DIGITAL VIDEO STABILIZATION ALGORITHM FOR CMOS IMAGE SENSOR

Jeong-A Im, Dae-Woong Kim, and Ki-Sang Hong*

Dept. of E.E. in POSTECH, Pohang, Korea
{jeonga, dwkim, hongks}@posetch.ac.kr

ABSTRACT

This paper presents a novel motion compensation approach for stabilizing video sequences captured by CMOS image sensors. Although a number of papers have been dedicated to the compensation of shaky camera motions, most of them have focused on CCD image sensors. By relating the homography model to the characteristics of CMOS video sequences, we estimated the dominant motion in image sequences degraded by random shake, simultaneously estimating the distortion parameters of CMOS image sensors. Also, we considered the real-time applications with a fast linear-programming approach for the estimation of camera motion.

Index Terms— CMOS, CCD, Digital video stabilization.

1. INTRODUCTION

The goal of video stabilization is to remove unintentional shaky motions, while preserving intentional global motion in a video sequence captured by hand-held devices such as camcorder, PDAs, and mobile phones. In particular, CMOS image sensors have mostly been used in mobile phones due to their small size and low power consumption and cost compared with those of CCD image sensors.

A number of papers have been proposed in the video stabilization in which accurate motion estimation is critical. Considering real-time performances, Paik et al.[1] and Guestrin et al.[2] estimated the motions using block matching technique, which is commonly used in MPEG coding. Ko et al.[3] improved these approaches using bit-plane matching. However, the previous methods just considered CCD image sensors. There have been a few efforts for CMOS image sensors where video sequences are distorted according to the camera motion.

In this paper, by relating CMOS image to CCD image with a convenient distortion model, we define a new homography based motion model between sequential images. The homography is used to estimate both original camera motions and CMOS distortions. Therefore, accurate estimation of homographies is very important for the stabilization of CMOS image sequences. Among various methods[4, 5, 6], we use the linear-programming (LP) approach[6] where the homograpy can be computed from point-to-line correspondences. This method is very robust to outliers and easy to implement in real-time due to the low computational complexity. Once the homography is obtained, it can be easily decomposed into the translational camera motion and CMOS distortion. After compensating the distortion, the translational camera motions are smoothed with a low-pass filter to remove shaky motions.

Assuming translational camera motions, we relate the motion parameters to the distortion parameters of CMOS image sensor, which is a key contribution of our method. The framework presented in this paper has an efficient structure to be implemented in real time hard-ware design because it is fast and simple.

Section 2 describes the homography motion model for CMOS video images and section 3 gives the overall flow of video stabilization proposed in this paper. With experiments in section 4, we conclude our work with some remarks in section 5.

2. CHARACTERISTICS OF CMOS VIDEO

Basically, the CMOS image sensors have a different shuttering mechanism from CCD image sensors. To achieve the lowest cost and the smallest size, most CMOS image sensors adopt line scanning. Because of the time difference between the first and the last line of the CMOS sensor array, video sequences from the CMOS camera can be distorted depending on directions and variations of the camera motion.

In this section, a new distortion model is presented, which is related to the homography model from the analysis of the characteristics of CMOS image sensors.

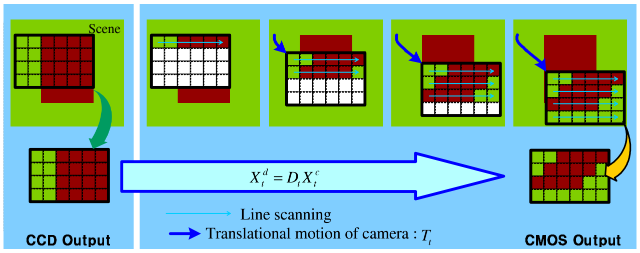


Fig. 1. Image acquisition models for CCD camera(left) and CMOS camera(right).

Fig. 2. CMOS video images. (a) A CMOS image without camera motion. (b) A compressed image when the camera moves up. (c) An extended image when the camera moves down. (d) A skewed image to the right when the camera moves left. (e) A skewed image to the left when the camera moves right.

2.1. CMOS distortion model

Figure 1 shows the image acquisition models for CCD and CMOS sensors. All CCD image sensors capture scene radiances simultaneously, while the CMOS image sensors capture the scene line by line (more precisely pixel by pixel). Thus, if the camera motion is very large compared with the scanning time, the captured image can be distorted as shown in the right part of Figure 1.

