Antiblurry Dejitter Image Stabilization Method of Fuzzy Video for Driving Recorders

Jing-Ying Xiong[1,2], Ming Dai[1], Chun-Lei Zhao[1], and Ruo-Qiu Wang[2]
[1] Key Laboratory of Airborne Optical Imaging and Measurement, Changchun Institute of Optics, Fine Mechanics and Physics Chinese Academy of Sciences, Dong Nan Hu Road 3888, Changchun, China, 130033
[e-mail: xiong_ing@163.com]
[2] University of Chinese Academy of Sciences, Changchun 130033
[e-mail: daim@vip.sina.com]
*Corresponding author: Jing-Ying Xiong

Received December 14, 2016; revised February 27, 2017; accepted March 20, 2017; published June 30, 2017

Abstract

Video images captured by vehicle cameras often contain blurry or dithering frames due to inadvertent motion from bumps in the road or by insufficient illumination during the morning or evening, which greatly reduces the perception of objects expression and recognition from the records. Therefore, a real-time electronic stabilization method to correct fuzzy video from driving recorders has been proposed. In the first stage of feature detection, a coarse-to-fine inspection policy and a scale nonlinear diffusion filter are proposed to provide more accurate keypoints. Second, a new antiblurry binary descriptor and a feature point selection strategy for unintentional estimation are proposed, which brought more discriminative power. In addition, a new evaluation criterion for affine region detectors is presented based on the percentage interval of repeatability. The experiments show that the proposed method exhibits improvement in detecting blurry corner points. Moreover, it improves the performance of the algorithm and guarantees high processing speed at the same time.

Keywords: Electronic image stabilization, antiblurry dejitter, feature detection, evaluation criterion, binary feature description

1. Introduction

Vehicle cameras can be used to capture any activity within the range of vision; however, the images may contain dithering or fuzzy frames. In general, by detecting and compensating the inter-frame motion using image processing methods, an electronic image stabilization (EIS) technique is utilized for disturbance attenuation. [1-3] Remote sensing, visual surveillance, walking robots, civil infrastructure, and unmanned aerial vehicles are examples of such applications that operate in dynamic environments.

A driving recorder is used as a record of a vehicle to provide evidence for traffic accidents or to memorialize the journey on private trips. In view of the restrictions on volume and the high cost of driving recorder models, adopting either mechanical or optical stabilizations method is less appropriate compared with implementing the EIS algorithms.

Video images captured by vehicle cameras often contain blurry or dithering frames due to inadvertent motion from bumps in the road or by insufficient illumination during the morning or evening. The human eye has persistence and continuity, enabling it to determine the movements of objects in videos when watching a blurry video without jitter; however, the human eye's visual coherence cannot be guaranteed if dithering is added, which greatly reduces the perception of objects expression and recognition. To solve the problem of dithering videos from vehicle cameras, feature points are selected as a characteristic quantity to obtain image information and feature descriptor performs matching.

Feature point detection employing feature descriptors for motion estimation has been widely explored [4,5]. The methods of descriptor matching, discriminated by Euclidean and Hamming distances, are roughly able to divide these descriptors into two categories, scale-invariant feature transform (SIFT)-like algorithms and binary feature descriptor algorithms.

In SIFT-like algorithms, to overcome scaling and rotation inefficiency, a local scale-invariant feature [6,7] was developed, which delivered competitive performance. Developed by Krystian Mikolajczyk et al., the method that used a gradient location orientation histogram [8] was considered more spatial regions for the histograms. Furthermore, it reduced the dimensionality of a descriptor by principal components analysis. The speed enhancing feature [9] exhibited similar performance to SIFT [6] while significantly promoting speed. Its feature points were obtained by computing the determinant of a Hessian matrix, whereas its accumulated description was accomplished using Haar-wavelet response.

Considering the computational complexity of feature point matching, binary descriptors and Hamming space have been proposed as substitutes for the exhaustive Euclidean distance calculation algorithms to perform fast similarity searches, where binary codes are defined based on the relationship between a given adjoining pixel and the central reference. The Binary Robust Independent Elementary Feature (BRIEF) [10], the Oriented Fast and Rotated BRIEF (ORB) [11], and the Binary Robust Invariant Scalable Keypoints (BRISK) [12] are well-known representatives. The first two binary descriptors are described by feature vectors that compare the intensity of 512 pairs of pixels after applying Gaussian smoothing. The positions of the pixels are preselected around one patch center. In addition, ORB allocates the vector directions. To build the descriptor bit stream, BRISK designed a template with a set of concentric circles. In reference to the distance from center to center, long-distance pairs were assigned for orientation, and short-distance pairs describe the keypoints. However, the enhancement of binary matching speed results in degradation of its effectiveness.

High-speeds and good performances in stabilization algorithms have a mutual exclusive relationship. In this paper, an antiblurry dejitter electronic image stabilization method (AD) is proposed, which tackles fuzzy vision records problem for vehicles. In contrast to classic binary description algorithms, more quality feature point are detected via utilizing coarse-to-fine inspection policy and discriminative descriptor is established with proven good performance. Further, the accuracy of unintentional motion estimation benefits from the feature points selection strategy. Finally, a new evaluation criterion is proposed to utilize a percentage interval of repeatability to appraise the performance of stabilization methods for videos. Precise timing experiments were conducted, which indicated that the proposed method can overcome dithering of fuzzy videos for driving records effectively.

The proposed method is presented in Section 2. The repeatability experimental setup and results are provided in Section 3.1. In Section 3.2 and 3.3, the proposed method is compared with the other methods in the literature on frame rate and ITF. The feature point selection experiment for unintentional motion estimation is exhibited in Section 3.4. The main conclusions are summarized in Section 4.

2. Description of the Proposed Algorithm

The overall procedure of the proposed video stabilization method is shown in **Fig. 1**. The proposed algorithm consists mainly of two parts: 1) extraction and description of the feature points, and 2) feature point selection strategy for unintentional motion estimation. The task of the extraction and description of feature points is to find more distinctive features from each video frame, whereas the tasks of unintentional motion estimation are eliminating mismatching correspondences to the greatest extent and precisely estimating the virtual geometric transformation between images.

