



Article

High-Precision Satellite Video Stabilization Method Based on ED-RANSAC Operator

Feida Zhang ^{1,2} , Xin Li ³, Taoyang Wang ^{2,*}, Guo Zhang ³ , Jianzhi Hong ², Qian Cheng ² and Tiancheng Dong ³

¹ School of Earth Science and Resources, China University of Geosciences, Beijing 100083, China; zhangfeida@email.cugb.edu.cn

² School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; hongjianzhi@whu.edu.cn (J.H.); cheng_qian@whu.edu.cn (Q.C.)

³ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; lixin_1995@whu.edu.cn (X.L.); guozhang@whu.edu.cn (G.Z.); 2021106190028@whu.edu.cn (T.D.)

* Correspondence: wangtaoyang@whu.edu.cn

Abstract: Video image stabilization technology is a crucial foundation for applications such as video image target identification, monitoring, and tracking. Satellite video covers a wide range of areas with complex and similar types of objects on the ground and diverse video types. However, currently, there is a lack of a general high-precision satellite video stabilization method (VSM) that can be applied to different land cover types and imaging modes. This paper proposes a high-precision VSM based on the ED-RANSAC, an error elimination operator constrained by Euclidean distance. Furthermore, a set of accuracy evaluation methods to ensure the reliability of video stabilization are sorted out. This paper conducted video stabilization experiments using optical video data from the Jilin-01 satellite and airborne SAR video data. Under the precision evaluation criteria proposed in this paper, the optical satellite video achieved inter-frame stabilization accuracy of better than 0.15 pixels in different test areas. The overall stabilization accuracy was better than 0.15 pixels. Similarly, the SAR video achieved inter-frame stabilization accuracy better than 0.3 pixels, and the overall stabilization accuracy was better than 0.3 pixels. These experimental results demonstrate the reliability and effectiveness of the proposed method for multi-modal satellite video stabilization.

Keywords: satellite video; video stabilization; movement estimation; ED-RANSAC



Citation: Zhang, F.; Li, X.; Wang, T.; Zhang, G.; Hong, J.; Cheng, Q.; Dong, T. High-Precision Satellite Video Stabilization Method Based on ED-RANSAC Operator. *Remote Sens.* **2023**, *15*, 3036. <https://doi.org/10.3390/rs15123036>

Academic Editor: Jon Atli Benediktsson

Received: 26 April 2023

Revised: 29 May 2023

Accepted: 8 June 2023

Published: 10 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, optical video satellites have thrived, since the first Skysat series of small video satellites launched by Skybox Imaging in the United States in 2013. The video camera carried on the satellite can continuously observe the moving process in the form of video recording [1], with observation times of up to 120 s. Compared with traditional satellite images, video images can capture changes in a specific area in a short amount of time and can be effectively applied to real-time monitoring of natural disasters such as volcanic eruptions, earthquakes, floods, and fires. However, the application of satellite video is severely constrained by factors such as satellite attitude control errors, satellite platform jitter, and differences in imaging viewpoints between adjacent frames. As a result, correct mapping relationships between pixels in inter-frames cannot be established [2]. Video stabilization aims to eliminate or reduce the relative deformation between adjacent frames, establish correct mapping relationships between homologous image elements, and generate stable and smooth videos.

There are relatively few studies on satellite video stabilization. The existing VSMs mainly include the technique based on the classical motion model and the rational polynomial model (RPC) process. Both methods first need to obtain the video inter-frame motion vector. The feature-based method [3,4] is mainly used. For example, this is completed using

the SIFT [5–7] or SAR-SIFT [8] algorithms or the deep learning method [9–11] to extract the homologous points between video frames. Secondly, various forms differ in the different transformation models adopted. Among the RPC-based methods, Zhou Nan [12] and others proposed a Digital Elevation Model (DEM)-assisted VSM for optical video satellites. At the same time, each frame of the video was given geocoding [13]. Zhang et al. [14] studied the image stabilization of satellite video with geometric model constraints; Wang Xia et al. [15] proposed a VSM considering image plane distortion. The high-precision VSM based on RPC and DEM is difficult to widely use because of the inaccuracy and easy loss of RPC information and the plane projection error caused by DEM. Therefore, a high-precision VSM based on the classical motion model is designed to expand the application range of satellite video formats in satellite video stabilization.