Figure 2 shows real COMS images. If the motion of camera is very small, the CMOS image will be the same as the output from the CCD camera as shown in Figure 2 (a). However, images from the CMOS camera will be distorted depending on directions and variations of camera motion as shown in Figure 2 (b)-(e).

Assuming translational motion of a line scanning camera as in [7], the point relationship between X_t^c in the CCD image and X_t^d int the CMOS image at t-th frame can be represented by a 3x3 distortion matrix D_t, which is given by

$$X_t^d = \begin{bmatrix} x_t^d \\ y_t^d \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{t_t^x}{Y+t_t^y} & 0 \\ 0 & \frac{Y}{Y+t_t^y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t^c \\ y_t^c \\ 1 \end{bmatrix} = D_t X_t^c, \tag{1}$$

where t_t^x and t_t^y are elements of the translation matrix given by

$$T_t = \begin{bmatrix} 1 & 0 & -t_t^x \\ 0 & 1 & -t_t^y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

In Equation (1), it should be noted that D_t depends only on T_t and the image height Y. Thus, the CMOS distortion can be estimated directly from the translation.

2.2. CMOS motion model

This section shows how a homography can be related to the global motion of the CMOS camera with the distortion model. For image sequences captured from the CCD camera, the displacement between corresponding points should be the same as the motion of camera if it is a pure translation. However, this is not true for CMOS images due to the distortion in Equation (1).

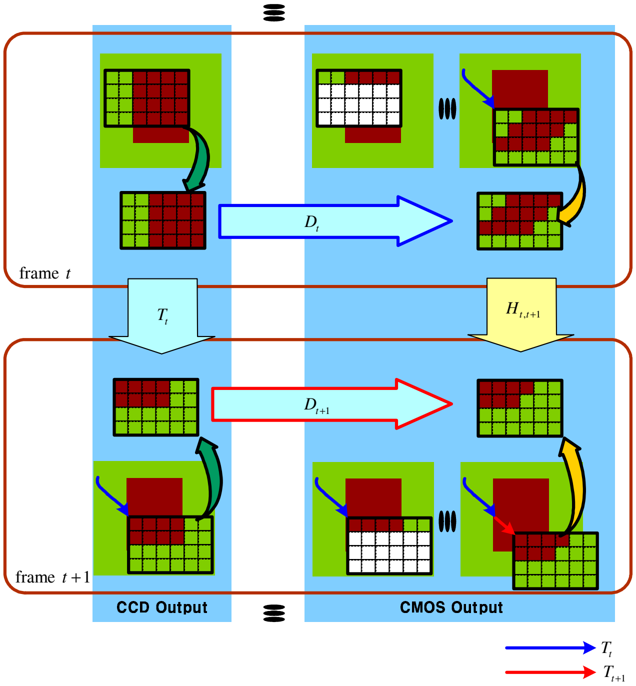


Fig. 3. Motion models for CCD camera(left) and CMOS camera(right).

Figure 3 shows an example describing the difference of motion between CCD and CMOS images. When the CCD camera moves with the translational motion between t-th and $t+1$-th frame, it is sufficient to use the translational motion model T_t as shown in the left part of Figure 3. However, the CMOS sensor can cause distortions according to the translational motion in each time period as shown in the right part of Figure 3. In this case, more complex motion model should be used to estimate an accurate motion of the camera.

Using the relations in Figure 3, a new motion model can be obtained by

$$X^d_{t+1} = H_{t,t+1} X^d_t = D_{t+1} T_t D_t^{-1} X^d_t. \tag{3}$$

Substituting Equation (1) and (2) into Equation (3), the matrix $H_{t,t+1}$ can be written as

$$H_{t,t+1} = \begin{bmatrix} 1 & \frac{t^x_t}{Y} - \frac{t^x_{t+1}(Y+t^y_t)}{Y(Y+t^y_{t+1})} & -t^x_t + \frac{t^x_{t+1} t^y_t}{Y+t^y_{t+1}} \\ 0 & \frac{Y+t^y_t}{Y+t^y_{t+1}} & -t^y_t(1 - \frac{t^y_{t+1}}{Y+t^y_{t+1}}) \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

Finally, we used Equation (4) as the motion model of the CMOS camera with pure translational motions. It should be noted that the translation matrix T_t can be estimated directly from $H_{t,t+1}$, and the distortion matrix D_t can be estimated from T_t using the relation in Equation (1). As a result, the computation of $H_{t,t+1}$ is essential for the recovery of motion parameters and distortion parameters. Moreover, it implies that the robust motion estimation is the core of the video stabilization system, because more accurate motion information provides a better performance.