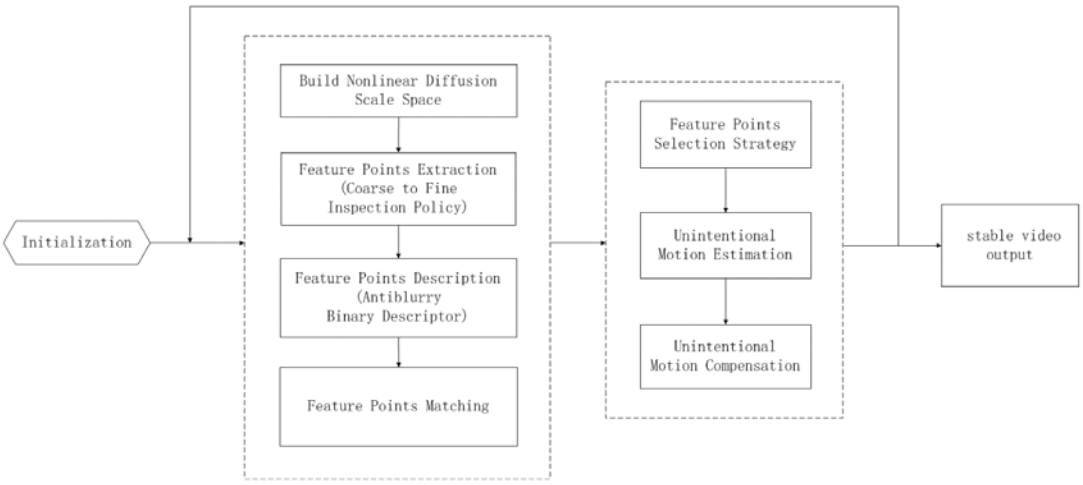


Fig. 1. Video stabilization algorithm process.

2.1. Nonlinear Diffusion Scale Space for Pre-processing

The Gauss-scale space pyramid constructed in SIFT is based on Gauss decomposition, which lead to fuzzy boundaries and sacrifices the detail of the original image. In contrast, the nonlinear diffusion scale space pyramid [13] applies the divergence of a luminance flow function to increase the scale levels and keep the accuracy of the original image for each layer..

Firstly, Gaussian filter is applied to the input frame to control image noise. Secondly, contrast parameter λ is calculated which is 70% of the gradient histogram value of the smoothed input frame. This empirical value brings in general good performances in our experiments. It is probable that for some scenes a more comprehensive analysis of the contrast parameter brings better results. Then the conductivity equation is calculated:

$$g = \frac{1}{1 + \frac{\left|\nabla L_\sigma\right|^2}{\lambda^2}} \tag{1}$$

where function ∇L_σ is the gradient of a Gaussian-smoothed version of original image L and parameter λ is the contrast factor that controls the diffusion level.

Given the set of evolution times t_i, it is straightforward to build the nonlinear scale space, as shown in **Fig. 2**, in an iterative way [13]:

$$\begin{cases} t_i = \frac{1}{2}\sigma_i^2 \\ L_{i+1} = \left[I - \left(t_{i+1} - t_i\right)\sum_{l=1}^{m} A_l\left(L_i\right)\right]^{-1} L_i \end{cases}, \quad i \in [0,1,\ldots,N] \tag{2}$$

where σ is Gaussian standard deviation, L_0 is the original image and used as the base scale level, I is identity matrix, A_l is a matrix that encodes the image conductivities achieved by equation (1) and N is the total number of filtered images.

The nonlinear diffusion filter is proposed to refine the repeatability [14] of the SIFT-like algorithms. However, it is more suitable for fast abstraction and binary description algorithms, and therefore nonlinear diffusion scale space is built to make it liable for high quality feature detection.

Fig. 2. Conductivity images from nonlinear diffusion scale space with several evolution times t_i.

2.2. Coarse-To-Fine Inspection Policy for Feature Detection

The process of detecting the characteristic quantities in each layer begins after establishing the scale pyramid. It has become a celebrated research method for fast feature extraction since the advent of the feature of accelerated segment test (FAST) [15] algorithm. As shown in **Fig. 3**, FAST corner detector uses a circle of 16 pixels (a Bresenham circle of radius 3.7) to classify whether a candidate point p is actually a corner. Each pixel in the circle is labeled from integer number 1 to 16 clockwise. If a set of 12 contiguous pixels in the circle are all brighter than the intensity of candidate pixel p or all darker than the intensity of candidate pixel p, then p is classified as corner.

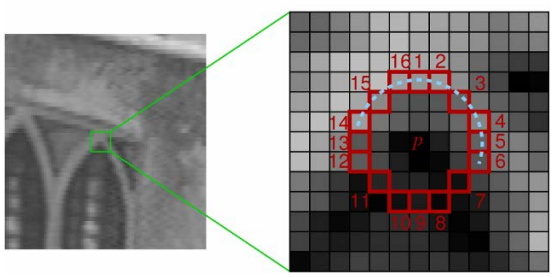


Fig. 3. Principle of FAST corner detection.

However, it fails to account for the magnitude characteristic points easily by utilizing this method, along with the shortage of discriminability. Therefore, except for the exploited nonlinear diffusion filter pyramid toward the Hamming space algorithms, a coarse-to-fine inspection policy is proposed that combined FAST with the strong scale point of the Shi–Tomasi [16] method.

Shi–Tomasi was an improved form of Harris [17-19] that was determined by gray variation of an image. The Taylor expansion of a gray variation of one image can be expressed as:

$$E(u,v)=\sum_{x,y}\left[(u,v)\begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix}\begin{pmatrix} u \\ v \end{pmatrix}\right] \tag{3}$$

Supposing that:

$$T=\sum\sigma\begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} \tag{4}$$

where $E(u, v)$ is the gray-level variation resulted from the translation of image window (u, v); I_x^2, I_xI_y and I_y^2 are second-order gradient components of the gray value of image pixels $I(x, y)$, and σ is the Gaussian smoothing coefficient.

Harris corner response function is defined as equation (5), and a pixel point is a corner when the result of R is larger than a predetermined threshold.

$$R=\lambda_1\lambda_2-k(\lambda_1+\lambda_2)^2 \tag{5}$$

where λ_1 and λ_2 are the eigenvalues of matrix T.

Shi–Tomasi affirmed that the corner determined as a strong corner point once the predetermined threshold was smaller than $\min(\lambda_1,\lambda_2)$. Compared with the judgment condition proposed by Harris, Shi-Tomasi algorithm is more sufficient and the detected strong corner points mostly lying on the intersection between different objects consequently.

However, the original Shi-Tomasi algorithm is a time-consuming operation and does not boast the scale invariance. Firstly, the distribution of potential feature points obtained via recognition using FAST, which passes through a nonmaximum suppression [20]. Secondly, in

order to improve the discrimination of these points, a scale - normalized determinant of σ - Harris is established to test the fineness and to select the top *N* feature points.

$$\begin{cases} \sigma_{i,norm} = \dfrac{\sigma_i}{2^{o^i}} \\ L^i_{Harris} = \sigma^2_{i.norm}\left(L^i_{xx}L^i_{yy} - L^i_{xy}L^i_{xy}\right) \end{cases} \tag{6}$$

where σ is the scale coefficient, L^i_{xx} and L^i_{yy} are the second-order horizontal and vertical derivatives, respectively, and L^i_{xy} is the second-order cross derivative.

