cost, the rolling shutter has a profound effect on image quality. Webcams, for example, normally have little or no local buffer and need to transfer the recorded image data to the host computer, where the image is processed, via a USB or Firewire. The readout time of such devices is limited by the transmission rate. In this case, the readout time is comparable to the exposure time and may dominate the entire imaging process, making the rolling shutter effect profoundly noticeable.

The rolling shutter effect can be prevented using a mechanical shutter. Alternatively, one can design a new circuit with a local sample-and-hold for each photo sensor so that the entire sensor array experiences a single exposure [3, 4]. On the other hand, this method reduces the fill factor of the photo sensor and degrades the efficiency of photo collection. As a result, the readout rate is still limited by the transmission rate.

The image deformation caused by the rolling shutter is corrected by estimating the camera motion when each scan-line is exposed, and by compensating for that motion. In [5], the motion for each scan-line was estimated using an accelerometer mounted rigidly on the camera. This approach, however, can be utilized only for the systems with an accelerometer, and its correction result depends on the accuracy of the sensor.

Approaches that address the rolling shutter effect by image analysis have been proposed. Geyer *et al*. [6] derived a time-dependent perspective model for the rolling-shutter camera by assuming that the perspective transformation of each scanline is a function of the relative location and velocity between the object and the camera. The condition under which the rolling shutter effect becomes negligible was also derived. Ait-Aider *et al*. applied the model further to estimate the camera parameters and the object dynamics, but the shape of the object and the correspondence of the object between frames must be known *a priori* [7, 8]. Wilburn *et al*. [9] suggested that the rolling shutter effect is not always undesirable and constructed a 2-D array of rolling-shutter cameras to achieve high-speed imaging. The final image was constructed by stitching the scanlines acquired by different cameras at the same exposure time. Assuming transitional and constant object motion, Sasaki and Nara [8] used several images acquired using a rolling-shutter camera to generate a single clean image.

In [10, 11], the rolling shutter distortion was corrected using only captured images by means of image processing. A global motion vector (GMV) between two successive image frames was first estimated. The line motion vector for each scan-line (LMV) was determined by smoothly interpolating the GMVs estimated from several successive frames.

On the other hand, the conventional block-based rolling shutter distortion compensation method requires a huge amount of computation for block motion estimation. In addition, the interpolated LMVs can abruptly change between two successive frames.

To address such problems, an efficient and accurate rolling shutter distortion removal method was proposed. In the proposed method, the GMV is obtained efficiently using a hierarchical searching approach that reduces the computational burden significantly, maintaining the accuracy of the obtained GMV.

The remainder of this paper is organized as follows. The next section describes a related work that the proposed algorithm is based upon. The proposed scheme section describes a hierarchical global motion vector estimation method that reduces the computation significantly. A scanline motion estimation producing more accurate line motion is then presented. The experimental results include performance evaluations and a discussion of the effects of the proposed scheme. The paper is concluded in the last section.

2. Related Work

This section presents a conventional rolling shutter removal method employing digital image processing [10]. Fig. 3 illustrates a block diagram of the conventional method. As mentioned before, the conventional method consists of three parts; GMV estimation, LMV estimation and rolling shutter distortion removal.

In the GMV estimation, the n^{th} image frame, f_n, is first divided into non-overlapped blocks and the motion vector (MV) of each block is searched for using the sum of the absolute difference (SAD) as the error criterion [10]. To

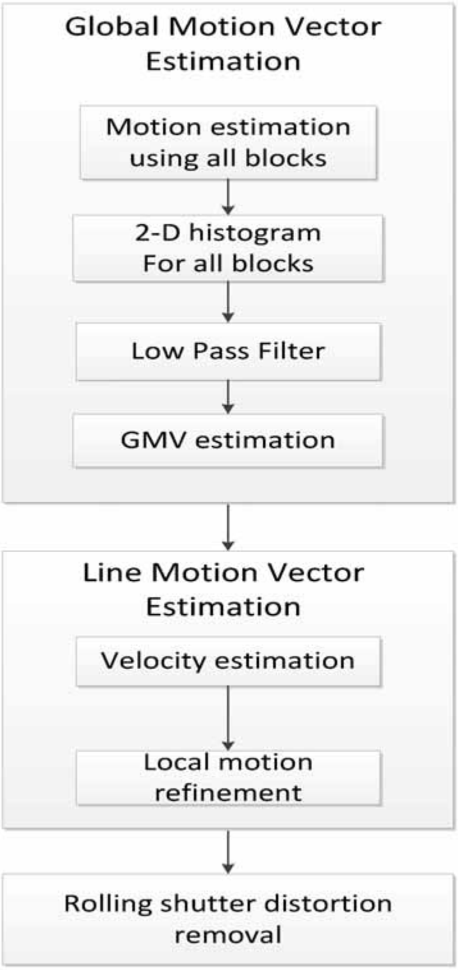


Fig. 3. Block diagram of the conventional method [10].

