# Real-Time Digital Image Stabilization Algorithm using Sage-Husa Filter

Juan-Juan Zhu and Bao-Long Guo

School of Aerospace science and technology
Xidian University
No.2 Taibai Road, xi'an, 710071, China
zhujoo@126.com;blguo@xidian.edu.cn

Chuang Ge

Changchun Institute of Optics, Fine Mechanics and Physics
Chinese Academy of Sciences,Changchun,100039,China
gczongheng@163.com

ABSTRACT. *We propose a digital image stabilization algorithm based on improved Sage-Husa filter. Global motion estimation is obtained by detecting interest points from reference frame, matching points in current frame, and solving motion equations. The motion includes intentional camera scan and unwanted jitter; therefore, we need to smooth global motion sequence to remove jitter. The Sage-Husa filter is improved by modifying process and observation noise using their statistical property. The prediction error matrix can be adjusted according to the property of innovation sequence to avoid filter divergence. Finally each current frame is warped by fast motion compensation in consideration of the linear storage of the image. We investigate the filter accuracy and visual quality by experimental evaluation. The results show that the proposed algorithm is capable of smoothing inter-frame jitter and tracking real scene at real time speed.*
**Keywords:** Digital image stabilization; Sage-Husa filter; Global motion estimation

1. **Introduction.** The camera on unstable platforms in surveillance applications will capture a shaky video with low perceptual quality. Digital Image Stabilization (DIS) is designed to detect and remove inter-frame global motion. It has been widely used in moving object detection and target tracking [1], walking robot, video compression and image mosaic.

In general, DIS system consists of Global Motion Estimation (GME) and Motion Compensation (MC). GME estimates inter-frame motion and its most popular method is the feature matching [2] method, which can obtain translation, rotation and zooming. MC is designed to correct motion by smoothing the motion parameters to reduce dithering. The previously proposed algorithms include motion attenuation [3], average filter [4]and Kalman filter [5, 6]. The attenuation coefficient is set by experimental results, which is not applicable to all videos. The average filter has simple computation but the extra low frequency noise still exists. The Kalman filter is based on the assumption that the noise is given and obeys the Gauss distribution with zero mean value, which is impossible in real applications.

Based on visual attention mechanism, the video is smooth if camera moves uniformly at low frequency speed. When the camera is placed on moving vehicles such as planes, cars

or ships, the jitter is assumed to be high-frequency random vibration, but the intentional camera scan is smooth at low-frequency. Based on the above analysis, a fast digital image stabilization algorithm using adaptive filter is proposed. It detects interest points using Harris operator in regions and matches points with template. The pairs are verified with the statistical property of distances and brought into the motion model to compute global motion. Then, motions are filtered by the proposed improved Sage-Husa filter, which can estimate process and observation noise of Kalman filter in real time. Finally, we use fast linear compensation method to ensure real-time performance. Experimental results show that the proposed method can reduce jitter and track the camera scan with improved visual quality.
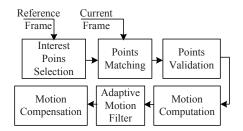


FIGURE 1. Flow chart of proposed algorithm

2. **Parametric motion model.** Various parametric motion models have been used for global motion estimation. The most commonly used motion model is similarity model, as established in equation (1).

$$P = T \cdot \tilde{P} + b \tag{1}$$

$$T = \begin{bmatrix} 1 & -\theta \\ \theta & 1 \end{bmatrix}, b = \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} \tag{2}$$

Here, $P = (x, y)^T$ and $\tilde{P} = (\tilde{x}, \tilde{y})^T$ are corresponding pixels in adjacent images. It describes rotation $\theta$ and translation $(\Delta X, \Delta Y)$ in horizontal and vertical direction.