A coarse-to-fine inspection policy is proposed for feature extraction, where coarse portion correspond fast detection and fine portion match σ - strong points.

2.3. Antiblurry Binary Descriptor for Feature Description

The BRISK descriptor performs invariant rotation. However, for video shooting using recorders, a little difference exists between two adjacent frames, which means that the increased time in the feature description annihilates the promotion of discriminability by setting up rotating characteristics. In contrast, the specificity of the new descriptor increases if the rotating characteristics of the original BRISK descriptor is removed, as shown in Fig. 4. It turns out that the rotation without BRISK acts better discrimination in the central area.

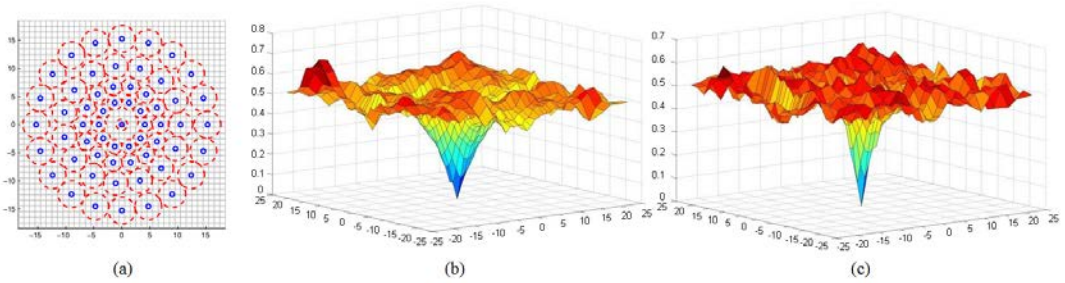


Fig. 4. (a) is the BRISK pattern. (b) and (c) shows the discrimination between the center point and neighboring patches of the original BRISK and the rotation removed BRISK, respectively.

In addition, a threshold for brightness contrast is established in order to have a more discriminant feature descriptor. The value of threshold in equation (7) is set to 8 according to empirical results. This empirical value brings in general good performances in our experiments. It is probable that for some scenes a more comprehensive analysis of the contrast parameter obtains better results. The improved descriptor is called an antiblurry binary descriptor.

$$Diff\left(threshold\right): \; b = \begin{cases} 1, abs\left[I\left(P_i\right) - I\left(P_j\right)\right] > threshold \\ 0, otherwise \end{cases} \tag{7}$$

where $I\left(P_i\right)$ and $I\left(P_j\right)$ are the intensities of point P_i and P_j.

2.4. Feature Points Selection Strategy for Unintentional Estimation

A driving recorder is generally installed at the vehicle ceiling adjacent to the inside rear-view mirror for convenience in observing the status of the front road and recognize various information. It does not recognize the targets in a manner similar to that in object identification because for the stabilization algorithm the random motion of foreground objects is not conducive in accurately estimating the global motion vector.

To eliminate unintentional random motion, the RANSAC [21] and position verification method is employed in the selection of correspondence pairs for affine transformation. In **Fig. 6**, outliers are eliminated by RANSAC via repeatedly subset selection from the observing data.

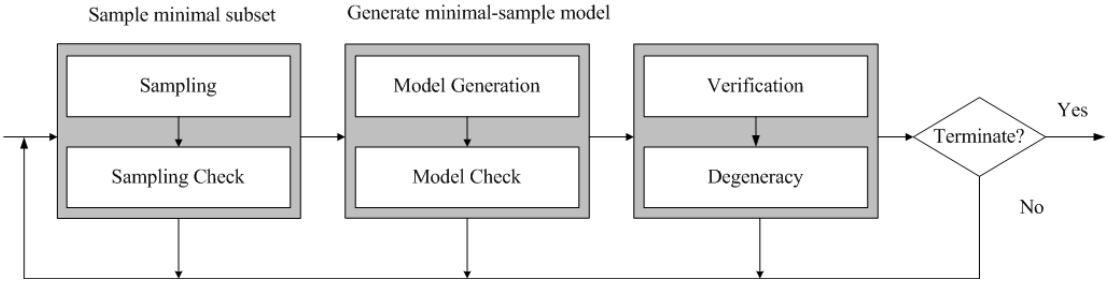


Fig. 6. RANSAC method process.

Furthermore, the displacement distance of the corresponding point in the current frame is usually located at a given interval compared to the certain point of the selected reference frame through the analysis of a great amount of experimental data:

$$\begin{cases} -150 \le k \times (x_2 - x_1) \le 150 \\ -30 \le k \times (y_2 - y_1) \le 30 \end{cases}, \ 1080\text{P}: k = \frac{2}{3}, \ 720\text{P}: k = 1, \ 480\text{P}: k = 1.5 \tag{8}$$

where (x_1,y_1) is the coordinate of a certain point of the selected reference frame, (x_2,y_2) is the coordinate of the corresponding point of the current frame, and k is the resolution coefficient. If the resolution of a video is non-standardized, the resolution coefficient selected is one of similar resolution.

3. Experimental Results and Analysis

The proposed approach was tested on five video sequences and a standard dataset. Four of the video sequences were separately taken by driving recorders in an underground parking garage, in the streets in the morning and evening without adequate light, on the highway with several moving vehicles and at night characterized by dazzling-light effect. The remaining video sequences are news report of smog. The dataset represents a standard evaluation set by Mikolajczyk[1], which is available in the Internet. These video sequences have slight changes in viewpoints, different depths of field, and certain dithering and blurring, as shown in **Fig. 7**.

[1] http://lear.inrialpes.fr/people/Mikolajczyk

(a) (b) (c) (d) (e) (f)

Fig. 7. Examples of video sequences. Both (a) evening (Video 1) and (b) night (Video 2) have 896 × 504 pixel resolution. The screenshot size of (c) smog (Video 3) is 716 × 532 pixels, and the resolution of (d) park (Video 4) is 640 × 480 pixels. (e) highway (Video 5) has 806×452 resolution. The width and height of (f) trees is 1000 and 700 pixels, respectively.

In this section, the testing of the proposed approach is presented in four series of experiments using state-of-the-art algorithms via C++ and OpenCV-based implementation. In the first series of experiments, the repeatability was compared and repeatability interval to distinguish the effectiveness of the feature detection and description. The second series of experiments was mainly related to the execution time and the operation of the collectivity matching. In the third series of experiments, by utilizing inter-frame transformation fidelity (ITF) [22], the practicability of the proposed method was tested and verified. The illustration of feature points selection strategy for unintentional estimation is in the fourth series experiments. Kalman Filter and Mean Filter are adopted for motion filtering. The specification of the testing PC is Intel i3 core with 3.30-GHz CPU and 4-G RAM.