obtain a GMV between two successive image frames, two histograms are generated using each directional component of all the block motion vectors (BMV). The majority bin in each histogram becomes the corresponding directional component of the GMV. To improve the accuracy, the noise in the histograms is alleviated prior to the majority selection.

After the GMV is determined for f_n, the LMV for each scanline is determined using the cubic spline interpolation with four GMVs obtained from the n-2th frame to the n+1th frame. Fig. 4 shows the LMV estimation. Let $\mathbf{g}(n)$, $\mathbf{v}(t)$ and T denote the GMV obtained for f_n, the camera motion at time, t, and the frame interval, respectively. $\mathbf{g}(n)$, which is represented by the blue arrows in Fig. 4, is then approximated by

$$\mathbf{g}(n) \cong \int_{(n-0.5)T}^{(n+0.5)T} \mathbf{v}(t)dt\ . \tag{1}$$

Therefore, the initial estimate of the velocity of the middle scanline can be obtained as follows:

$$\mathbf{v}((n+0.5)T) = \mathbf{g}(n)/T\ . \tag{2}$$

To estimate the LMV of each scanline, a Bezier curve

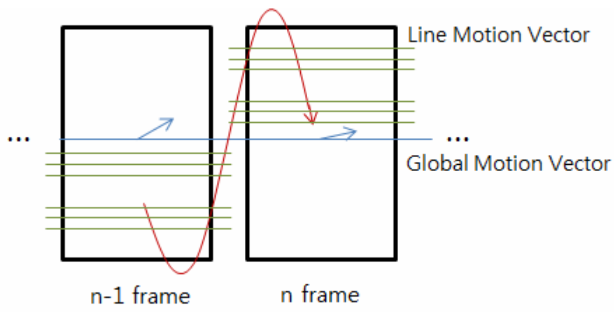


Fig. 4. Line Motion Vector and Global Motion Vector in frames.

$\mathbf{q}(\lambda)=(q_x(\lambda),q_y(\lambda))$ was utilized, where the x component of $\mathbf{q}(\lambda)$ is described by

$$q_x(\lambda) = (1-\lambda)^3 c_1 + 3\lambda(1-\lambda)^2 c_2 + 3\lambda^2(1-\lambda)c_3 + 3\lambda^3 c_4 \tag{3}$$

The following four nearby velocity samples can be obtained;

$$\begin{aligned} q_x(0) &= v_x((n-1.5)T) \\ q_x(1/3) &= v_x((n-0.5)T) \\ q_x(2/3) &= v_x((n+0.5)T) \\ q_x(1) &= v_x((n+1.5)T). \end{aligned} \tag{4}$$

After calculating the control points, c_1, c_2, c_3, and c_4, the velocity of the i^{th} scanline in the n^{th} image is calculated using the following equation:

$$\mathbf{v}(nT+i) = \mathbf{q}\left(\frac{(i/T)+1.5}{3}\right). \tag{5}$$

Finally, rolling shutter effect removal was performed by re-aligning the scanlines. The corrected point, $\mathbf{p}$, corresponding to the point, $\mathbf{s}$, at the s_y^{th} scanline in the distorted image can be obtained by

$$\mathbf{p} = \mathbf{s} - \int_{(n+0.5)T}^{nT+s_y} \mathbf{v}(t)dt\ . \tag{6}$$

3. The Proposed Algorithm

Fig. 5 presents a block diagram of the proposed method. Although the overall pipeline is similar to that of the method [10], both the computational efficiency and accuracy are improved in the proposed algorithm as follows. First, hierarchical GMV estimation of the current frame is performed, which consists of a motion estimation (ME) for a group of blocks and a motion refinement of each block to improve the motion accuracy. After motion refinement, $\mathbf{g}(n)$ is estimated by averaging the MVs of all the blocks. In the proposed method, the LMV for each scan line of the current frame is estimated using a 5-point cubic spline interpolation with the GMVs of five