Among the methods based on the classical motion model, Feng Li [16] used the rigid transform model as the inter-frame motion model to perform video stabilization on the infrared video satellite, but this method can only be applied to video images with rotation and translation; Kumar et al. [17] and Maolei Zhang et al. [18] used the affine transform model for video stabilization; Murthy et al. [19] used the perspective transform model as the inter-frame motion model to perform video stabilization on the SkySat-1 satellite, but the accuracy obtained was low; Hui Xing et al. [20] and Walha et al. [21] used the similar transformation model for video stabilization. All of the above methods have advantages only for a particular data condition, cannot be widely used for many types of satellite video data, and have other limitations.

The stabilized image of Synthetic Aperture Radar (SAR) video has been less studied, and it is only stabilized to some extent in the generation stage of SAR video. For example, Yan et al. [22] obtained the stabilized SAR video by stabilizing the rotation and trajectory of the platform in real time; Robert Linnehan et al. [23] generated the stabilized SAR video by introducing the concept of map drift to compensate for the platform motion.

This paper presents a general satellite VSM based on the traditional transformation model. It addresses the limitations that the existing satellite VSMs are only applicable to specific data and improves the error elimination process. To enhance the stabilization accuracy of the satellite VSM based on the traditional motion model, an improved error elimination algorithm based on RANSAC using the Euclidean distance constraint, ED-RANSAC, is proposed. Furthermore, we propose evaluation indexes for assessing the stabilization accuracy of satellite video since the current satellite VSMs lack a systematic approach for evaluation.

The innovations of this paper include the following:

- (1) An improved error rejection algorithm: the Euclidean distance-constrained RANSAC algorithm (ED-RANSAC) is proposed to achieve high-precision homologous feature extraction. Additionally, the limitation that existing satellite VSMs are only applicable to specific data is solved.
- (2) The stabilization accuracy of optical video is improved to better than 0.15 pixels to achieve stable and smooth video, providing a reliable database for subsequent applications such as target detection based on video data. Additionally, SAR video stabilization accuracy of better than 0.3 pixels can also be achieved.
- (3) The paper proposes evaluation indexes for assessing the stabilized image accuracy of satellite video.

2. Methods

The main methods currently applied for satellite video stabilization are the adjacent frame method and the fixed frame method. The fixed frame method, also called the master frame method, is a method that aligns the auxiliary frames to the master frame by using the first or middle frame of the video as the master frame, and the other frames as the auxiliary frames. This method suits satellite video with comprehensive coverage and minor changes between video frames. Still, it is easy to produce the phenomenon that the error gradually worsens as the number of frames increases.

In this paper, the adjacent frame method is used as the video stabilization method process. The adjacent frame method is to take the former frame of the video as the main frame and the last frame as the auxiliary frame. The SIFT algorithm [5] is used to detect homologous points for two frames. The proposed ED-RANSAC algorithm is used to eliminate the mismatched data to improve the homonymous point alignment accuracy. The homologous points are used to calculate the model transformation parameters between the two frames. Finally, the auxiliary frame is corrected to the correct position of the main frame by the calculated model transformation parameters, and the video frame image sequence is obtained. Figure 1 shows the experimental flow.