3. DIGITAL VIDEO STABILIZATION (DVS) ALGORITHM

The proposed DVS algorithm consists of four stages: motion estimation, distortion estimation, motion filtering, and compensation.

In our approach, motion parameters are recovered from the linear-programming approach[6] where they used point-to-line correspondences to calculate homography between images given by

$$\hat{H}_{t,t+1} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}. \tag{5}$$

Using the equality condition between Equation (4) and (5), we can estimate the motion parameters t^x_t and t^y_t as a function of the elements in the estimated homography $\hat{H}_{t,t+1}$. Then, the camera motion T_t and the CMOS distortion D_t can be obtained using Equation (1) and (2).

After compensating the CMOS distortion with D_t in each time step, unintentional shaky motions are smoothed out with a low-pass filter[8]. Finally, a stabilized video sequence can be generated with a simple warping procedure using the filtered motion parameters.

4. EXPERIMENTAL RESULTS

In this section, we present the experimental results of applying our method to real CMOS sequences. In each experiment, images(176x144) are captured by a mobile phone(Curitel PD-6500) with a frame rate of six.

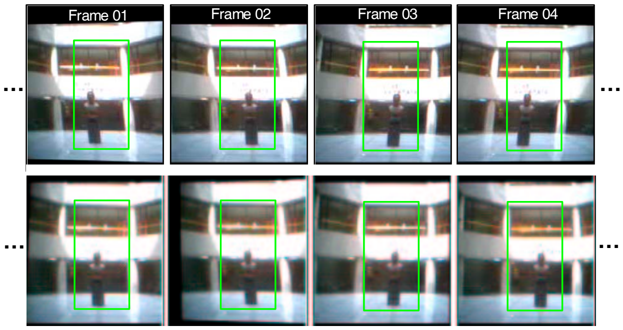


Fig. 4. The original sequence with shaky motion(top row) and the stabilized sequence(bottom row).

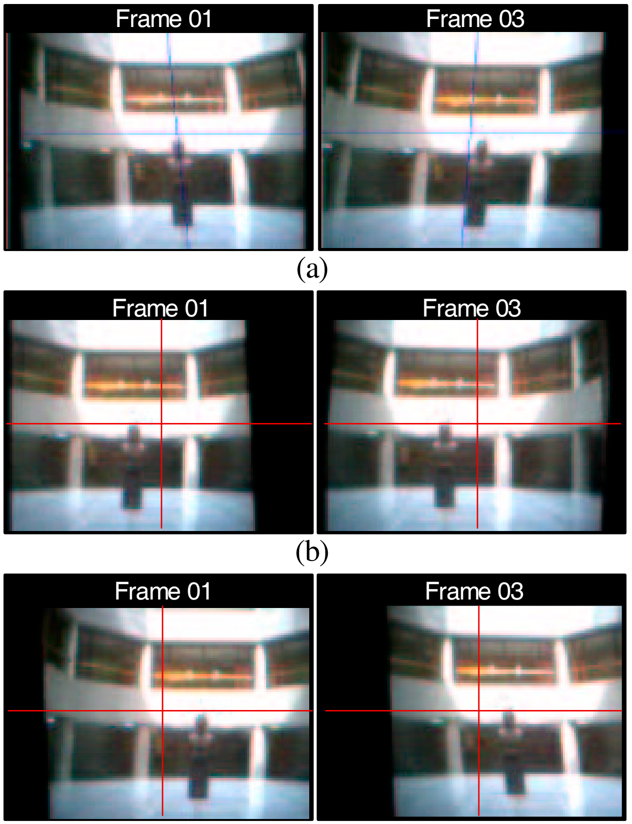
Figure 4 shows the stabilization results where the skewed lines are successfully compensated. It is noted that an accurate motion estimation is very important for a successful video stabilization. Figure 5 (a) shows the estimated distortion from motion estimation results, and Figure 5 (b) shows the stabilized images. However, if the motion estimation is not successful, the compensation will fail as shown in Figure 5 (c) where the motion is estimated using the block matching method[4]. In Figure 6, we compared optical flows of the original sequence and the stabilized sequences. From this figure, we expect that the proposed motion model can improve many computer vision algorithms using the CMOS image sensors.