3.1. Detection Quality

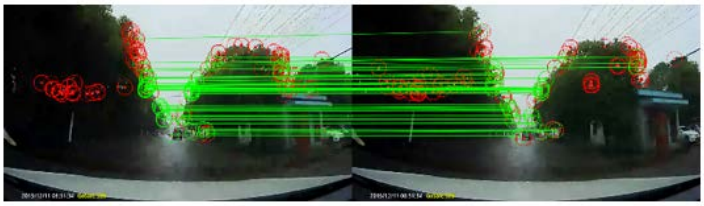
3.1.1 Features Detection of fuzzy images

The fuzzy images for feature detection in this experimental stage are the screenshots from Video 1 to 5. The value of threshold in equation (7) is set to 8 according to empirical results. To ensure approximately the same number of features abstracted in each algorithm thresholds of detectors was adjusted for fairness in the experiments.

From **Fig. 8** to **Fig. 12**, the red circles represent single points, while green circles and lines stand for correct matching point pairs. That means more low quality or marginal feature were detected in SIFT [6], SURF [9], ORB [11] and BRISK [12], which were prone to disappearing, be distorted or difficult to match. On the contrary, by adopting the proposed method AD, most of the keypoints located at the intersection between different objects and less internal distribution of objects, which emerged more saliency and were liable for matching.

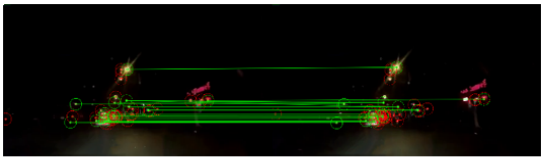
(a)SIFT (b)SURF

(c)ORB (d)BRISK

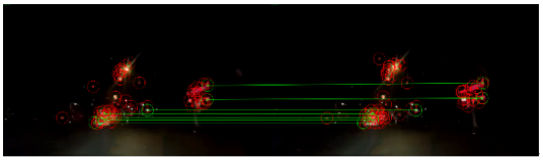


(e) AD (the proposed method)

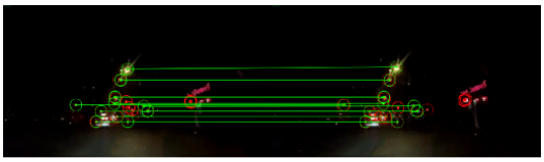
Fig. 8. The results of feature detection and matching for fuzzy images in Video 1. The red circles represent single points or mismatch correspondences, while green circles and lines are correct matching points. The number of keypoints in this group tests is around 280.



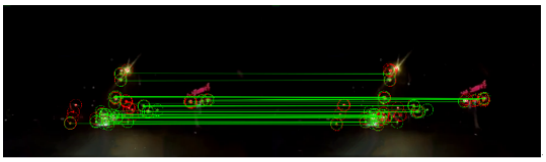
(a)SIFT



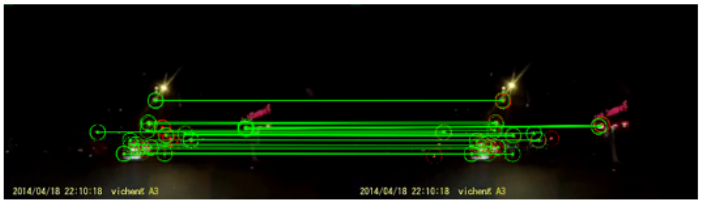
(b)SURF



(c)ORB

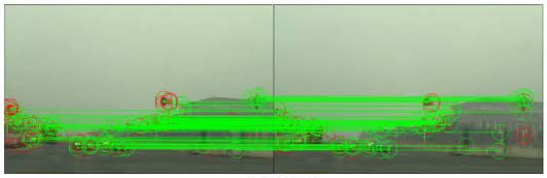


(d)BRISK

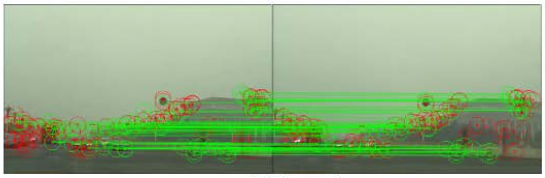


(e) AD (the proposed method)

Fig. 9. The results of feature detection and matching for fuzzy images in Video 2. The red circles represent single points or mismatch correspondences, while green circles and lines are correct matching points. The number of keypoints in this group tests is around 76.



(a)SIFT



(b)SURF

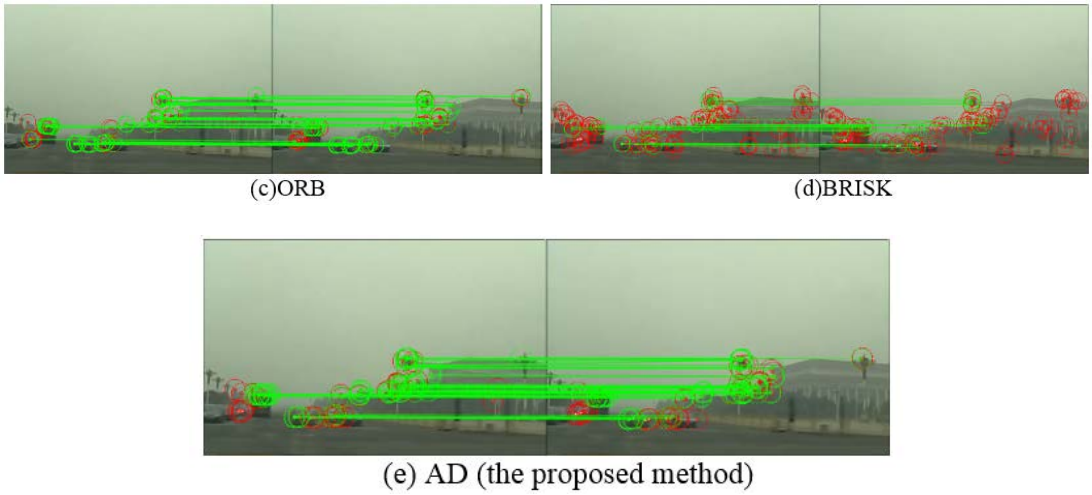


Fig.10. The results of feature detection and matching for fuzzy images in Video 3. The red circles represent single points or mismatch correspondences, while green circles and lines are correct matching points. The number of keypoints in this group tests is around 128.