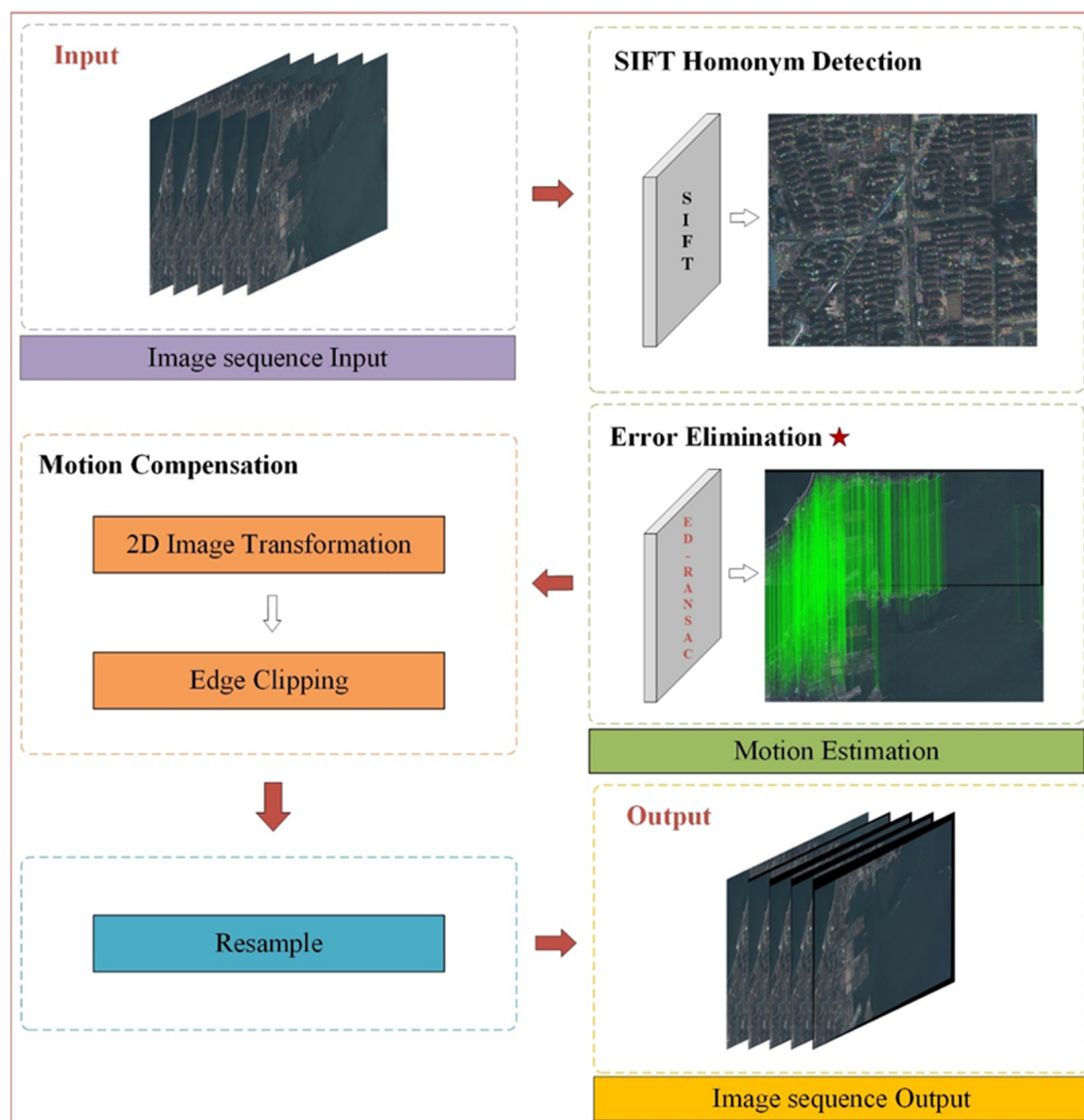


Figure 1. Flow chart of experimental method. The red star part is the main part of this article.

2.1. Homologous Feature Detection Algorithm

The SIFT algorithm was adopted to achieve homologous feature extraction in this paper. This algorithm can be roughly divided into four steps: creating a scale space, feature point localization, key point direction distribution and generate descriptors.

Creating a scale space: The SIFT algorithm uses Gaussian kernel functions $G(x, y, \sigma)$ of different scales convolved with a two-dimensional image $I(x, y, \sigma)$ to create the scale space $L(x, y, \sigma)$. The convolution operation is represented as follows:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y, \sigma) \quad (1)$$

where σ is the scale factor, which indicates the blurring degree of the image.

Differentiate two adjacent Gaussian images to obtain the difference of Gaussian pyramid (DOG), expressed as:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - D(x, y, \sigma)] * I(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (2)$$

Feature point localization: Each pixel in the DOG is compared with its 26 adjacent points to determine whether it is an extreme point; then, the detected extreme points are fitted using Taylor expansion to find the correct position of the feature point on the image.

Key point direction distribution: To achieve rotation invariance and reduce the impact caused by image rotation on descriptors, Formulas (3) and (4) are used to calculate the gradient values and directional parameters of different feature points (x, y) at their respective scales. Within the neighborhood range of each feature point, every 10 degrees is a direction, and the gradient histograms of 36 directions between 0 degrees and 360 degrees are counted. The direction with the peak value in the histogram is determined as the main direction of the feature point.

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \quad (3)$$

$$\theta(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (4)$$

Among them, $m(x, y)$ and $\theta(x, y)$ denote the gradient values and direction parameters of the feature points (x, y) on their respective scales, respectively.

