Video stabilization with sub-image phase correlation

Juanjuan Zhu (朱娟娟) and Baolong Guo (郭宝龙)

School of Electromechanical Engineering, ICIE Institute Xidian University, Xi'an 710071

Received May 18, 2006

A fast video stabilization method is presented, which consists of sub-image phase correlation based global motion estimation, Kalman filtering based motion smoothing and motion modification based compensation. Global motion is decided using phase correlation in four sub-images. Then, the motion vectors are accumulated to be Kalman filtered for smoothing. The ordinal motion compensation is applied to each frame with modification to prevent error propagation. Experimental results show that this stabilization system can remove unwanted translational jitter of video sequences and follow intentional scan at real-time speed.

OCIS codes: 330.4150, 040.7290, 100.20000.

Electronic image stabilization (EIS) aims at removing camera jitter from a video sequence to obtain a compensated sequence that displays smooth camera movements only. It has two typical application areas: consumer video cameras and hand held wireless visual communication equipments. Up to now, many algorithms[1−3] have been proposed. In all, the key problems are how to obtain reliable global motion vector and how to remove (or to be precise, reduce) camera jitter while following the intentional camera pan. Considering these two problems, we proposed a fast video stabilization system, as shown in Fig. 1.

Motion estimation is realized in the Fourier domain[4] to operate on the translational global motion component. To keep the computation load low and at the same time sufficient image content for correct estimation, four blocks of 64×64 are selected in each corner of the original 320×240 image, as shown in Fig. 2. Let G_1 andG_2 represent the two-dimensional (2D) discrete Fourier transform (DFT) of images I_1 and I_2 respectively, their inverse DFT (IDFT) of cross-power spectrum[5] can be defined as

$$F^{-1}(\mathrm{e}^{j(\omega_x \mathrm{d}x+\omega_y \mathrm{d}y)}) = F^{-1}\left(\frac{G_2(\omega_x,\omega_y)G_1^*(\omega_x,\omega_y)}{|G_2(\omega_x,\omega_y)|\,|G_1^*(\omega_x,\omega_y)|}\right) = \delta(x+\mathrm{d}x, y+\mathrm{d}y), \qquad (1)$$

where $(\mathrm{d}x, \mathrm{d}y)$ respectively stands for horizontal and vertical displacements. Image I_2 is a translation $(\mathrm{d}x, \mathrm{d}y)$ of I_1

$$I_2(x,y) = I_1(x+\mathrm{d}x, y+\mathrm{d}y). \qquad (2)$$

Accordingly, the transformation between G_1 and G_2 can be defined as

$$G_2(\omega_x,\omega_y) = G_1(\omega_x,\omega_y)\mathrm{e}^{j(\omega_x \mathrm{d}x+\omega_y \mathrm{d}y)}. \qquad (3)$$

Hence, if I_2 is a spatially shifted version of I_1, the phase correlation surface is zero everywhere except for a delta function at location $(-\mathrm{d}x, -\mathrm{d}y)$, which corresponds to the displacement between images. So, for all four sub-images of an image frame, local motion vectors are estimated from the respective sub-images of the previous frame based on the above phase correlation. All local motion vectors can then be weighted proportionally to their peak amplitude values and the result can be assigned as the global inter-frame motion vector.

Motion smoothing is then carried on to smooth the original accumulated motion vectors, i.e., the smoothed displacement representing deliberate camera movements of the stabilized sequence. It has been demonstrated that the Kalman filter can be used to effectively stabilize frame displacements in real-time. The Kalman filter provides a state system in the form of $X(k+1) = \Phi(k)*X(k)+w(k)$ and the observation system in the form of $Z(k) = H(k)*X(k)+v(k)$, where $w(k)$ represents the process noise and $v(k)$ shows the measurement noise. For the image displacement, with (x_1, x_2) representing the horizontal and vertical translation and $(\mathrm{d}x_1, \mathrm{d}x_2)$ representing the corresponding velocity, the state system and the observation system are defined as

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \mathrm{d}x_1(k+1) \\ \mathrm{d}x_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \mathrm{d}x_1(k) \\ \mathrm{d}x_2(k) \end{bmatrix} + \begin{bmatrix} wx_1(k) \\ wx_2(k) \\ w\mathrm{d}x_1(k) \\ w\mathrm{d}x_2(k) \end{bmatrix}, \qquad (4)$$

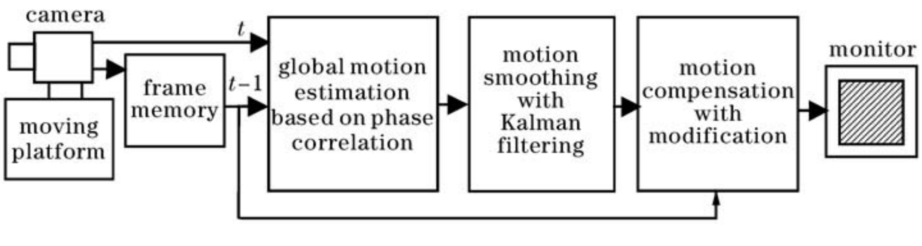

Fig. 1. Block diagram of the stabilizatin system.