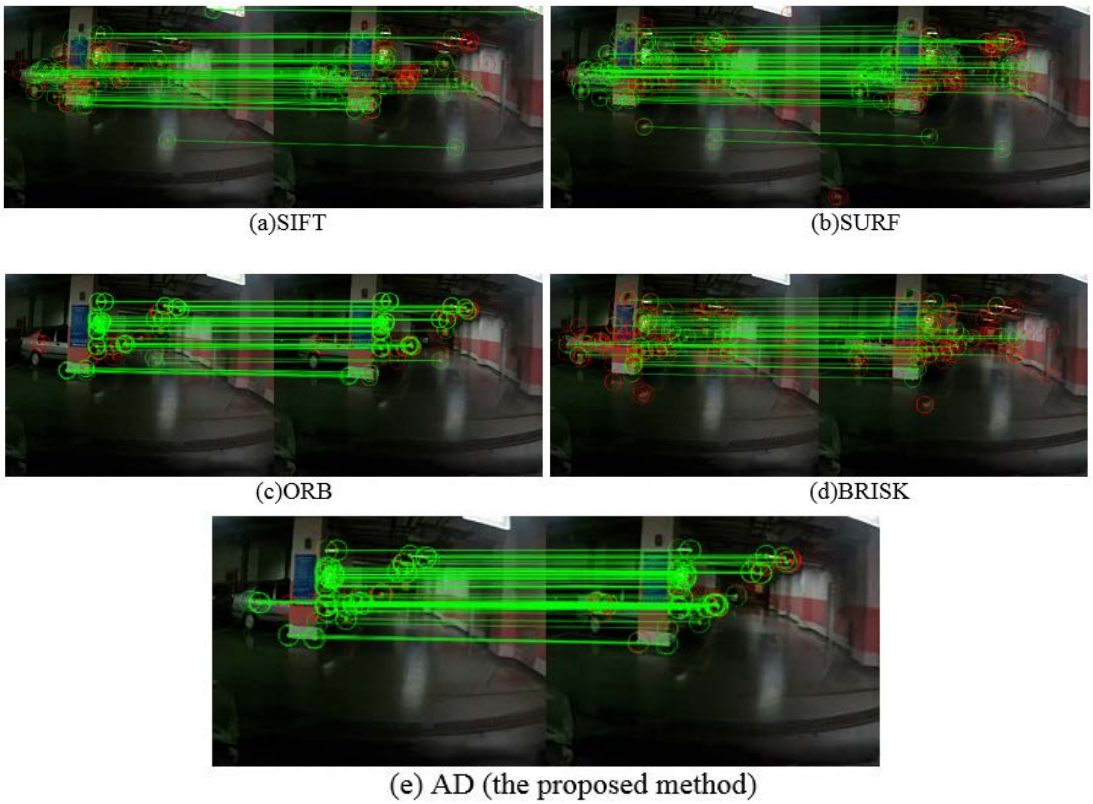


Fig. 11. The results of feature detection and matching for fuzzy images in Video 4. The red circles represent single points or mismatch correspondences, while green circles and lines are correct matching points. The number of keypoints in this group tests is around 140.

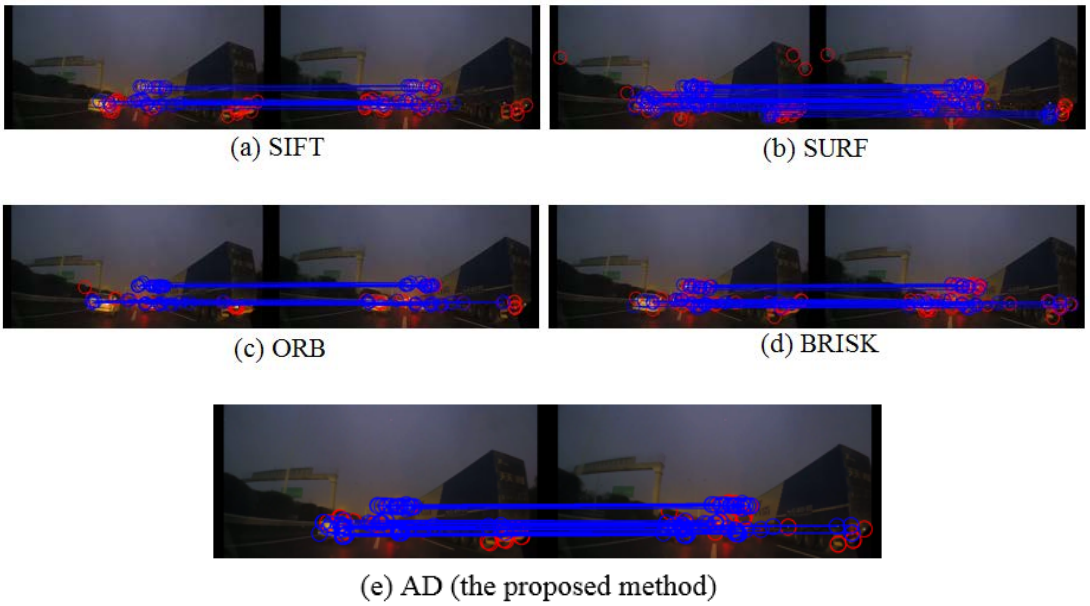


Fig. 12. The results of feature detection and matching for fuzzy images in Video 5. The red circles represent single points or mismatch correspondences, while blue circles and lines are correct matching points. The number of keypoints in this group tests is around 94.

3.1.2 Repeatability of fuzzy images

To make a quantitative evaluation of the detection quality, the repeatability comparison measure proposed in [14] was adopted:

$$repeatability = \frac{num\ of\ \ corresponding\ keypoint\ pairs}{\min\left(N_i, N_j\right)} \tag{9}$$

where N_i and N_j are the number of visible keypoints measured in images i and j, respectively. A correspondence is considered if the overlap region of an identical keypoint between two frames is larger than 50%. The higher the repeatability value is, the stronger ability of anti-blur an algorithm has.

By altering the threshold of the detector, we can control the number of extracted features from fuzzy image, and then it is capable to achieve the return repeatability value at this threshold value. A large number of experiments have been made by adjusting the threshold in a wide range. It turns out one or more local peaks of repeatability obtained via changing the threshold value. At some peaks, the number of feature points extracted by different algorithms are close, but sometimes there are some deviations. In the comparison experiments, suitable thresholds of each algorithm are set to guarantee approximately the same number of features and high repeatability, and the filters in all algorithms are removed for fairness.