Generate descriptors: The coordinate axis is rotated to the main direction of the feature point, using the feature point as the center. A 4×4 window is then set, with every 45 degrees representing an interval. The 8 directional intervals are evenly divided between 0 and 360 degrees. For each unit within the window, the gradient histograms of the 8 directions are calculated. These histograms are then Gaussian-weighted to cover the entire window and normalization is performed to generate a 128-dimensional descriptor vector.

2.2. ED-RANSAC Algorithm

The Randomized Sampling Consistency (RANSAC) algorithm, proposed by Fischler and Bolles in 1981 [24], is a stochastic parameter estimation algorithm that iteratively fits the mathematical model parameters from a set of sample points.

An error threshold needs to be set as an upper limit of iterations for traditional RANSAC algorithm. The algorithm can fail to converge if the maximum number of iterations is not set in advance. The upper limit on the number of iterations is closely related to the probability of obtaining the best model. As the upper limit on the number of iterations increases, the probability of obtaining the best model increases. However, a larger upper limit setting leads to an increase in computational cost, which reduces the speed of execution of the algorithm. It has been argued that the RANSAC process was too time-consuming since an assumption of RANSAC rarely holds in practice: namely, the assumption that the model parameters are calculated from an uncontaminated sample. An improved method called LO-RANSAC [25] is therefore proposed, which takes advantage of the fact that the model assumptions from the smallest uncontaminated sample are almost always sufficiently close to the optimal solution. An algorithm that is almost identical to the theoretical performance is produced when applied to the local optimization step of the chosen model. Lo-RANSAC increases the number of outlier detections and thus speeds up the overall solving process by allowing for the early termination of the RANSAC iterative process, ultimately achieving the aim of obtaining a higher quality model. The big trouble, however, is that this method requires the identification of a pure sample as a basis, and finding a pure sample is usually uncertain. In addition to RANSAC and its related variant methods, the use of the Pauta criterion (3sigma) to reject outliers is a valid method. It assumes that the sample obeys a normal distribution and 99.7% of the correct values are

within three standard deviations. This method is suitable for data with a large number of samples, making this method suitable for data with a large sample size.

To improve the overall stabilization accuracy while considering the time cost, this paper proposes an improved RANSAC algorithm. The algorithm selects matching pairs randomly as samples to calculate the transformation matrix. The algorithm calculates the consistent set that satisfies the current transformation matrix based on the transformation matrix, the sample set, and the error metric function. Then, it iteratively updates the optimal consistent set. The spatial distance between two points is calculated, and the Euclidean distance (ED) was introduced as a threshold to filter the optimal consistent set twice. The matching pairs that satisfy the threshold condition in the optimal set are retained as the final set of homologous points to calculate the transformation matrix.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \rightarrow \begin{cases} x_s = \frac{X}{Z} = \frac{a_{11}x+a_{12}y+a_{13}}{a_{31}x+a_{32}y+a_{33}} \\ y_s = \frac{Y}{Z} = \frac{a_{21}x+a_{22}y+a_{23}}{a_{31}x+a_{32}y+a_{33}} \\ Z' = \frac{Z}{Z} = 1 \end{cases} \quad (5)$$

Equation (5) represents the two-dimensional plane coordinates (x_s, y_s) obtained after the transformation from the original coordinates consisting of the homogeneous coordinate $(x, y, 1)$. Among them, a_{ij} is the transformation parameter obtained by least-squares decomposition of four randomly selected points, $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ represents the linear image transformation, $\begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$ is the translation on x, y , respectively, and $\begin{bmatrix} a_{31} & a_{32} \end{bmatrix}$ is used to generate the image perspective transformation. a_{33} is usually set to 1.

$$ED = \sqrt{(x_r - x_s)^2 + (y_r - y_s)^2} \quad (6)$$

In Equation (6), (x_r, y_r) are the coordinates of the homonymous points on the reference image, and (x_s, y_s) are the coordinates of the homonymous points on the image to be aligned after the transformation of Equation (5). Figure 2 shows the algorithm flow.