3. **Global motion estimation.**

3.1. **Interest point selection.** In this paper, global motion estimation is realized by extracting and tracking interest points. The classic Harris operator is robust to noise and light variation [7]. The principle of Harris corner detection is as follows:

$$H_C = \det \hat{C} - K \cdot trace\hat{C} \tag{3}$$

$$\det \hat{C} = \lambda_1 \cdot \lambda_2 , trace\hat{C} = \lambda_1 + \lambda_2 \tag{4}$$

$$\hat{C} = \frac{1}{2\pi\sigma^2} \exp(-\frac{u^2 + v^2}{2\sigma^2}) \otimes \begin{bmatrix} I_u^2 & I_u I_v \\ I_u I_v & I_v^2 \end{bmatrix} \tag{5}$$

Where, symbol $\tilde{I}$ represents applying Gauss filter on image $I.I_u$ is the gray difference in horizontal direction and $I_v$, vertical direction. $[\lambda_1, \lambda_2]$ is the eigenvalue of the auto-correlation matrix $\tilde{C}$.It is supposed that point with large eigenvalue has high interest value $H_C$ and is taken as the interest point.

The interest points that detected by Harris operator are too dense in edge or corner region, as shown in Fig. 2(a). The points are mostly detected in the moving object, which are defined local points. Therefore, the Harris operator is modified to select interest points

in divided areas. The reference image is divided into $r \times s$ non-overlapped regions, and in each region the point with the maximal $H_C$ is selected as interest point. Thus these independent points distribute uniformly, as shown in Fig.2(b).



(a) Result of Harris operator    (b) Result of sub-area Harris operator

FIGURE 2. Result of Harris points detection

3.2. **Point matching and validation.** Each interest point in reference frame is matched using template matching to get corresponding point in current frame. The template is selected as a block around each interest point. Then, the sequential similarity detection algorithm (SSDA) is used to find the best matched block and its center is the corresponding point in current frame.

Considering local points will interfere with global motion, the matched points should be validated to delete mismatched or local points. The distance criterion [8] is used to verify points. The Euclidean distance $d_i = \sqrt{(x_i - \tilde{x}_i)^2 + (y_i - \tilde{y}_i)^2}$ is defined between point pair in adjacent frames. According to the statistic of experimental data, the verification is as the following steps.

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

2) Accumulate the point pairs number $n_i(i = 1, 2, \cdots, k)$ of each kind.

3) Find $N = \max\{n_i | i = 1, \cdots, k\}$ and then its corresponding points pairs are the correctly matched global features.

3.3. **Global motion computation.** All the verified interest points are then brought into equation (1) to get 2N-linear equations. The final function $B = Am$ is in form of matrix, as shown in Equation (6). The global motion parameter is defined as $m = [\,\theta, \Delta X, \Delta Y\,]^T$.

$$B = \begin{bmatrix} x_1 - \tilde{x}_1 \\ y_1 - \tilde{y}_1 \\ \vdots \\ x_N - \tilde{x}_N \\ y_N - \tilde{y}_N \end{bmatrix}, A = \begin{bmatrix} -\tilde{y}_1 & 1 & 0 \\ \tilde{x}_1 & 0 & 1 \\ & \vdots & \\ -\tilde{y}_N & 1 & 0 \\ \tilde{x}_N & 0 & 1 \end{bmatrix} \tag{6}$$

We can solve the above over-determined linear equation with three unknowns. The least-square solution of the equation is defined as $m = (A^T A)^{-1} A^T B$.

4. **Motion compensation.** Global motion composes of intentional camera scan and unwanted jitter. The basic hypothesis of Sage-Husa filter is that scan is intended to move smoothly at low speed towards one direction, while jitter is random in altitude and direction. So, the intentional scan is of low frequency while undesired jitter is of higher frequency. We can get smooth motion vector through filter ($MV_{filter}$), and then jitter is the difference between original and smooth vector, $MV_{jitter} = MV_{filter} - MV_{raw}$.

4.1. **Sage-Husa filter.** Global motion vectors are filtered to get dithering component by Sage-Husa filter. The accuracy of classic Kalman filter reduces and even diffuses, when the distribution of process and observation noise is unknown. However, in the process of Sage-Husa filter, observation data estimate the predictive value automatically and correct the process noise and observation noise simultaneously, which can reduce model error and improve accuracy. The linear discrete system model is constructed as follows:

$$S(k) = F \cdot S(k-1) + w$$
$$Z(k) = H \cdot S(k) + v$$
$$F = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{7}$$

where $S(k) = [x(k), y(k), dx(k), dy(k)]^T$ is state in horizontal and vertical displacement and their instantaneous velocity. $Z(k) = [x(k), y(k)]^T$ is observation vector. $w - N(0, Q)$ and $v - N(0, R)$ represent the process noise and observation noise respectively [9]. $F$ is state transition matrix, and $H$ is observation matrix. The process of Sage-Husa state prediction and update is as the following steps:

1) Build the state prediction equation;

$$S(k|k-1) = F \cdot S(k-1|k-1) \tag{8}$$

2) Predict covariance matrix $P(k|k-1)$;

$$P(k|k-1) = F \cdot P(k-1|k-1) \cdot F^T + Q \tag{9}$$

3) Build the state update equation and compute the result of optimal estimated value;

$$S(k|k) = S(k|k-1) + K_g(k) \cdot \varepsilon(K) \tag{10}$$

$$K_g(k) = P(k|k-1)H^T(HP(k|k-1)H^T + R)^{-1} \tag{11}$$

$$\varepsilon(k) = Z(k) - H \cdot S(k|k-1) \tag{12}$$

Where, $K_g(k)$ is Kalman gain, and $\varepsilon(k)$ is innovation sequence.

4) Update the filter error matrix $P(k|k)$ of $S(k|k)$.

$$P(k|k) = (I - K_g(k) \cdot H) \cdot P(k|k-1) \tag{13}$$

In the above process, we cannot estimate $Q$ and $R$ when they are unknown [10]. And when $Q$ and $R$ lose positive definiteness, the stability and convergence reduce or even diffuse.

4.2. **Improved adaptive filter.** To prevent filter diffusion, the adaptive filter (AF) based on Sage-Husa is improved by correcting covariance matrix $P(k|k-1)$ to increase performance of tracking abrupt variation. We determine whether the filter is divergent using the property of innovation sequence $\varepsilon(k)$.

$$\varepsilon(k)^T \varepsilon(k) \leq \gamma \cdot Trace[HP(k|k-1)H^T + R] \tag{14}$$

Here, $\gamma$ is an adjustable coefficient and $\gamma > 1$. If formula (14) holds, the filter is in normal state and the optimal estimated value is obtained by equation (9). Otherwise, the actual error is $\gamma$ times more than theoretical estimate value, and the filter will diverge. So, we make advantage of the fading factor [11] of Kalman filter to improve the convergence of Sage-Husa filter. $P(k|k-1)$ is corrected by weighted coefficient $C(k)$ as follows.

$$P(k|k-1) = C(k)FP(k-1)F^T + Q \tag{15}$$

$$C(k) = \frac{\varepsilon(k)^T \varepsilon(k) - Trace[HQH^T + R]}{Trace[H \cdot F \cdot P(k) \cdot F^T \cdot H^T]} \tag{16}$$

The weighted coefficient $C(k)$ is designed to modify $P(k|k-1)$ in equation (9) if the filter lose stability and convergence when Q and R lose positive definiteness. In real applications, the variance matrix Q of process noise and R of observation noise are unknown, they can be predicted as follows.

$$R(k) = (1 - d_k)R(k-1) + d_k[\varepsilon(k) \cdot \varepsilon(k)^T - H \cdot P(k) \cdot H^T] \tag{17}$$

$$\begin{aligned} Q(k) &= [1 - d(k)]Q(k-1) \\ &+ d(k)[K(k) \cdot \varepsilon(k) \cdot \varepsilon(k)^T K(k)^T] \\ &+ d(k)[P(k|k) - F \cdot P(k-1|k-1) \cdot F^T] \end{aligned} \tag{18}$$

Here $d(k) = (1-b)/(1-b^k)$; $b$ is forgetting factor and $0 < b < 1$. So, the process of proposed adaptive Sage-Husa filter is as follows.

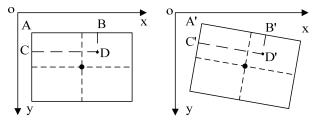Step 1) Build state prediction equation by equation (8);

Step 2) Predict Q and R by equation (17) and (18);

Step 3) Judge the filter divergence by equation (14); if it holds, $P(k|k-1)$ is predicted by equation (9); otherwise, it is corrected by equation (15)-(16);

Step 4) Build state-update equation and compute optimal estimated value by equation (10)-(12);

Step 5) Update $P(k|k)$ by equation (13) and continue step 2).