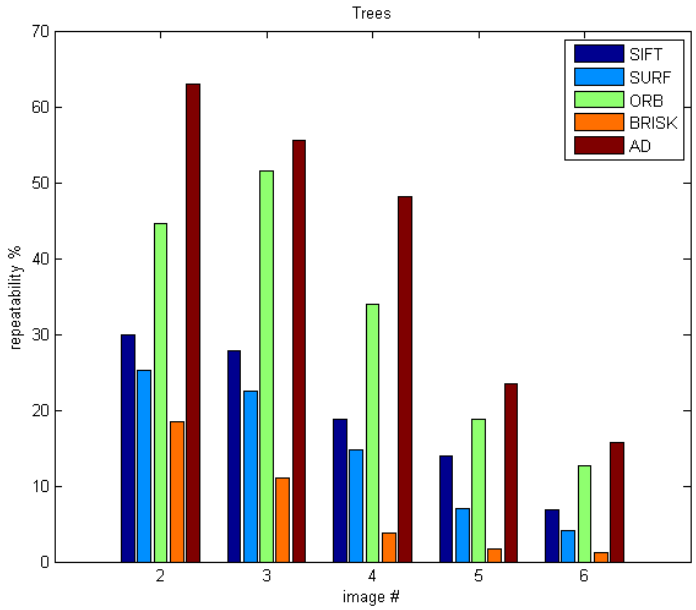


Fig. 13. Repeatability over 50% of the overlap region of the SIFT, SURF, ORB, BRISK, and AD in the standard blurring evaluation set.

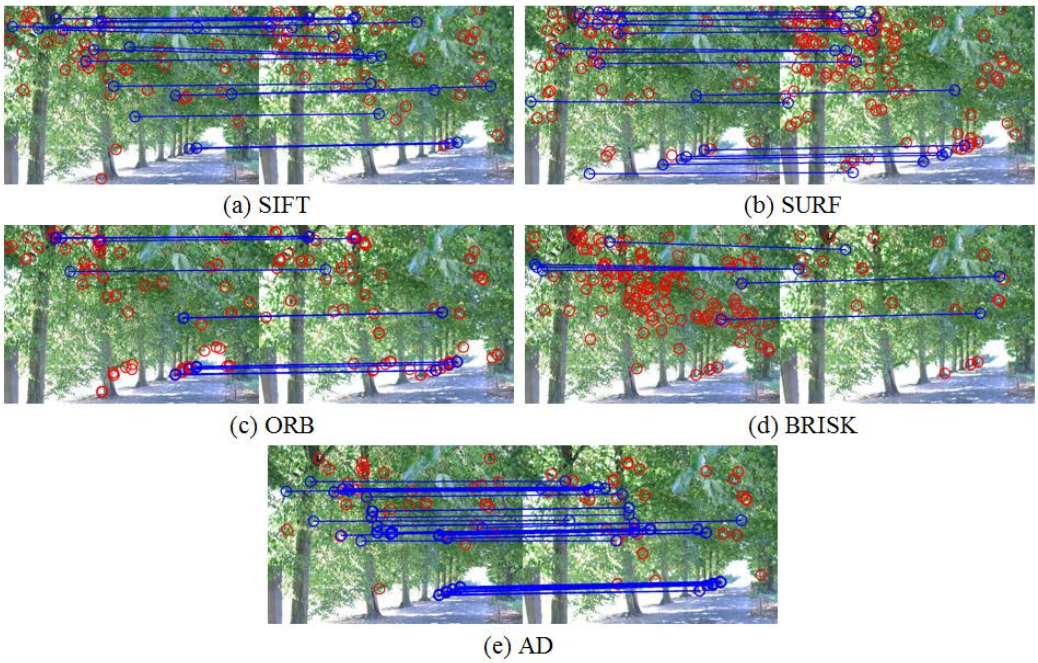


Fig. 14. The results of fourth-grade fuzzy comparison experiments. The red circles represent single points or mismatch correspondences, while blue circles and lines are correspondences. The number of keypoints in this group tests is around 100.

Fig. 13 shows that the values of the repeatability plot exhibit a downward trend accompanied with gradually increasing blurring in the tree dataset. The proposed method displays obvious advantage compared with the other methods, and it maintains a high degree of repeatability even in the case of fourth-grade fuzzy, as shown in **Fig. 14**.

3.1.3 Repeatability interval of videos

A new evaluation criterion is proposed for videos by utilizing a percentage interval of repeatability. The value of repeatability between two adjacent frames is the ratio of the correspondences and participant keypoints. Accordingly, one percentage interval of repeatability is the result of a recording. The method adopted to monitor the percentage interval of repeatability for a video sequence is only dependent on digital image processing. The experiments in the first step are related to binary descriptor matching distance. It is supposed a correct match when the matching distance of a keypoint pairs is satisfied:

$$dist \le min\_dist + 0.3 \times (max\_dist - min\_dist) \tag{10}$$

where *max_dist* represents the farthest matching distance of all matchers, and *min_dist* is the nearest. RANSAC method is adopt as the second step of the image processing method.

By means of a great number of tests, the number of corresponding keypoint pairs after binary descriptor distance test and RANSAC filters is presumed to be the real number and the results of repeatability by this method have a margin of error of ±3 percent. In **Fig. 15**, their average value represents the slightly fluctuate values of repeatability in one phase as a data note in **Fig. 16**. When a video content is notable changed, the value of repeatability results in a step change in the phase. A narrower interval and a higher occupied percentage position imply more effectiveness and adaptability.

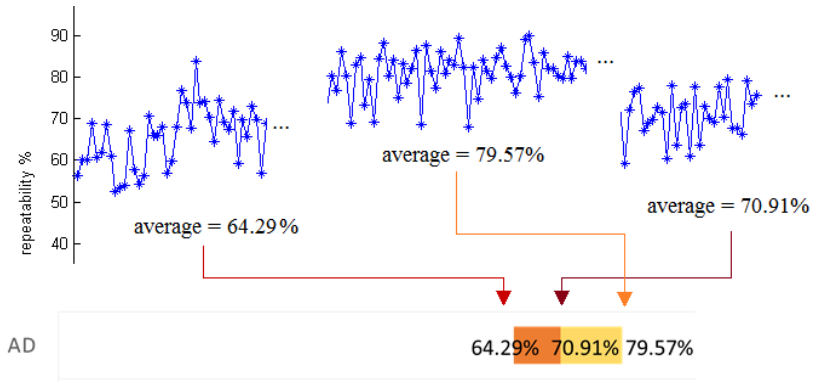


Fig. 15. Illustration tests of repeatability intervals by AD.