1671-7694/2006/090553-03 http://www.col.org.cn

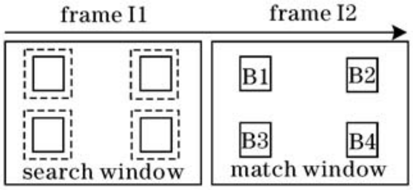

Fig. 2. Sub-images used for phase correlation.

$$\begin{bmatrix} Z_1(k) \\ Z_2(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \mathrm{d}x_1(k) \\ \mathrm{d}x_2(k) \end{bmatrix} + \begin{bmatrix} v_1(k) \\ v_2(k) \end{bmatrix}. \tag{5}$$

The Kalman filter estimates the process state from Eq. (4) (prediction stage) and then obtains feedback in Eq. (5) (update stage). The filter output is obtained recursively through these stages enabling real time operation of the filter.

After filtering, the proposed ordinal modified motion compensation is applied frame by frame. With the preserved smooth displacement representing deliberate camera movements of the stabilized sequence, each stabilized frame is built upon previous stabilized frame. However, error at the beginning will propagate to subsequent frames. For the case of camera translation, the accumulation of such errors occurs at the frame boundaries and appears as stripe-pattern artifacts, as shown in Fig. 3. To remove these annoying artifacts, modification technique is used as follows.

At each frame k, the accumulated difference is calculated as

$$D = \left| \sum_{l=1}^{k-1} u_{l,\mathrm{o}} - \sum_{l=1}^{k-1} u_{l,\mathrm{s}} \right|^2 + \left| \sum_{l=1}^{k-1} v_{l,\mathrm{o}} - \sum_{l=1}^{k-1} v_{l,\mathrm{s}} \right|^2, \tag{6}$$

where $[u_l, v_l]$ is the estimated motion vector, the subscripts "o" and "s" imply "original" and "smoothed", respectively. If $D <$ threshold, the frame $k+1$ is obtained by compensating the original frame k with $[u_{k,\mathrm{s}}, v_{k,\mathrm{s}}]$ and the motion between the stabilized frames $k+1$ and k is estimated to be the modification; else the frame $k+1$ is obtained by compensating the previous stabilized frame k with $[u_{k,\mathrm{s}}, v_{k,\mathrm{s}}]$. In practical, when threshold is equal to 12 the motion shows good performance. From Fig. 3, the motion is modified from $(19, -4)$ to $(7, -2)$, hence, boundary error decreases greatly. In addition, we also choose to force a frame synchronization every 12 frames in the case where the above scheme is not invoked.


(a) (b)

Fig. 3. Illustration of boundary error. (a) Artifacts near the boundary (right) due to error propagation; (b) result of error modification.

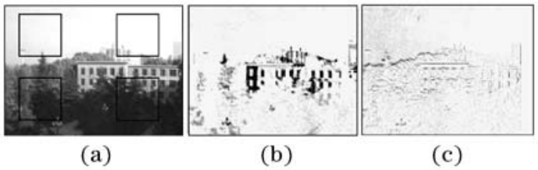
(a) (b) (c)

Fig. 4. Illustration of image difference. (a) An original frame in a sequence; (b) the difference between (a) and its previous frame; (c) the according difference of the stabilized frames at the same time.

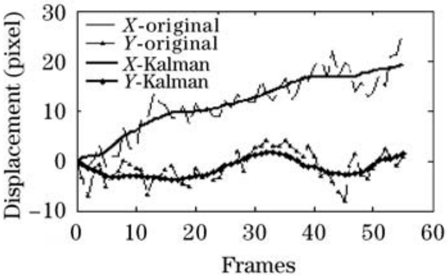

Fig. 5. Results of motion smoothing with Kalman filtering.

Experiments are carried on video sequences recorded from a camera mounted on a dithering test platform, whose dithering frequency is between 1—5 Hz. The experiments are realized at 25 fps with computer and the video capture device (model, 10Moons SDK2000). The video sequence is captured at frame rate of 30 fps with RGB 24 image size of 240×320. Figure 4 gives a sample frame from a viedo sequence. The sub-images used for local motion vectors are shown superimposed in Fig. 4(a). The performance is evaluated using difference between two consecutive frames from original and stabilized sequences, as shown in Figs. 4(b) and (c), respectively. In both case, only the pixels with absolute difference higher than 25 are displayed. Since the algorithm attempts to remove large and rapid camera, the decrease of the difference between images indicates that more smooth motion is resulted. Figure 5 shows absolute frame displacements of the sequence, of which the horizontal and vertical absolute motion are referred to as X-original and Y-original respectively. It can clearly be seen that the sequence contains relatively high-frequency jitter in both dierections. The smoothed displacements obtained by the Kalman filter output are also displayed in Fig. 5, and likewise the smoothed ones are referred to as X-kalman and Y-kalman. Kalman filter successfully removes high-frequency jitter and smoothly follows the global motion trajectory.

In this paper, phase correlation is used in four sub-images to detect motion, Kalman filter is applied to smooth displacements, and error control is considered. From these experiments, it can be seen that the system can stabilize imge fluctuations caused by camera translational jitter and hence improves visual quality of the image sequence.

This work was supported by the National Science Foundation of China under Grant No. 60572152. J. Zhu's e-mail address is zhujoo@126.com.

References

1. F. Vella, A. Castorina, M. Mancuso, and G. Messina, IEEE Trans. Consum. Electron. **48,** 796 (2002).
2. S.-J. Ko, S.-H. Lee, and K. H. Lee, IEEE Trans. Consum. Electron. **44,** 617 (1998).
3. J. Zhu and B. Guo, Acta Opt. Sin. (in Chinese) **26,** 516 (2006).
4. S. Ertürk and T. J. Dennis, IEEE Proc. on Image Vision and Signal Processing **147,** 95 (2000).
5. Y. He and X. Yang, Chin. Opt. Lett. **2,** 323 (2004).