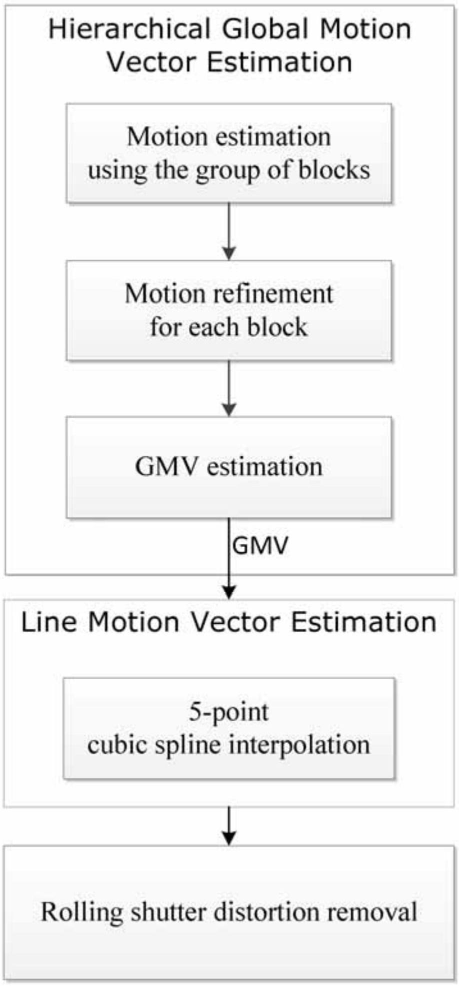


Fig. 5. Block diagram of the proposed method.

successive frames. Finally, rolling shutter distortion is removed using the LMVs obtained, as in the method reported elsewhere [10].

3.1 Hierarchical Global Motion Vector Estimation

Fig. 6 shows the configuration of the proposed hierarchical GMV estimation. In the proposed method, a group of sixteen equally separated blocks is used to estimate the GMV, instead of all the non-overlapped blocks. In contrast to conventional block matching, where a motion estimation for each block is performed independently, block matching for all sixteen blocks is performed simultaneously, i.e. an MV yielding the minimum sum of SADs of the group of blocks is determined to be the initial GMV in the proposed method.

This approach reduces the computation for the GMV estimation significantly. In addition, the approach produces a reliable result because the aperture problem can be avoided.

On the other hand, the computational burden can still be huge if the search range for block matching is too large. To alleviate the computation for GMV estimation further, the sum of SADs of the block group can be obtained at

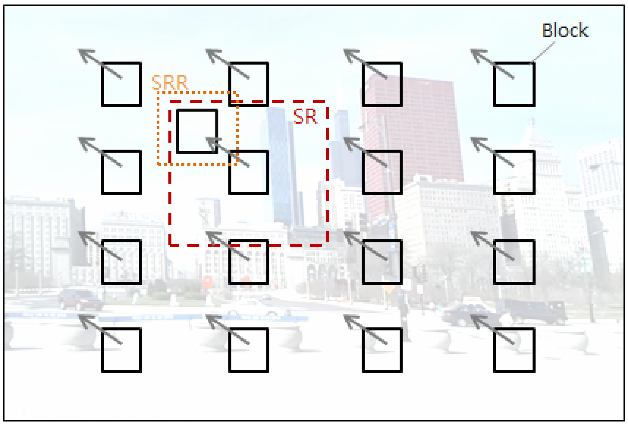


Fig. 6. Configuration for the proposed global motion vector estimation.

every 8th position over the initial search range (SR), which is depicted as a red dashed box in Fig. 6.

After the initial GMV $\mathbf{g}(n)$ is obtained, the MV for each block is refined independently by searching for the optimal position over the much smaller search range for refinement (SRR) than SR, which is illustrated as an orange box in Fig. 5. In the motion refinement, every position within the SRR is traversed to improve the accuracy of each MV. A final GMV is obtained by averaging the sixteen MVs.

3.2 Line Motion Vector Estimation

As $\mathbf{g}(n)$ represents motion of the center row, after $\mathbf{g}(n)$ is determined, the LMVs of other rows need to be found. In the proposed method, the LMV for each scan-line is obtained using the cubic spline interpolation with five GMVs obtained from the n-2th frame to the n+2th frame. To remove the LMV discontinuity between the LMV of the last line of the previous frame and that of the first line of the current frame, the LMVs are calculated twice. In particular, each LMV is calculated with four GMVs of f_i, $n-2 \leq i \leq n+1$. Each LMV is calculated again with four GMVs of f_i, $n-1 \leq i \leq n+2$. Two estimated LMVs for each scan-line is averaged to obtain the final LMV.