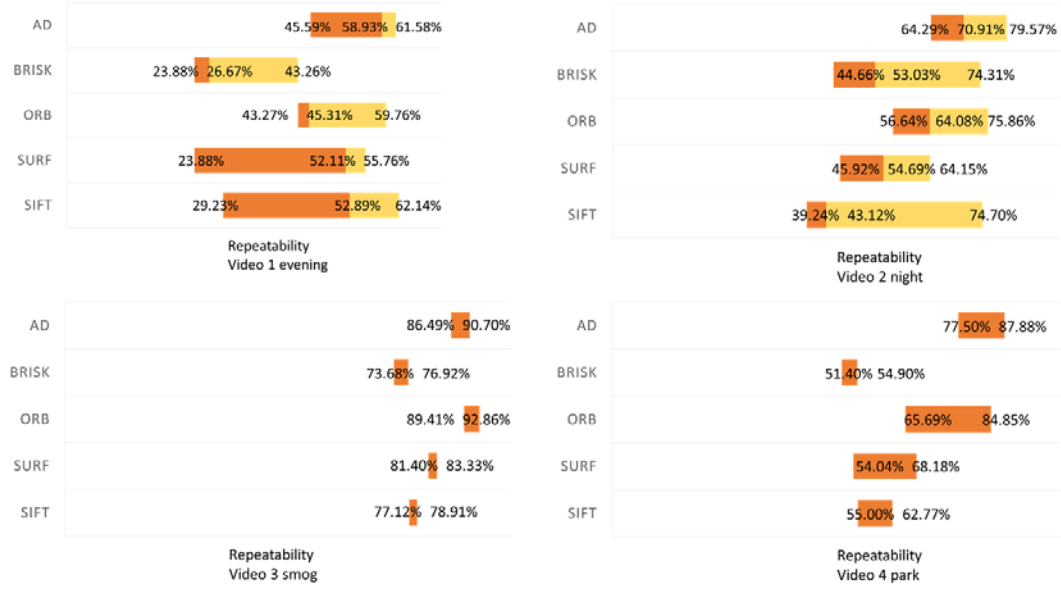


Fig. 16. Repeatability intervals of the SIFT, SURF, ORB, BRISK, and AD in four fuzzy videos.

Lack of illumination occurs both at Video 1 evening and Video 2 night, whereas the large difference in content between frames in Video 1 evening leads to decreasing repeatability and increasing interval width of all algorithms. The value of repeatability in Video 2 night is mainly dependent on the position of the light source. In addition, the dust and water on the lens creates an undesirable effect on processing after exposure. Although the smog in Video 3 causes blurring of distant contents, it has little effect on the closer contents and consequently seldom influences the repeatability. Meanwhile, it reduces the accuracy of estimation of the global motion vector.

The contents of Video 1 evening and Video 2 night vary with the movement of the carriers. Both have two significant stages of change, which provide three values of repeatability. On the other hand, Video 3 smog and Video 4 parking merely have one segment and two repeatability values. The maximum repeatability determines the number of keypoints. Moreover, an approximately the same quantity of features is present for each algorithm to ensure objectiveness in the entire period.

As expected, in **Fig. 16** the proposed approach has the shortest interval and the highest percentages. Although the performance of SURF and SIFT are similar, SURF has an edge over SIFT in terms of repeatability. BRISK entirely depends on the intensity information, which results in a poor expression of repeatability.

3.2. Frame Rate

The two important factors that determine the average frames per second (fps) of the algorithms are the video contents and pixel dimensions. Classically, they are considered to have met the real-time requirement when the processing time reaches up to 30 fps. In the experiments, two conditions need to be satisfied for the motion estimation: the repeatability of the algorithms maintained at a high level and the quantity of correspondence for affine transformation. Hence, the number of detection feature points is selected as 140.

Table 1. Frame rate of the stabilization methods (fps)

Device	Content	Resolution	Average fps				
			SIFT	SURF	ORB	BRISK	AD
Gosafe520	(a)	896 × 504	1.37	5.23	18.20	40.02	25.25
vichengA3	(b)	896 × 504	1.36	4.79	23.13	36.66	30.43
PAPAGO!	(d)	640 × 480	2.30	5.75	31.17	41.77	37.97
Goluk T1	(e)	806 × 452	1.47	6.05	20.37	37.46	33.48

The results listed in **Table 1** show that the binary matching algorithms are several times faster than the SIFT-like algorithms and can satisfy the request of real time depends on the pixel dimensions and video content. In addition, the video shooting at the underground parking garage suffers from a lower degree of ambiguity, which makes detection of high-quality feature points easier. Consequently, the discrepancy in the processing speeds between BRISK and the new method AD is significantly reduced.

3.3. ITF

In this series of experiments, ITF is adopted to test and verify the practicability of this algorithm for video source. A higher ITF imply more practicability.

$$ITF = \frac{1}{N_{frame-1}} \sum_{b=1}^{b=N_{frame-1}} PSNR(I_a, I_b) \tag{11}$$

where N_{frame} is the total number of video sources, I_a is the reference frame, and I_b is the current frame after stabilization. $PSNR(I_a, I_b)$ is defined as the peak signal-to-noise ratio [23]:

$$PSNR(I_a, I_b) = 10\log_{10}\frac{255^2}{MSE(I_a, I_b)} \tag{12}$$

where $MSE(I_a, I_b)$ represents the mean square error of the comparison.

Table 2. ITF of the stabilization methods

Video Seq.	ITF					
	Original	SIFT	SURF	ORB	BRISK	AD
(a)	11.51	15.91	14.82	17.11	14.46	17.16
(b)	22.45	24.49	24.25	24.68	24.39	25.08
(d)	16.28	20.97	18.96	23.21	20.37	25.61
(e)	28.49	34.16	33.65	32.92	30.84	36.88

Table 2 lists the comparative results of ITF. A higher ITF score indicates a more practical measure. The total number of feature points in a frame is 140 and all results are tested for more than 20 times. From **Table 2**, the proposed approach shows better results than the others. The ORB algorithm is also impressive, although block matching makes it sensitive to large content change as well as rotation. Obviously, SURF makes more deviation between two frames than SIFT. The key reason for the worse performance of BRISK is that it focuses on higher speed.

Besides, it is clear that AD is superior in ITF but inferior in frame rate to BRISK. However, the data from Table 1 indicate AD is capable to achieve the real-time requirement for low computation power of mobile processor on vehicles. AD with more effective performance is consequently more suitable than BRISK for driving records.

3.4. Unintentional Motion Estimation

In this stage, two frames captured from Video 1 are employed as an example to illustrate the proposed feature selection strategy for unintentional motion estimation. As mentioned above, the background matching point pairs should be obtained in order to compensate the dithering caused by unintentional camera movements. **Fig. 17** shows the matching results of correspondences before and after using feature points selection strategy with AD algorithm. First, the mismatching correspondences are filtered out. The feature points located at the moving target are usually low quality ones, and the experimental results show that the proposed method can effectively avoid the detection of feature points on moving targets. If there are feature points on the moving target, they will be removed by feature points selection strategy, and only the background feature points survive.