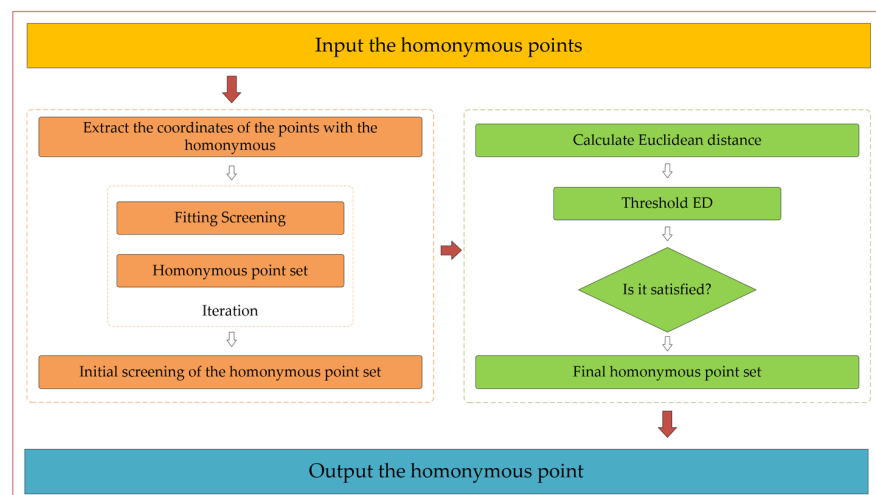


Figure 2. ED-RANSAC algorithm flow chart.

2.3. Evaluation Indicators

In this paper, the Root Means Square Error (RMSE) is used as the evaluation index of steady image accuracy, and its formula is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_{ri} - x_{si})^2 + (y_{ri} - y_{si})^2}{N}} \quad (7)$$

In Equation (7), $(x_{ri} - y_{ri})$ are the coordinates of the homonymous points detected on the primary image, $(x_{si} - y_{si})$ are the coordinates of the homonymous points detected on the auxiliary image transformed by Equation (5), and N is the number of homonymous points.

3. Experiment and Analysis

In this section, we conduct video stabilization experiments and evaluate the accuracy using different regions of optical satellite video and SAR video to verify the effectiveness and generalizability of the proposed method.

3.1. Experiment Data

This experiment was conducted using Jilin-01 optical video satellite data for verification. The Jilin-01 video satellite orbits at an altitude of 656 km, the ground resolution of 1.13 m. The single shot video can last up to 120 s, and the frame rate of the video is 25 frames per second [26–28]. To demonstrate the wide applicability of the method proposed in this paper, satellite video data from three different land cover types were used for the experiments, namely, sea area (Zhifu Bay in Yantai), desert (Jiayuguan in Gansu), and mountainous area (Leibo County in Sichuan). The details of the satellite video data used in this experiment are shown in Table 1. Figure 3 shows the satellite video images for the three different land cover types used in this experiment. It can be observed from the figure that the three feature types varied greatly. For instance, the data from Yantai Zhifu Bay and Gansu Jiayuguan have a majority of areas with an inconspicuous texture inside, which could have an impact on the detection of key points.

Table 1. Details of the experimental data.

Data Source	Area	Size (Pixels)	Frames
Jilin-01 video 03 star	Zhifu Bay in Yantai	12,000 × 5000	94
	Jiayuguan in Gansu		100
	Leibo County in Sichuan		33

3.2. Threshold ED Determination

The error rejection process is iterative, and the obtained homologous points are utilized to compute the transformation parameters for correcting the frame images. The smaller the Euclidean distance between the coordinates of the homologous points on the main frame and the coordinates of the homologous points on the auxiliary frame after transformation, the higher the correction accuracy between the main frame and the auxiliary frame and the higher the accuracy of video stabilization when pushed to the entire video frame sequence. Figure 4 illustrates the application of this method using data from Leibo County in Sichuan Province. In this example, the threshold ED value is set to 0.2, which maximizes the elimination of false match pairs while preserving sufficient correct match pairs to calculate the model transformation parameters. The relationship between the ED value and RMSE is shown on the left side of the image axis, while the relationship between the ED value and Correct Matching Number (CMN) is depicted on the right side. To demonstrate the balance between steady-state accuracy and correct matching points, and to emphasize the rationality of the chosen threshold, the RMSE is inverted in the figure.