4. Experimental Results

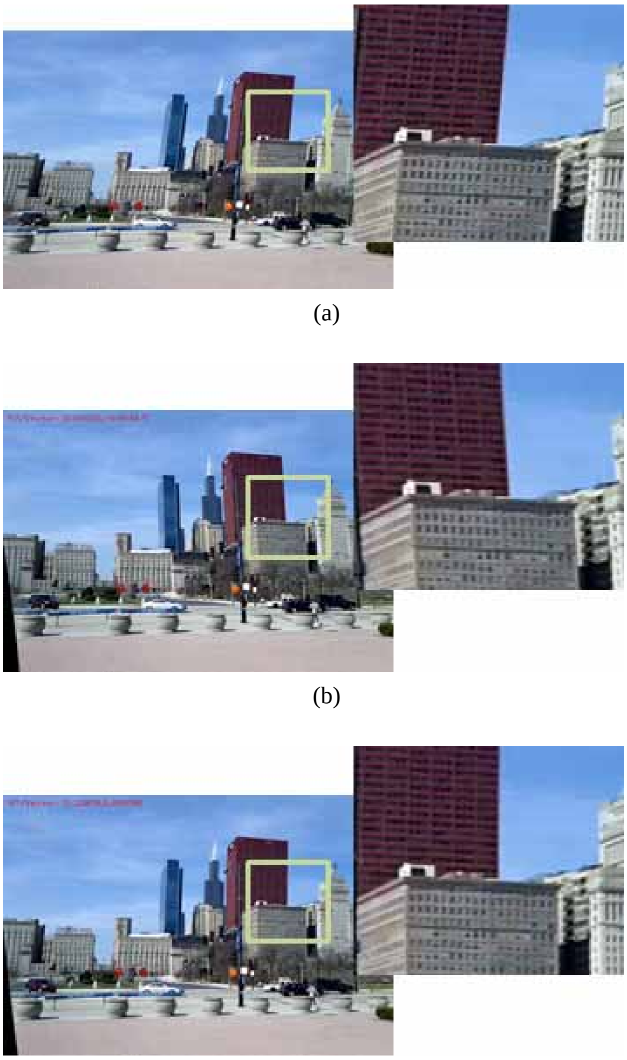
To compare the performance of the rolling shutter distortion removal, the method reported elsewhere [10] was also tested. In the experiments, SR, SRR and the block size were set to ±128, ±16 and 32×32 pixels, respectively, for the test video, the size of which as 720×480.

Table 1 lists the computational load reduction of the proposed method with respect to the number of SAD operations. The hierarchical approach drastically reduces the amount of computation for the GMV estimation by three orders of magnitude.

In the conventional method [10], ME was performed for all non-overlapping blocks independently to obtain the GMV. For the test video, there was approximately 330

Table 1. Computational performance comparison of the conventional method [2] and the proposed method.

	Method [10]	Proposed
Number of SAD operations per frame	21,796,170	**34,848**
Computation time (sec)	407	**0.45**



(a)

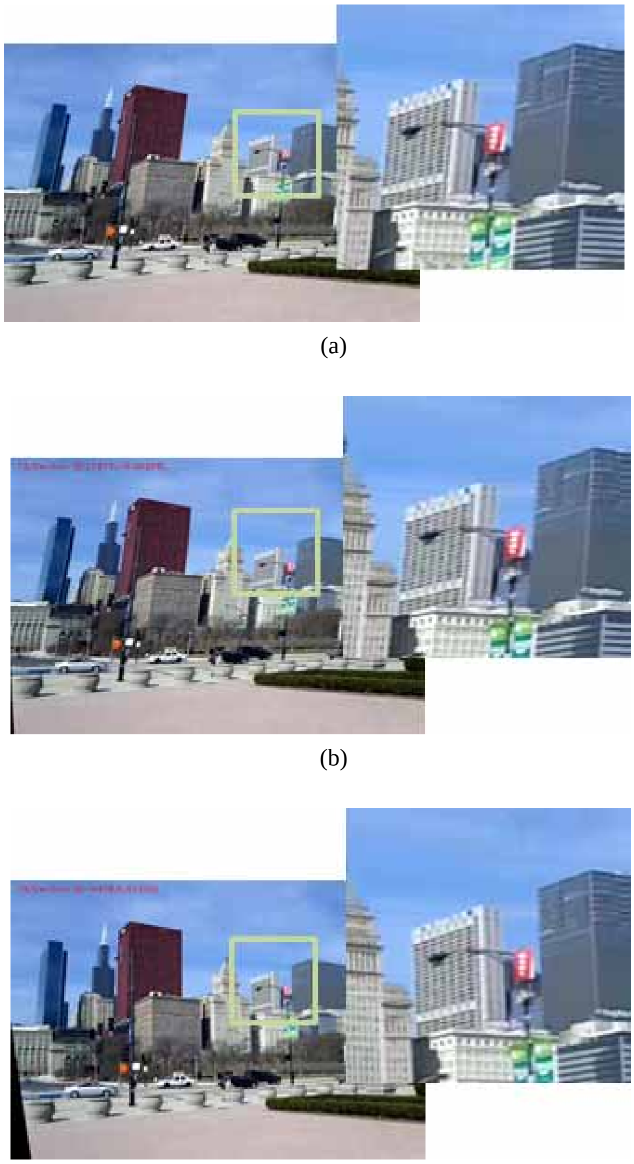
(b)

