An Aerial Video Stabilization Method Based on SURF Feature

Hao WU [1] and Shao-Yang HE [1]

[1]*School of Information and Electronics, Beijing Institute of Technology*

Abstract: The video captured by Micro Aerial Vehicle is often degraded due to unexpected random trembling and jitter caused by wind and the shake of the aerial platform. An approach for stabilizing the aerial video based on SURF feature and Kalman filter is proposed. SURF feature points are extracted in each frame, and the feature points between adjacent frames are matched using Fast Library for Approximate Nearest Neighbors search method. Then Random Sampling Consensus matching algorithm and Least Squares Method are used to remove mismatching points pairs, and estimate the transformation between the adjacent images. Finally, Kalman filter is applied to smooth the motion parameters and separate Intentional Motion from Unwanted Motion to stabilize the aerial video. Experiments results show that the approach can stabilize aerial video efficiently with high accuracy, and it is robust to the translation, rotation and zooming motion of camera.

1 Introduction

Videos of specified target area can be acquired flexibly and efficiently by Micro Aerial Vehicle (MAV) which is wildly used in military and civilian fields. Because of its small size, lightweight and poor stability, MAV jitters easily because of wind blows and mechanical vibration, which results in the unstability of image sequence captured by MAV camera. This unstability of image sequence is harmful to the subsequent image processing. Therefore it's necessary to stabilize the image sequence in order to reduce the influence of random trembling of the image system on MAV. Compared to Mechanical Stabilization and Optical Stabilization, Electronic Stabilization has the advantage of high accuracy, low-power consumption, small volume and low-cost, and Electronic Stabilization has become the most important video stabilization technique.

Electronic Stabilization consists of two processes: Global Motion Estimation and Motion Compensation. Global motion estimation estimates the global motion vector between frames, and removes the interference of local motion at the same time. Motion compensation separates the intentional motion from random jitter in the global motion obtained by global motion estimation, acquires compensation vector, and moves the frame in the opposite direction of the compensation vector equivalently to obtain stabilized video.

Global motion estimation acquires transformation relation between frames by using gray information or feature points. The methods using gray information directly include Block Matching Algorithm[1], Bit Plane Algorithm[2] and Gray Projection Algorithm. These methods have the advantages of fast computation speed and accurate estimation for translation motion. But they are sensitive to illumination change, and could not estimate rotation and zooming motion. The methods based on feature points mainly use the results of feature points matching to obtain the transformation relation between frames, and the feature points generally include edge points, Harris corner points and SIFT feature points[3]. These methods have robustness, and can estimate motion parameters well. In these methods, the performance of SIFT algorithm is outstanding, and SURF algorithm which derives from SIFT is enhanced in the computation speed. Therefore, SURF algorithm is applicable for the real-time requirements of MAV video stabilization.

Motion compensation is mainly used to adjust the global motion parameters, and keep the intentional motion of camera. The motion compensation methods include mean filtering[4], curve fitting, Gaussian Mixture filtering[5] and Kalman filtering[6]. Kalman filtering which smooths the motion parameters by low-pass filtering the motion vectors is outstanding in video stabilization.

According to the advantages of SURF algorithm and Kalman filter, in this paper SURF is used to estimate the motion vector between frames, and Kalman filter is adopted to adjust the similar transformation array to estimate the intentional motion in order to stabilize video.

2 SURF Feature Extraction and Matching

SURF Feature is a rapid local feature points detection algorithm which is proposed by Bay based on SIFT feature[7]. SURF speeds up by using integral image and Boxlets filtering to approximate the Gauss-Laplace second-order differential response of image in order to simplify the process of different size convolution to several additions and subtractions. The contrast experiments which were done by Bauer indicated that the speed of SURF is enhanced three times compare to SIFT, and the key performances such as repeatability and resolving ability are equivalent[8].