4.3. **Fast motion compensation.** Motion sequence is smooth through the above filter, and the jitter component is computed as the difference between original motion vector and smooth vector, that is $MV_{jitter} = MV_{filter} - MV_{raw}$ . The jitter component is taken into similarity model (1) to compute new pixels in current image. This point-to-point pixel computation is time-consuming. So, the fast linear compensation is proposed to reduce computation cost.



(a) Before rotation     (b) After rotation

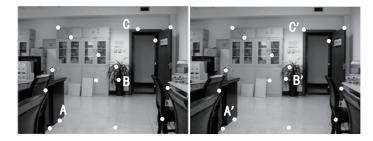FIGURE 3. Rotation-invariant structure of image

In Fig.3(a), $A(x_A, y_A)$ , $B$ ,$C$ and $D$ are four vertices of a rectangle. Based on the linear storage structure of image, their relative position does not change after linear transformation. So, new pixel $D'$ in Fig.3(b) can be calculated as:

$$\begin{aligned} x'_D &= x'_B + (x'_C - x'_A) \\ y'_D &= y'_B + (y'_C - y'_A) \end{aligned} \tag{19}$$

This simple algebraic addition can avoid matrix multiplication and reduce time-cost greatly. The compensation is made by two ways. If pixel is at first row and first column, similarity transformation (1) is made to get its new coordinates; for all the other pixels, coordinates addition is made with formula (19).

5. **Experimental results.** In this section, robustness of global motion estimation, convergence of adaptive motion filter and validity of linear compensation is compared to demonstrate the performance of proposed algorithm. We select 80 adjacent frames with the size of $640 * 480$ from the test video on a moving camera. The camera scans slowly in horizontal direction and the platform dithers in horizontal and vertical direction.



(a) Point selection     (b) Point matching and validation

FIGURE 4. Interest points selection and validation

5.1. **Global motion estimation.** Fig.4 is the result of point selection, matching and validation. The points A, B and C are validated as mismatching points through distance criterion. By deleting these mismatched points, Global Motion Estimation (GME) is improved.

The validated motion is compared with direct motion which is computed by points without distance validating. We select two videos and give comparison in average time per-frame and average error with the real motion. In Table 1, the average error between validated motion and real motion is below half a pixel, which reduces 80% comparing to the error of direct motion. The accuracy improves greatly while the processing speed can also achieve real time realization.

TABLE 1. Results of time and accuracy of GME

| Comparison results | Global Motion | Average Error(pixel) | Time (ms) |
|---|---|---|---|
| **Direct(1)** | 0.0982 −20.0360 11.0341 | 1.6 | 34.45 |
| **Validated(1)** | 0.0970 −18.5840 10.6860 | 0.3 | 35.25 |
| **Direct(2)** | 1.0200 −0.0223 1.0035 | 1.7 | 31.89 |
| **Validated(2)** | 1.0000 −1.0190 0.1927 | 0.4 | 32.10 |

5.2. **Results of motion smooth.** Fig.5 shows convergence of our proposed AF (Adaptive filter) versus Kalman filter at different process noise Q. The original horizontal motion curve 5(a) accumulates at a steady increase because of camera scan. Curve 5(b) fluctuates

at zero-position due to jitter in vertical direction. As Fig.5 shows, different process noise Q in Kalman filter affects smooth result obviously. A Large Q shows no effect in smoothing; while a small Q leads to over smooth deviating from real motion, and even results in filter divergence. It can be seen that the adaptive filter can smooth jitter component as well as track camera scan effectively.



