

Electronic Image Stabilization based on Motion Classification

Zhu Juanjuan¹, Guo Baolong

Abstract. A novel electronic image stabilization based on motion classification is presented. It achieves image stabilization with high accuracy in complicated camera motion. Firstly, the feature points are selected evenly using the Harris operator and then matched based on the feature window. Hence, the corresponding points are obtained and the statistic features of their motions are analyzed according to different motion kinds including translation, rotation and zoom. Then, the fast motion classification method is proposed to validate all points and delete mismatched or unreliable points. Thirdly, the remained global feature points are brought to the affine model to compute global motion. Lastly, the Kalman filter is used to obtain dithering motion and each current frame is compensated. Experimental results show that the algorithm can correctly detect global motion in dynamic scenes with camera scan and various dithering. Furthermore, its estimation error is below 1/2 pixel and it also accomplishes real-time stabilization, which can greatly improve the stability and fidelity of videos.

Keywords: electronic image stabilization, global motion estimation, motion compensation, feature points

1.1 Introduction

Due to the camera dithering of moving vehicles, it will capture fuzzy or blurring videos, which need to be stabilized. Electronic Image Stabilization^[1] (EIS) is to detect and remove or reduce unwanted motion by image processing methods. It can improve the stability and fidelity of videos and will offer foundation for moving targets detecting and tracking^[2]. It has been widely used in military and civil camera systems.

The EIS system mainly consists of global motion estimation and motion compensation. The traditional algorithms including block matching^[3], projection algorithm^[4] and bit-plane matching^[5] can only deal with simple translational jitter with low accuracy. Nowadays, many feature points matching algorithms^[6-8] are proposed. The points at fixed position are likely to be mismatched. The detection

¹ Juanjuan Zhu (✉) This work is supported by the National Science Foundation of China 61003196
School of Electromechanical Engineering, Xidian Univ, 710071, Xi'an, China
e-mail: zhujuo@126.com

process of scale-invariant points is too complicated to obtain real-time speed. The motions in videos include translation, rotation, and zoom of the camera and local motion of objects. The accuracy will be reduced if simple algorithms are used to detect complex motions, meanwhile, the computation cost will be increased if complicated algorithms are used to deal with simple translation. So, it is necessary to classify and process motion types respectively.

This paper proposed a Motion Classification based Image Stabilization algorithm (MCIS), which combined features of camera motions with statistic property of feature points' motions. **Fig. 1.1** shows the flow chart of MCIS algorithm. According to camera translation, rotation and zoom determination, different methods are used to validate point pairs, which can reduce bad effect of local motion or mismatched points. All the validated pairs are brought into motion model to compute global motion parameters. Then, the motion filter is applied to get dithering motion to compensate each current frame. The MCIS algorithm can stabilize videos at real-time speed and shows adaptability for complex scenes.

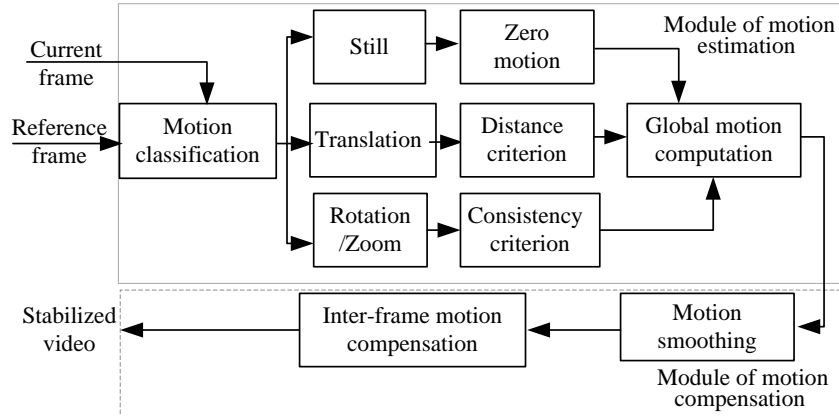
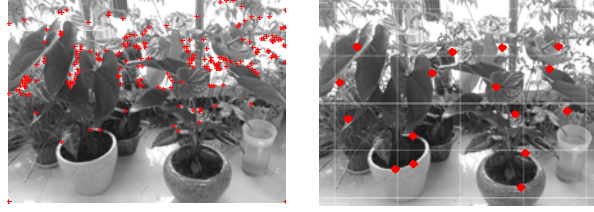


Fig. 1.1 The flow chart of MCIS

1.2 Global Motion Estimation by Motion Classification

1.2.1 Feature Points Selection

The Harris operator^[9] is used to detect feature points in reference frame. To avoid points clustering, the reference is divided into non-overlapping blocks. The local extreme Harris point is selected as feature point. As **Fig. 1.2** shows, the traditional Harris points focus in edges, nevertheless, the modified points distribute uniformly and these independent points can stand for the global motion.