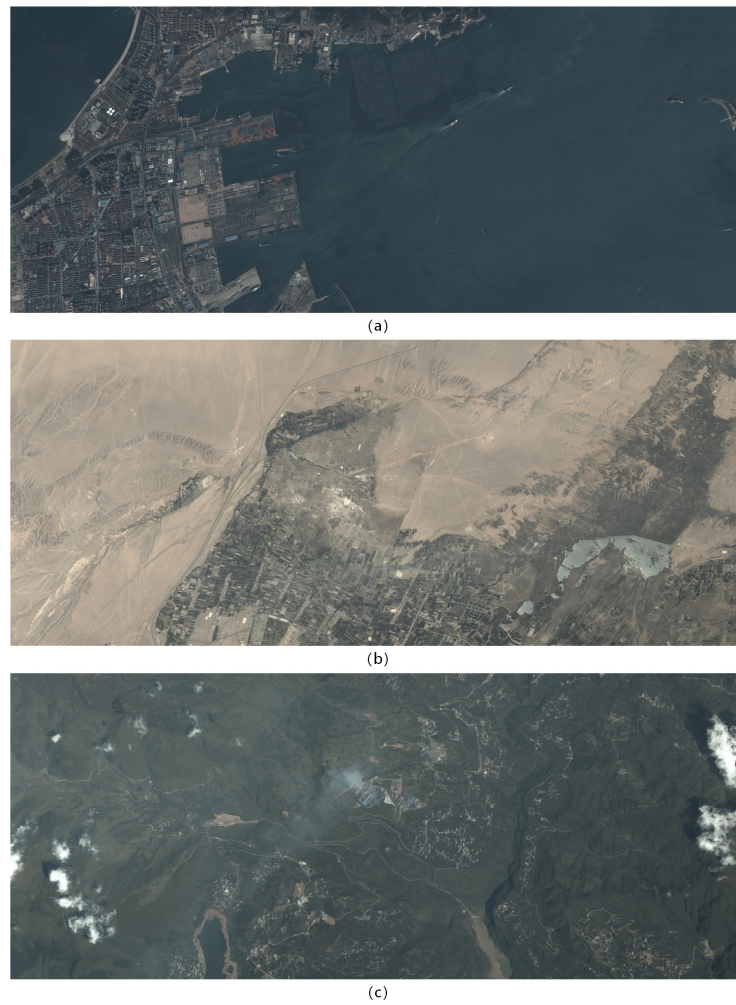


Figure 3. Diagram of video images of three feature types from Jilin-01. (a) Zhifu Bay in Yantai. (b) Jiayuguan in Gansu. (c) Leibo County in Sichuan.

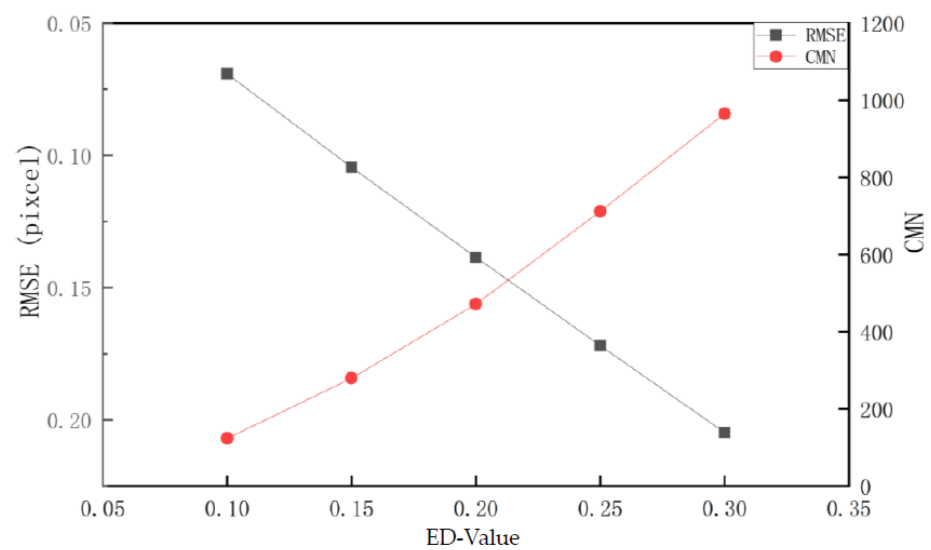


Figure 4. ED value verification chart. The relationship between the ED value and RMSE is shown on the left side of the image axis, and the relationship between the ED value and the Correct Matching Number (CMN) is established on the right side. ED-Value is the threshold value used for secondary screening of the homologous points.