Figure 7 shows the comparison results using three different motion estimations: the block matching approach [1], the linear-programming (LP) approach[6], and the pixel based Levenberg-Marquardt (LM) method[4]. We considered the results from the LM method as the true value because the ground truth is not given in the experiments. The accuracy of LP method is very high while the computational cost is very low compared with the simple block matching method. In this sense, the framework presented in this paper can be implemented in real time using the LP approach and the proposed motion model.

5. CONCLUSION

By introducing a new homography based motion estimation method, we robustly stabilized a shaky video from the CMOS image sensors. By incorporating characteristics of the CMOS image sensor into the distortion model and using the homography model, we could calibrate the CMOS distortion parameters and translational camera motions. Also, we considered real-time applications with a fast linear-programming approach for the estimation of homographies. We expect that the proposed motion model can be used in many applications using sequences taken from the CMOS image sensors.

At the moment, we are considering pure translational mo-



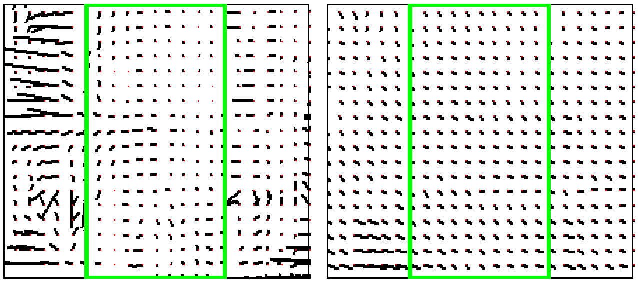
(a)

(b)

(c)

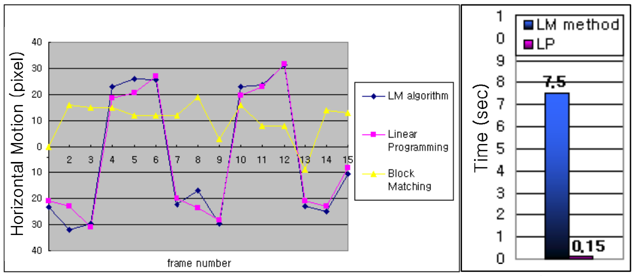
Fig. 5. Video stabilization results for frame01 and frame03 in Figure 4. (a) Distortion estimation from the estimated homographies with the LP approach[6]. Input images with blue lines representing the estimated distortion. (b) Compensation results from the proposed method. (c) Compensation results from the estimated homographies with the block matching approach[1].

tions of the CMOS cameras. However, other general motions such as rotation will be considered in the future research. Also, we need further research on the effect of image blur caused by fast camera motions.



(a) (b)

Fig. 6. Optical flow between frame03 and frame04 of the original sequence(left) and the stabilized sequence(right). The green rectangles represent the selected regions in Figure 4.



(a) (b)

Fig. 7. Comparison results using various motion estimation methods. (a) Horizontal motion. (b) Computational cost.

6. REFERENCES

[1] J. K. Paik, Y. C. Park, and D. W. Kim, "An adaptive motion decision system for digital image stabilizer based on edge pattern matching," *IEEE Transactions on Consumer Electronics*, vol. 38, pp. 607–615, 1992.

[2] C. Guestrin, F. Cozman, and M. G. Simoes, "Industrial applications of image mosaicing and stabilization," *Knowledge-Based Intelligent Electronic Systems*, vol. 2, pp. 174–183, 1998.

[3] S. J. Ko, S. H. Lee, and K. H. Lee, "Digital image stabilizing algorithms based on bit-plane matching," *IEEE Transactions on Consumer Electronics*, vol. 44, pp. 617–622, 1998.

[4] R. Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications*, vol. 16, pp. 22–30, 1996.

[5] J. Shi and C. Tomasi, "Good features to track," *CVPR*, pp. 593–600, 1994.

[6] M. Verman, "A real-time motion analysis with linear-programming," *ICCV*, pp. 703–709, 1999.

[7] J. H. Hwang, "Cmos image sensor distortion compensation method," *Korea Patent 10-0092224*, 2003.

[8] A. J. Crawford, H. Denman, F. Kelly, F. Pitie, and A. C. Kokaam, "Gradient based dominant motion estimation with integral projections for real-time video stabilization," *ICIP*, pp. 3371–3374, 2004.