(a) Traditional Harris points (b) Modified sub-area points

Fig. 1.2 Comparison of point selection

1.2.2 Feature Points Matching

Feature points' matching is made between adjacent frames, as shown in **Fig. 1.3**. Each point in reference frame needs to be matched in current frame. The similarity measurement of feature windows is applied to find the local optimum matching feature points. The feature point $[x, y]^T$ in reference is set as the center of each feature window. Then, the diamond search with SAD^[10] criterion is used to find the best matched window. Its center is the corresponding point $[\hat{x}, \hat{y}]^T$ in current frame. The motion model between points is shown in Equation (1.1).

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} m_0 & m_1 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_2 \\ m_5 \end{bmatrix}, \quad M^T_{affine} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \end{bmatrix} \quad (1.1)$$

where m_2 and m_5 represent the translation, and m_0, m_1, m_3 and m_4 reflect the scale and rotation.

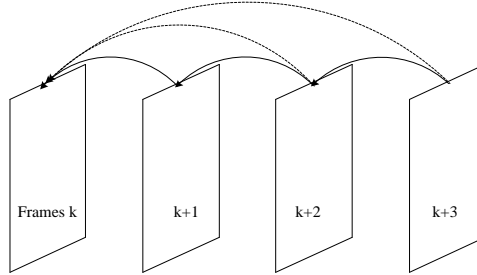


Fig. 1.3 Illustration of inter-frame matching

1.2.3 Global motion computation based on motion classification

1.2.3.1 Fast Classification of Motion Types

Assume that the camera is viewing a flat still scene, the inter-frame motions are classified into four types: still, translation, zoom and rotation. Suppose $[x, y]^T$ and $[\hat{x}, \hat{y}]^T$ is the matched pair, $MV_{i,j} = (\Delta x, \Delta y)$ is defined as $\Delta x = x_i - \hat{x}_i, \Delta y = y_i - \hat{y}_i$. The direction $\Delta x / \Delta y$ and amplitude $\sqrt{\Delta x^2 + \Delta y^2}$ are selected to determine motion type.

(1) Still

If $\sqrt{\Delta x^2 + \Delta y^2}$ is smaller than the threshold, the motion is regarded as still. If the number of the still motions reaches half of all the motions, the camera motion is considered as still.

(2) Translation

If the motion of camera is pan, pitch or yaw, the inter-frame motion will be translation with uniform direction and amplitude, as shown in **Fig. 1.4(a)**. If the direction $\Delta x / \Delta y$ and amplitude $\sqrt{\Delta x^2 + \Delta y^2}$ of each point pair is consistent, the motion type is considered as translation.

(3) Zoom

Fig. 1.4(b) gives the motion filed of camera zoom-out or dolly. Almost all the motions point at the FOE (focus of expansion). If FOE can be found, the motion type can be regarded as zoom. According to the statistic property of the motion filed, the process of searching FOE can be as follows. The motion with $\Delta y=0$ is first found at the symmetry line from left to right in the field. Then, we will search from left to right in this line to find FOE, whose Δx changes from positive to negative.

(4) Rotation

Fig. 1.4(c) gives the motion filed of camera clockwise roll. Almost all the motions distribute around the FOE in a counterclockwise circle. The process of searching FOE can be as follows. The motion with $\Delta y=0$ is first found at the symmetry line from top to bottom in the field. Then, we will search from top to bottom in this line to find FOE, whose Δx changes from positive to negative.

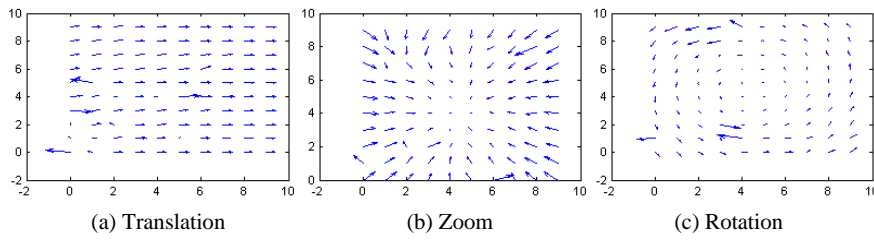


Fig. 1.4 Illustration of camera motion fields

1.2.3.2 Points Validation based on Motion Types

After motion classification, the feature points need to be validated to delete mismatched or local points. And all the global points can stand for the global motion.

(1) Still

If the motion type is still, the global motion is $M_{affine}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

(2) Translation

The distance criterion is proposed to delete the local points. The distance is still defined as $\sqrt{\Delta x^2 + \Delta y^2}$. The distance criterion takes the global coherence of all feature pairs as a whole and intends to find the globally matched pairs, defined as global features. The verification is as the following steps.