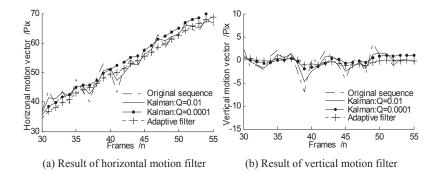(a) Result of horizontal motion filter          (b) Result of vertical motion filter

FIGURE 5. Comparison between AF and Kalman

Fig.6 shows convergence of our proposed AF (Adaptive filter) versus Motion Attenuation. It can be seen that attenuation filter has a time-delay and its smooth effect is not obvious.
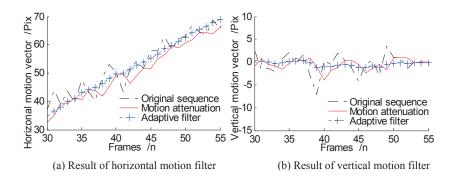


(a) Result of horizontal motion filter          (b) Result of vertical motion filter

FIGURE 6. Comparison between AF and Attenuation

5.3. **Result of linear compensation.** According to experimental statistic data, linear compensation time cost is 4.025ms per-frame, which reduces 76.2% comparing with 17.633mspf by traditional compensation method.
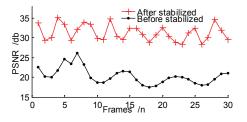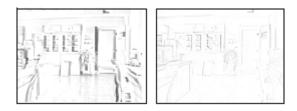


FIGURE 7. Comparison of PSNR

The $PSNR$ (Peak Signal to Noise Ratio) is used to test the global fidelity of inter-frames. From Fig.7, we can see that the $PSNR$ after stabilization is increased greatly, which means the difference between frames is reduced and the video is stabilized.

Fig.8 gives difference image from stabilized video and original video. The difference gray value is not zero due to camera scan. Comparing to original difference, the stabilized difference reduces greatly after removing jitter motion with compensation.



(a)Original difference    (b) Stabilized difference

FIGURE 8. Comparison of image difference

6. **Conclusions.** An efficient digital image stabilization algorithm is presented, capable of real-time performance. The global motion estimation improves by interest points validation and motion is filtered based on adaptive motion filter, which is independent of the noise variance matrix and has fast convergence. The algorithm is applicable to hand-held cameras and surveillance system in moving vehicles, which need to track scan and reduce jitter. And, for more complex applications with camera zoom or shaky, we plan to use a complex parametric motion model for global motion estimation and design a more effective motion filter for motion compensation.

**REFERENCES**

[1] J. Chu, L. H. Du, L.F. Wang, C.H. Pan, Local background-aware target tracking, *Acta Automatica Sinica*, vol. 35, no. 12, pp. 1985-1995 2012.
[2] S. K. Kim, S. J. Kang, Feature point classification based global motion estimation for video stabilization, *IEEE Trans. on Consumer Electronics*, vol. 59, no. 1, pp. 267-272, 2013.
[3] M. H. Shakoor, M. Moattari, A new fast method for digital image stabilization, *International Conference on ACTE*, vol. 2, pp. 350-354, 2010.
[4] D. Y. Huang, C. H. Chen, W. C. Hu, S. C. Yi, Y. F. Lin, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 3, no. 3, pp. 282-296, 2012.
[5] C. T. Wang, H. K. Jin, Robust digital image stabilization using the Kalman filter, *IEEE Trans. on Consumer Electronics*, vol. 55, no. 1, pp. 6-14, 2009.
[6] K. Y. You, L. H. Xie, Kalman filtering with scheduled measurements, *Journal of IEEE Trans. on Signal Processing*, vol. 61, no. 6, pp. 1520-1530, 2013.
[7] D. O. Kim, R. H. Park, New image quality metric using the Harris response, *Journal of IEEE on Signal Processing Letters*, vol. 16, no. 7, pp. 616-619, 2009.
[8] J. J. Zhu, B. L. Guo, Electronic image stabilization system based on global feature tracking, *Journal of Systems Engineering and Electronics*,vol. 19, no. 2, pp. 228-233, 2008.
[9] C. Y. Zhang, Approach to adaptive filtering algorithm, *Journal of Acta Aeronautica et Astronautica Sinica*, vol. 19, no. 7, pp. 96-99,1998.
[10] R. L. Kirlin, M. Alireza, Robust adaptive Kalman Filtering for systems with unknown step inputs and non-Gaussian measurement errors, *Journal of IEEE Trans. on Acoustrics*, vol. 34, no. 2, pp. 252-263, 1986.
[11] J. S. Xu, Y. Y. Qin, R. Peng, New method for selecting adaptive Kalman filter fading factor, *System Engineering and Electronics*, vol. 26, no. 11, pp. 1152-1154, 2004.