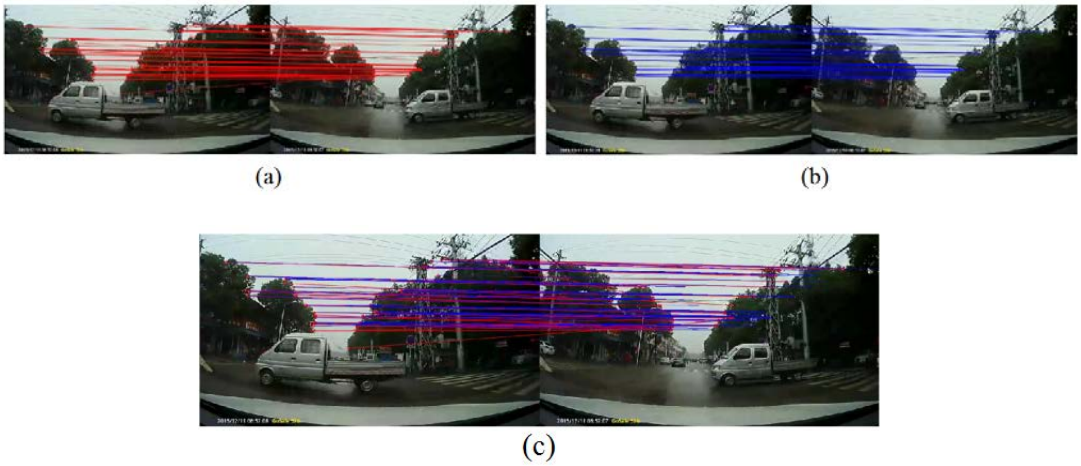


Fig. 17. The results of feature selection for fuzzy images in Video 1. (a) and (b) shows the matching result before and after using feature points selection strategy, respectively. (c) performs the comparison of (a) and (b).

4. Conclusion

In this paper, an antiblurry dejitter video stabilization algorithm is proposed based on quality feature point detection and accurate description to constrain the dithering influence. Further, the accuracy of unintentional motion estimation benefits from the feature points selection strategy. Moreover, repeatability interval is proposed to delineate the different sections of videos. The experiments have confirmed the feasibility and practicability of the proposed antiblurry dejitter stabilization algorithm.

References

[1] K. Uomon, A. Morimura, H. lshii, T. Sakaguchi, and Y. Kitamura, "Automatic image stabilizing system by fully-digital signal processing," *IEEE Trans. on Consumer Electronics*, vol. 36, no. 3, Aug, 1990. Article (CrossRef Link)

[2] T. Kinugasa, N. Yamamoto, and l. Komatsu, "Electronic image stabilizer for video camera use," *IEEE Trans. on Consumer Electronics*, vol. 36, no. 3, Aug, 1990. Article (CrossRef Link)

[3] J.-J. Zhu and B-l. Guo, "Electronic image stabilization algorithm based on adaptive motion filter," *Journal of Theoretical and Applied Information Technology*, vol. 50, no. 1, Apr. 2013. Article (CrossRef Link)

[4] O. Urhan and S. Erturk, " Single sub-image matching based low complexity motion estimation for digital image stabilization using constrained one-bit transform," *IEEE Trans. on Consumer Electronics*, vol. 52, no. 4, pp. 1275–1279, Nov, 2006. Article (CrossRef Link)

[5] H. Okuda, M. Hashimoto, K. Sumi, and S. Kaneko, "Optimum motion estimation algorithms for fast and robust digital image stabilization," *IEEE Trans. on Consumer Electronics*, vol. 51, no. 1, pp. 276–280, Feb, 2006. Article (CrossRef Link)

[6] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. of the seventh IEEE International Conference on Computer Vision*, 1150–1157, 1999. Article (CrossRef Link)

[7] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. of International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 506–513, Jun, 2004. Article (CrossRef Link)

[8] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 10, 27, pp 1615–1630, 2005. Article (CrossRef Link)

[9] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Computer Vision–ECCV*, 404–417, 2006. Article (CrossRef Link)

[10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," *Computer Vision–ECCV 2010*, pp. 778–792, 2010. Article (CrossRef Link)

[11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. of 2011 International Conference on Computer Vision*, 2011. Article (CrossRef Link)

[12] S. Leutenegger, M. Chli, R. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. of 2011 International Conference on Computer Vision*, 2011. Article (CrossRef Link)

[13] P. F. Alcantarilla, A. Bartoli, A. J. Davison, "KAZE Features," in *Proc. of the European Conference on Computer Vision*, 2012. Article (CrossRef Link)

[14] K. Mikolajczyk, T Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," in *Proc. of the International Journal of Computer Vision*, 2005. Article (CrossRef Link)

[15] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. of European Conference on Computer Vision*, 2006. Article (CrossRef Link)

[16] J. Shi and C. Tomasi, "Good features to track," in *Proc. of 9th IEEE Conference on Computer Vision and Pattern Recognition*, *Springer*, June, 1994. Article (CrossRef Link)

[17] H. Moravec, "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," *Tech Report CMU-RI-TR-3 Carnegie-Mellon University, Robotics Institute*, 1980. Article (CrossRef Link)

[18] C. Harris and M. Stephens, "A combined comer and edge detector," in *Proc. of the 4th Alvey Vision Conference*, 1988. Article (CrossRef Link)

[19] T. Lindeberg, "Image matching using generalized scale-space interest points," *Journal of Mathematical Imaging and Vision*, 2015. Article (CrossRef Link)

[20] A. Neubeck, L. Van Gool, "Efficient non-maximum suppression," in *Proc. of International Conference on Pattern Recognition (ICPR)*, 2006. Article (CrossRef Link)

[21] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communication ACM*, 24(6): 381-395, 1981. Article (CrossRef Link)
[22] S. Yao, G. Parthasarathy, and D. Thyagaraju, "Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework," *IEEE Trans. Consumer Electronics*, vol. 55, no. 3, pp. 1714-1721, Aug, 2009. Article (CrossRef Link)
[23] S. B. Balakirsky, R. Chellappa. "Performance characterization of image stabilization algorithms," in *Proc. of 3rd IEEE International Conference on Image Processing*, 1996. Article (CrossRef Link)

Jing-Ying Xiong graduated from the Nanjing University of Aeronautics and Astronautics, China, in 2012. She is currently a Ph.D candidate in the University of Chinese Academy of Sciences, China. Her main research interests are digital image processing and data communication.

Dai Ming received the B.S. degree from Changchun University of Science and Technology, China, in 1990 and the M.S. degree from the Chinese Academy of Sciences, China, in 1993. He is a Professor with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. He is mainly devoted to airborne image processing and stabilizing foundation bed.

Chun-Lei Zhao received the B.S. degree from Changchun University of Science and Technology, China, in 2007 and the Ph.D. degree from Chinese Academy of Sciences, China, in 2016. Her main research interests are digital image processing and video compression.

Ruo-Qiu Wang graduated from the Nanjing University of Science and Technology, China, in 2012. She is currently a Ph.D candidate in the University of Chinese Academy of Sciences, China. Her main research interests are diffractive optics.