1) Calculate the distances between N corresponding pairs and then classify these distances into $k(k \leq N)$ kinds.

2) Accumulate the point pairs' number $c_i (i = 1, 2, \dots, k)$ of each kind.

3) Find $c = \max\{c_i | i = 1, \dots, k\}$ and then its corresponding pairs are the correctly matched global features.

(3) Zoom or rotation

If the motion is zooming, the motions are divided into eight groups reflecting eight directions, as shown in **Fig. 1.5(a)**. The consistency in direction and amplitude is checked to validate points. The motion in group (3a) disagrees with other directions, so, the corresponding feature point is deleted.

If the motion is rotation, the motions are divided into three groups reflecting three circles, as shown in **Fig. 1.5(b)**. With the consistency criterion, the motion in group (2) disagrees with others, so, the corresponding feature points are deleted.

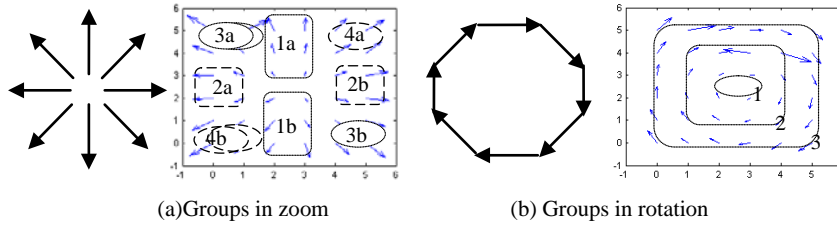


Fig. 1.5 Property and groups of zoom and rotation

1.2.3.3 Global Motion Computation

All the verified c pairs of feature points are then brought into equation (1.1) to get $2c$ -linear equations. The final function $B = AM_{affine}$ is in form of matrix, as shown in Equation (1.2). The global motion matrix is M_{affine} . The least-square solution of the over determined linear equation is defined as $M_{affine} = (A^T A)^{-1} A^T B$.

$$B = \begin{bmatrix} \hat{x}_1 & y_1 \\ \hat{x}_2 & y_2 \\ \vdots & \vdots \\ \hat{x}_c & y_c \end{bmatrix} = AM_{affine} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_c & y_c & 1 \end{bmatrix} \times \begin{bmatrix} m_0 & m_3 \\ m_1 & m_4 \\ m_2 & m_5 \end{bmatrix} \quad (1.2)$$

1.3 Inter-frame compensation

After global motion estimation, the motion vectors are filtered to get dithering component. It has been demonstrated that the Kalman filter is effective to stabilize frame displacements in real-time. For each frame k , the accumulated motion vector is $Z_{raw}(k)$, and the Kalman filtered motion vector is $Z_{Kal}(k)$. So, the correction vector representing jitter is calculated as $Z_{cor}(k) = Z_{Kal}(k) - Z_{raw}(k)$.

Motion compensation is made on each current frame with $Z_{cor}(k)$ to reduce unwanted camera jitter. Because the transformed coordinates are not always at the integer pixels, we use the bilinear interpolation to determine the gray value of non-integer point at $(i + p, j + q)$.

$$f(i + p, j + q) = (1 - p)(1 - q) \cdot f(i, j) + (1 - p)q \cdot f(i, j + 1) + p(1 - q) \cdot f(i + 1, j) + pq \cdot f(i + 1, j + 1) \quad (1.3)$$

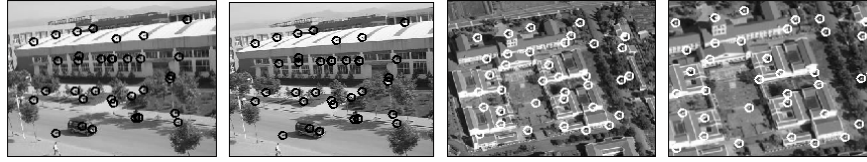
Where i and j are integer numbers, p and q are float numbers in $[0, 1)$.

1.4 Experimental Results Analysis

This section presents experimental results obtained from videos in a camera mounted on a dithering test platform. The system achieves 25.4fps on PC and the image size is 480×640 with a maximum displacement of ±40 pixels, rotation of 15° and zoom of 30%.

1.4.1 Results of Feature Points Selection and Matching

Fig. 1.6(a) shows two consecutive frames in video 1 with a small moving object. **Fig. 1.6(b)** is the result of video 2 with scale zoom. The Harris feature points are selected and matched in 6×6 sub-areas. It can be observed that most feature points are correctly matched.