SURF feature has rotational invariance by assigning main orientation for each feature point. First, select a certain number of sample points in the neighbourhood of the feature point, do statistic of their Harr wavelet filtering response components along x-axis and y-axis, named d_x and d_y, and map to a point (d_x, d_y) in response coordinates, as shown in Figure 1. Then, rotate the fan-shaped window whose opening angle is 60°around the origin with fixed step size as 15°. The accumulation of all the response values in the fan-shaped window is regarded as the centering direction response value of the window. Assign the max response value of all directions as the main orientation of the SURF feature point.

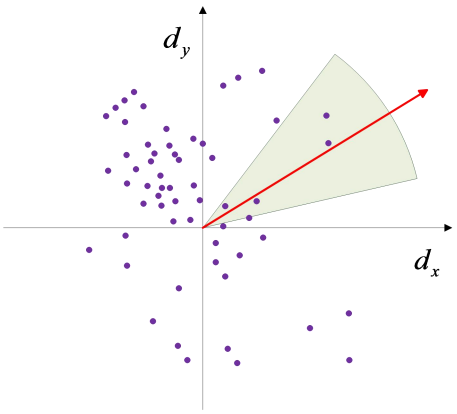

Figure 1.Orientation Assignment

SURF descriptor is constructed by doing statistics to the distribution of the sample points Haar wavelet effect. First, construct a 4×4 square mesh region in the neighbourhood of the feature point. Then, rotate the region to the main orientation. Next, select 5×5 sample points in each grid, and apply Haar wavelet filtering to every sample point to calculate their response d_x and d_y along x-axis and y-axis. As a result, a four-dimensional description vector is constructed in each grid, that is

$$v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \qquad (1)$$

Combine these 16 description vectors to a 64-dimensional SURF descriptor of the feature point, as shown in Figure 2.

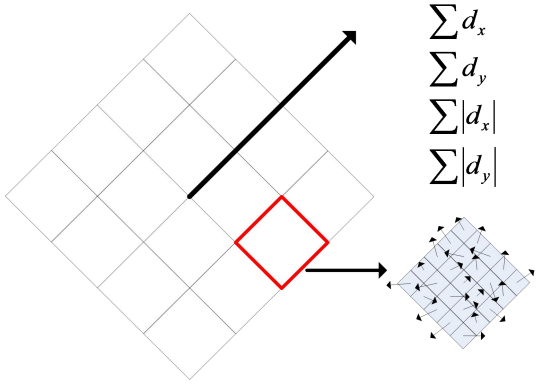

Figure 2.Build SURF Descriptor

3 Global Motion Estimation

According to the interframe feature points matching relations acquired by Fast Library for Approximate Nearest Neighbors searching algorithm, the following motion model is used to describe the translation, rotation and zooming motion of image sequence.

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = M \begin{bmatrix} x_j \\ y_j \\ 1 \end{bmatrix} \qquad (2)$$

$$M = \begin{bmatrix} s\cos\theta & -s\sin\theta & \Delta x \\ s\sin\theta & s\cos\theta & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \qquad (3)$$

Where (x_i,y_i) and (x_j,y_j) are the matching points coordinates of t_i and t_j frames, M is the similar transformation matrix, ($\triangle x$, $\triangle y$) is the translation motion vector between frames, θ is the rotation angle between frames, and s is the zoom coefficient of camera.

In the process of SURF feature point extraction and matching, some outliers are produced due to mismatching and local motion. If this original feature points set which contains outliers is used directly to estimate the translation, rotation and zooming motion parameters, the consequence is completely wrong or has greater error.

In order to eliminate the influence of these outliers, RANSAC (RANdom SAmple Consensus) algorithm is used to modify the matching points pairs set[9], and then Least Square Method is used to estimate model parameter. RANSAC uses minimum data set to estimate model parameter repeatedly for seeking the data set supported by most estimated model to obtain inlier set. The steps are as follows

Step 1 Sample n times repeatedly, randomly pick up two matching point pairs to form a sample set P each time, calculate similarity transformation matrix M, then calculate the error of matching points pair in complementary set Q based on M, select points whose error less than certain threshold t to form inlier set.

Step 2 Select a inlier set which contains greatest number of matching points pairs as the modified matching points pairs set.