(c)

Fig. 7. Subjective comparison of tolling shutter distortion removal (a) Original image, (b) Corrected result of the conventional method [10], (c) Corrected result of the proposed method.

blocks. For each block, the SAD operation was executed 66,049 (257×257) times. In addition, the conventional method required additional operations for the 2D-vector histogram and low pass filtering.

In the proposed method, however, only sixteen blocks were considered. Furthermore, because the initial GMV is estimated over sparse MV candidates at every 8^{th} position within the SR, the SAD operation was performed only 1,089 (33×33) times for each block. During the motion refinement, the same number of SAD operations was performed over the SRR. As a result, the total SAD



(a)

(b)

(c)

Fig. 8. Subjective comparison of tolling shutter distortion removal (a) Original image, (b) Corrected result of the conventional method [10], (c) Corrected result of the proposed method.

operation was performed 34,848 (16×33×33×2) times in the proposed method.

Fig. 7 shows the resulting images corrected by both methods. In the captured image, the buildings were distorted severely by the rolling shutter effect. In Fig. 7(b), the conventional method [10] failed to correct the image by over-compensating for the rolling shutter effect, while the proposed method successfully removed the geometric distortion, as shown in Fig. 7(c).

Fig. 8 also shows other resulting images. In Fig. 8(b), the results of the conventional method were corrected by under-compensating for the rolling shutter effect. The rolling shutter distortion still remained, whereas the proposed method corrected the distortion successfully by straightening the street lamp and the buildings, as shown in

(a)

(b)

(c)

Fig. 9. Subjective comparison of tolling shutter distortion removal (a) Original image, (b) Corrected result of the conventional method [10], (c) Corrected result of the proposed method.

Fig. 8(c).

In Fig. 9(b), because of the inaccurate LMV, the street lamp was bent when the conventional method was used, whereas the pillar in Fig. 9(c) appeared better than that in Fig. 9(b). The proposed method compensates for the rolling shutter distortion error of the LMV and GMV.

This successful results stems not only from the high accuracy of the proposed GMV estimation method, but also the proposed LMV calculation using a five-point cubic spline interpolation. In contrast to the four-point interpolation used in the method [10], the interpolation in the proposed method produces smoothly-connected LMVs.

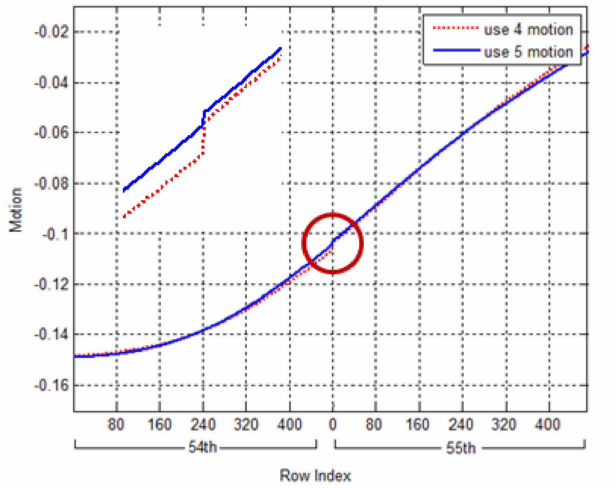


Fig. 10. Graphs of the horizontal displacement for all the horizontal lines of the 54^{th} and 55^{th} frames obtained using the conventional method [10] and the proposed method.