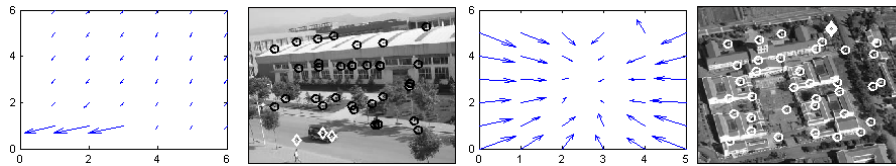


(a)Points selected and matched in video1 (b) Points selected and matched in video2

Fig. 1.6 Result of feature selection and matching

1.4.2 Results of Feature Points Validation

From the statistic property of feature points' motion, the camera motion type of video 1 is translation and its motion filed is shown in **Fig. 1.7(a)**. Almost all the motions are consistent in direction and amplitude except the three motions in the left-down area, which stand for the local moving car. The local points are deleted with the distance criterion, as “ \diamond ” marked in **Fig. 1.7(b)**. Similarly, the camera motion type of video 2 is zooming and its motion filed is shown in **Fig. 1.7(c)**. The local point is deleted, as “ \diamond ” marked in **Fig. 1.7(d)**.



(a)Motion field of video1 (b) Validated points (c) Motion field of video 2 (d) Validated point

Fig. 1.7 Result of feature points deletion

1.4.3 Results of Global Motion Computation and Compensation

Table 1.1 Results of time and accuracy of motion estimation

video	Direct Computation	Validated Computation	Time (ms)	Original PSNR(dB)	Compensated PSNR(dB)
1	$\begin{bmatrix} 1.000 & -0.054 \\ 0.0508 & 0.9822 \\ -20.036 & 11.0341 \end{bmatrix}$	$\begin{bmatrix} 1.0138 & -0.044 \\ 0.0610 & 0.9809 \\ -19.5840 & 11.6860 \end{bmatrix}$	28.45	19.87	29.32
2	$\begin{bmatrix} 1.0012 & -0.041 \\ 0.0011 & 1.0200 \\ -0.0223 & 1.0035 \end{bmatrix}$	$\begin{bmatrix} 0.9993 & -0.0027 \\ 0.0004 & 1.0000 \\ -1.0190 & 0.1927 \end{bmatrix}$	21.89	22.96	33.29

In **Table 1.1**, global motion is computed respectively and compared between direct computation by all points without validation and validated computation by motion classification. And PSNR is used to evaluate the inter-frame fidelity before and after compensation. It can be seen that the processing speed achieves up to 25fps and PSNR improves 46.5% comparing to original videos.

1.5 Conclusion

In conclusion, a fast image stabilizing algorithm based on motion classification is proposed to deal with complex motion in camera system. It improves global motion estimation by analyzing statistic property of motion field to classify motion type, based on which points are validated to delete local points. And Kalman filter accomplishes motion smooth following the intentional scan. The MCIS achieves successful stabilization performance at 25fps. Further work will be directed to deal with depth-variant image scene considering depth information.

1.6 References

1. Jinhee Lee, Sangkeun Lee and Joonki Paik. Digital image stabilization based on statistical selection of feasible regions [J]. IEEE Trans on Consumer Electronics, 2009, 55(4): 1748-1755.
2. Xing Zhuoyi, Zhu Qidan, Lin Zhuang. Target recognizing and tracking algorithm based on Gabor wavelet [J]. Control and Decision, 2008, 23(2): 162-166.
3. Tico M. Adaptive block-based approach to image stabilization [C]. In 15th IEEE International Conference on Image Processing, 12-15 Oct. 2008: 521-524.
4. Moshe Y and Hel-Or H. Video block motion estimation based on gray-code kernels [J]. IEEE Trans. on Image Processing, 2009, 18(10): 2243-2254.
5. S.-J. Ko, S.-H. Lee, et al. Fast digital image stabilizer based on gray-coded bit-plane matching [J]. IEEE Trans. on Consumer Electronics, 1999, 45(3): 598-603.
6. Tang Jin, Han Xiaowei, Yuan zhonghu, et al. An approach of electronic image stabilization based on the representative point matching [C]. In Third International Conference on Genetic and Evolutionary Computing, IEEE, 2009: 347-350.
7. David G Lowe. Distinctive image features from scale-invariant key-points [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
8. Yao Shen, Guturu P, Damarla T, et al. Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework [J]. IEEE Trans on Consumer Electronics, 2009, 55(3): 1714-1721.
9. C. Harris, M. Stephens. A combined corner and edge detector [C]. Fourth Alvey Vision Conference, 1988: 17-151.
10. S. Zhu and K. K. Ma. A new diamond search algorithm for fast block-matching motion estimation [J]. IEEE Trans. on Image Processing. 2000, 9(2): 287-290.