Step 3 Use Least Square Method to calculate similarity transformation matrix parameters based on the modified matching points pairs set.

This method eliminates the influence of mismatching and moving foregrounds by removing the outliers that do not meet the criteria, and obtains precise translation, rotation and zooming motion parameters.

4 Motion Compensation

According to the processing sequence of observations, Kalman filtering can be divided into Fixed Point Filtering, Fixed Delay Filtering and Recursive Filtering. Considering the calculation speed requirement, this paper adopts Recursive Filtering algorithm to modify the estimated global motion in order to remove random jitter and reserve the intentional motion.

Variable θ and s respectively describe the rotation and zooming motion of aerial camera. For most aerial video, the stability of θ and s is commonly influenced by Gaussian white noise. Therefore, the dynamic model is set as follows,

$$\theta^{k+1} = \theta^k + N(0,\sigma_\theta) \tag{4}$$

$$s^{k+1} = s^k + N(0,\sigma_s) \tag{5}$$

Where $N(0, \sigma_\theta)$ and $N(0, \sigma_s)$ are Gaussian white noise.

Variables $\triangle x$ and $\triangle y$ describe the translation motion of camera. For most aerial video, $\triangle x$ and $\triangle y$ reflect the intentional motion of the camera, and the variation of speed follows certain random distribution. Therefore, the dynamic model is as follows

$$\begin{bmatrix}\Delta x\\ v_x\end{bmatrix}^{k+1} = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}\Delta x\\ v_x\end{bmatrix}^{k} + \begin{bmatrix}0\\ N(0,\sigma_x)\end{bmatrix} \tag{6}$$

$$\begin{bmatrix}\Delta y\\ v_y\end{bmatrix}^{k+1} = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}\Delta y\\ v_y\end{bmatrix}^{k} + \begin{bmatrix}0\\ N(0,\sigma_y)\end{bmatrix} \tag{7}$$

Where vx is the variation of horizontal movement $\triangle$x, vy is the variation of vertical movement $\triangle$y, N(0,σx) and N(0,σy) are Gaussian white noise.

In conclusion, the Kalman state space model of camera is

$$\begin{bmatrix}\theta\\ s\\ \Delta x\\ v_x\\ \Delta y\\ v_y\end{bmatrix}^{k+1} = \begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\theta\\ s\\ \Delta x\\ v_x\\ \Delta y\\ v_y\end{bmatrix}^{k} + \begin{bmatrix}N(0,\sigma_\theta)\\ N(0,\sigma_s)\\ 0\\ N(0,\sigma_x)\\ 0\\ N(0,\sigma_y)\end{bmatrix} \tag{8}$$

Where σθ, σs, σx and σy are mutually independent and decided by the smoothness of the camera intentional motion. The greater the variance of the observation noise, the greater the variability of state variables is, which results in the stronger the randomness of intentional motion and the less stability of the compensated image sequence. Otherwise, if the variance is 0, the state variable is immutable, and it can be compensated completely.

For the translation, rotation and zooming motion, the observation equation of state space is

$$\begin{bmatrix}\theta\\ s\\ \Delta x\\ \Delta y\end{bmatrix}^{k+1} = \begin{bmatrix}\theta\\ s\\ \Delta x\\ \Delta y\end{bmatrix}^{k} + \begin{bmatrix}N(0,\sigma_{obs,\theta})\\ N(0,\sigma_{obs,s})\\ N(0,\sigma_{obs,x})\\ N(0,\sigma_{obs,y})\end{bmatrix} \tag{9}$$

Where N(0, σobs,θ), N(0, σobs,s), N(0, σobs,x) and N(0, σ obs,y) are observation Gaussian white noise. σ obs, θ , σ obs,s, σ obs,x and σ obs,y are mutually independent, and describe the variability of the interframe unintentional motion. Their influences are opposite to the process noise, the greater the variance of the observation noise, the greater the unintentional movement variability is, and the more stable the compensated image is. Otherwise, if the variance is 0, the observation variable varies randomly, and the image is uncompensated completely.