Fig. 10 shows the better smoothness of the proposed LMV estimation by plotting the horizontal displacement of the estimated LMVs for all the scanlines of the 54^{th} and 55^{th} frames. Although the four-point interpolation also produced smoothly changing LMVs within a frame, there was some discontinuity between the displacement of the last scanline of the 54^{th} frame and that of the first line of the 55^{th} frame. On the other hand, such discontinuity was alleviated clearly by the proposed LMV estimation.

5. Conclusions

This paper presented an efficient global motion vector estimation method for rolling shutter distortion removal. The proposed method ran more than 900 times faster than the conventional method in execution time. Despite this, the velocity of each scanline was obtained more accurately using the hierarchical GMV estimation and five-point cubic spline interpolation. As a result, the rolling shutter distortion can be removed effectively, producing natural videos.

References

[1] A. El Gamal and H. Eltoukhy, "CMOS image sensors," *IEEE Circuits and Devices Mag.*, vol. 21, no. 3, pp. 6–20, Mar. 2005. Article(CrossRef Link)

[2] Z. Zhou, B. Pain, and E. R. Fossum, "Frame-transfer CMOS active pixel sensor with pixel binning," *IEEE Trans. Electron Devices*, vol. 44, no. 10, pp. 1764–1768, Oct. 1997. Article(CrossRef Link)

[3] S. Lauxtermann, P. Schwider, P. Seitz, H. Bloss, J. Ernst, and H. Firla,"A high speed CMOS imager acquiring 5000 frames/sec,," in *Proc. Int. Electron Devices Meeting, Tech. Dig.*, Washington, DC, 1999, pp. 875–878. Article(CrossRef Link)

[4] M. Wäny and G. Paul, “CMOS image sensor with NMOS-only global shutter and enhanced responsivity,” *IEEE Trans. Electron Devices*, vol. 50, no. 1, pp. 57–62, Jan. 2003. Article(CrossRef Link)

[5] C. Jia, L.E. Brian, “Probabilistic 3-D motion estimation for rolling shutter video rectification from visual and inertial measurements,” *IEEE International Workshop on Multimedia Signal Processing*, pp. 203-208, Sep. 2012. Article(CrossRef Link)

[6] C. Geyer, M. Meingast, and S. Sastry, “Geometric models of rollingshutter cameras,” in *Proc. Omnidirectional Vision, Camera Networks and Non-classical Cameras*, Oct. 2005, pp. 12–19. Article(CrossRef Link)

[7] O. Ait-Aider, N. Andreff, J. M. Lavest, and P. Martinet, “Exploiting rolling shutter distortions for simultaneous object pose and velocity computation using a single view,” in *Proc. 4th IEEE Int. Conf. Computer Vision Systems*, Jan. 2006, pp. 35–41. Article(CrossRef Link)

[8] O. Ait-Aider and A. Bartoli, “Kinematics from lines in a single rolling shutter image,” in *Proc. Conf. Computer Vision and Pattern Recognition*, Minneapolis, MN, Jun. 2007, pp. 1–6. Article(CrossRef Link)

[9] G. Sasaki and Y. Nara, “Image processor and camera system,” U.S. Patent Application Publication, No. 2007/0120997 A1, May 31, 2007. Article(CrossRef Link)

[10] C.-K. Liang, L.-W. Chang, and H H. Chen, “Analysis and compensation of rolling shutter effect,” *IEEE Trans. on Image Processing*, vol.17, pp. 1323-1330, Aug. 2008. Article(CrossRef Link)

[11] Y. J. Koh, J.-H. Kim, and C.-S. Kim, “Removing rolling shutter distortion using the block-based motion interpolation,” *IEEK Conf. on Electronics*, pp. 451-452, Oct. 2013. Article(CrossRef Link)

Donggeun Lee is an undergraduate student at Korea University of Technology and Education (KOREATECH) where he is currently pursuing his B.S. His research interests include image processing and object tracking for UAV.

Kang-Sun Choi received his Ph.D. degree in nonlinear filter design in 2003, M.S. in 1999, and B.S. in 1997, in electronic engineering from Korea University. In 2011, he joined the School of Electrical, Electronics & Communication Engineering at Korea University of Technology and Education, where he is currently an Assistant Professor. From 2008 to 2010, he was a Research Professor of the Department of Electronic Engineering at Korea University. From 2005 to 2008, he worked in Samsung Electronics, Korea, as a Senior Software Engineer. From 2003 to 2005, he was a visiting scholar at the University of Southern California. His research interests are in the areas of multimedia compression, video processing, and computational photography. He is the recipient of an IEEE International Conference on Consumer Electronics Special Merit Award (2012).