5 Experiments and Analysis

To verify the accuracy and usefulness of the method which is proposed in this paper, we made a set of air-to-ground video whose resolution is 640×480.

Extract SURF feature points separately from the reference frame and the current frame, and get the matching relations between feature points by using FLANN searching algorithm, then eliminate outliers by RANSAC. The consequence is shown in Figure 3, which indicates SURF feature has better recognition capability, and can get satisfying matching results.

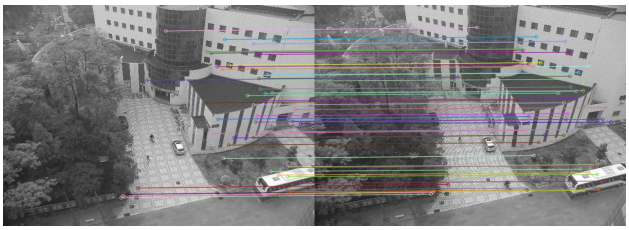

Figure 3.SURF feature detection and matching

After getting the translation, rotation and zooming motion parameters of every frame relative to the reference frame, the parameters are smoothed by Kalman filter to separate the intentional motion and random jitter. Figure 4 and Figure 5 are the movement curves of horizontal motion and vertical motion after Kalman filtering, in which the dotted line is the original estimated parameters curve and solid line is the motion parameters curve after Kalman filtering. The experiment results show that the original data varied intensely, and the differences between frames are obvious. But the movement curves become smoother after Kalman filtering, and keep the original movement tendency, which indicates the stabilization performances are very good.

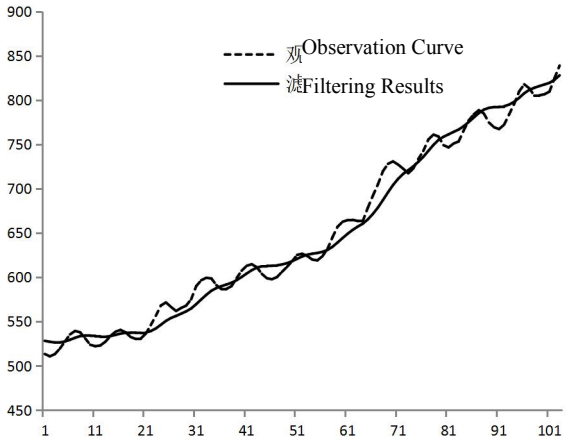

Figure 4.Measured Curve and Filtered Result for △x

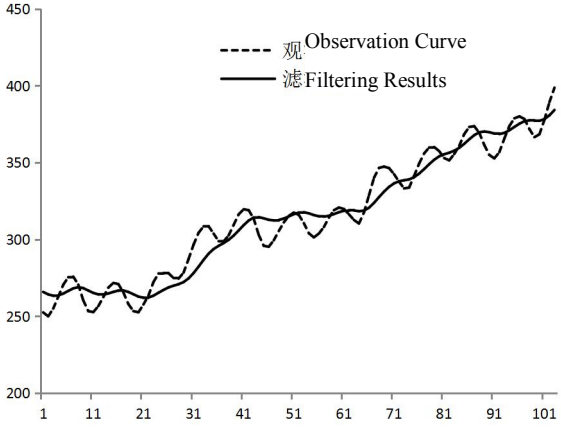

Figure 5. Measured Curve and Filtered Result For △y

Based on the consequences of Kalman filtering, transform the coordinate of each image in the video to acquire stabilized aerial video, as shown in Figure 6. For the sake of observing easily, select one frame in every five frames. The (a) to (d) are the original video frames, and the (e) to (h) are the stabilized video frames, from which we can see that the Horizontal and Vertical motions are all stabilized well.

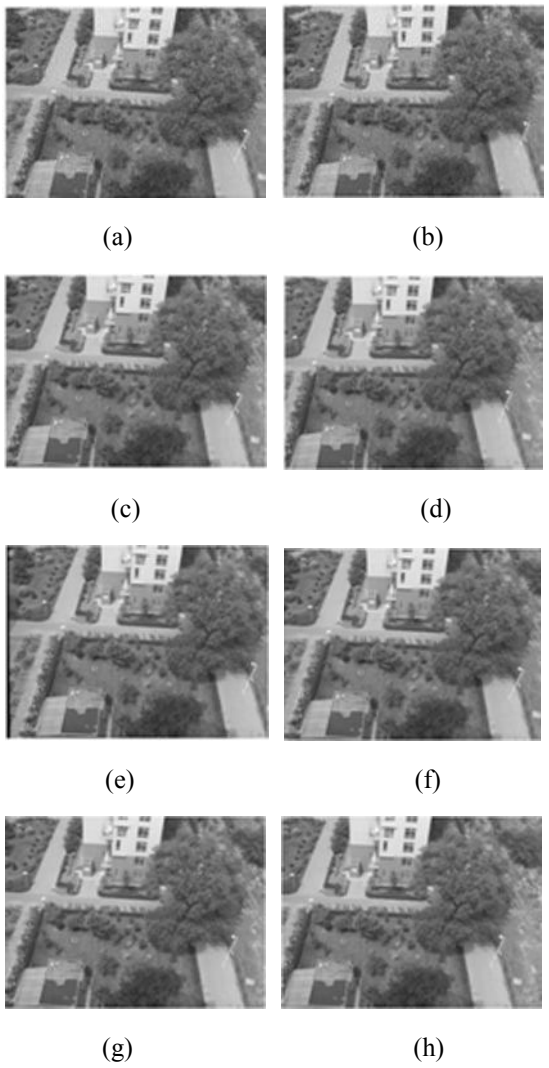

Figure 6. Experimental Results of Aerial Video Stabilization

6 Conclusion

According to the movement feature of the aerial camera on MAV, this paper proposed a method based on SURF feature and Kalman filter to acquire stabilized aerial video. SURF algorithm is used to extract and match feature points, then RANSAC and Least Square method are used to get the similar transformation matrix between frames, which can eliminate the influence of the mismatching points pair and local movements to acquire high-accuracy translation, rotation and zooming motion parameters. On this basis, Kalman filter is used to smooth the motion parameters, and separate the intentional motion and random jitter to stabilize the aerial video.

Experiments results show that the method has high calculation accuracy and can satisfy most requirements of aerial video stabilization.

The deficiency of this paper is that three-dimensional information of the scene is not considered, which lead to the limitations of motion model setting and computational accuracy. The follow-up work will be the intensive study of stabilization method based on three-dimensional information.

References

1. X. Li-dong, L. Xing-gang. Digital image stabilization based on circular block matching. IEEE Transactions on Consumer Electronics, **52**, 566(2006)
2. S. Ko, Sung-Hee, K. Lee. Digital image stabilizing algorithms based on bit-plane matching. IEEE Transactions on Consumer Electronics, **44**, 617 (1998)
3. D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, **60**, 91(2004)
4. Z. Yu-wen, W. Jun, J. Yun-de. A Feature Tracking Based Method for Image Stabilization. Transactions of Beijing Institute of Technology, **23**, 596(2003)
5. Z. Min, Z. Meng, J. Yun-de, Wang Jun. Image Stabilization Based on Adaptive Gaussian Mixture Model. Transactions of Beijing Institute of Technology, **24**, 897(2004)
6. S. Erturk. Image sequence stabilization based on Kalman filtering of frame positions. Electronics Letters, **37**, 1217(2001)
7. H. Bay, T. Tuytelaars, L. Van Gool. SURF: Speeded Up Robust Features. in Proc. of European Conference on Computer Vision, 404(2006)
8. J. Bauer, N. Sunderhauf, P. Protzel. Comparing several implementations of two recently published feature detectors. Proceedings of Proc. of the International Conference on Intelligent and Autonomous Systems(2007)
9. Y. Shen, P. Guturu, T. Damarla Buckles, B. Namuduri K. Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework [J]. IEEE Transactions on Consumer Electronics, **55**, 1714(2009)