3.3. Inter-Frame Motion Model

Scholars have used many transformation models, such as rigid, similar, affine, and perspective transformations, as inter-frame motion models for satellite video stabilization. The rigid transform only translates and rotates the image without changing the shape of the graph, so the rigid transform is unsuitable for satellite video stabilization because of the deformation between satellite video frames. The similarity transform is an extension of the rigid transform, and the similarity transform is a rigid transform when the scaling factor is 1. The affine transformation is a linear transformation from two-dimensional coordinates to two-dimensional coordinates, which responds to the mapping relationship between image coordinates before and after the shift [29] and is widely applied to image transformation. The perspective transformation is a transformation that takes advantage of the condition that the perspective center, image point, and target point are co-linear and rotates the shadow-bearing surface (perspective surface) around the trace (perspective axis) by a certain angle according to the law of perspective rotation, which destroys the original projection ray beam and still keeps the projection geometry on the shadow-bearing surface unchanged [30]. It is more widely applicable than the affine transformation.

In order to find a suitable transformation model, the following experiments are designed in this paper.

- (1) Two adjacent frames of the condition data of the three land cover types listed in Section 3, sea, desert, and mountain, are selected for homologous point detection to obtain 19,599, 18,197, and 12,169 homologous point pairs each, respectively.
- (2) The homologous point pairs were input into the ED-RANSAC operator combined with three transformation models (affine, perspective, similarity) for screening.
- (3) The First Select (FS), Final Point (FP), Correct Matching Ratio (CMR), and RMSE are plotted as discriminators. This is used to discriminate which model is more suitable for satellite video stabilization.

Figure 5 and Table 2 show the statistical results.

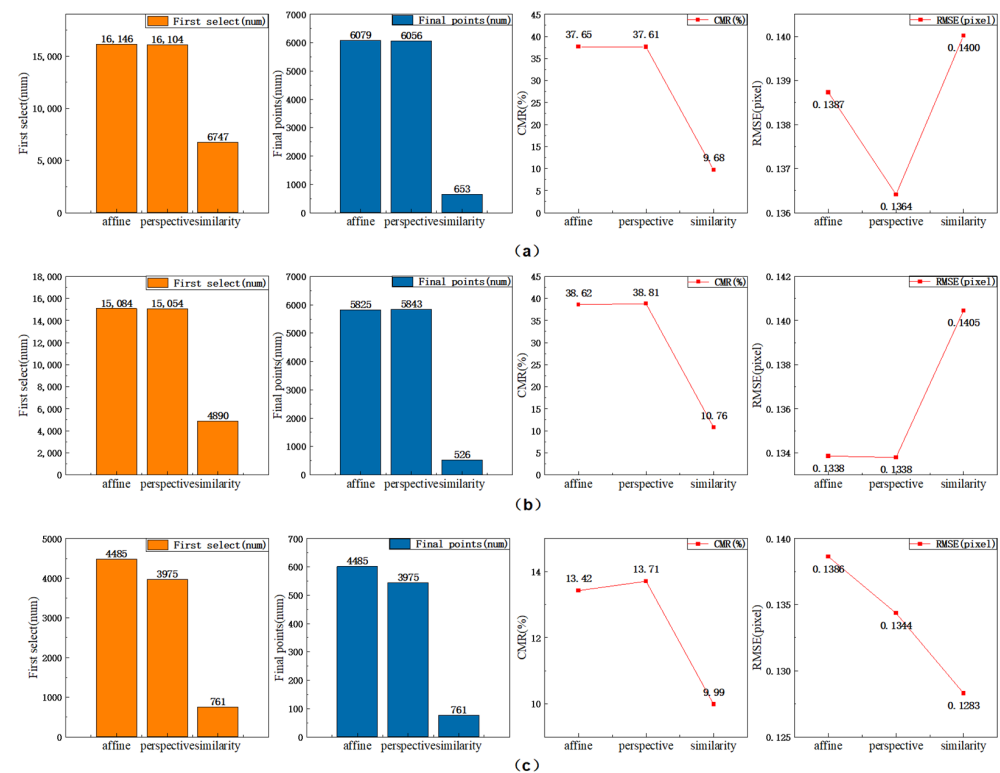


Figure 5. Transformation model discriminant statistics chart. (a) Zhifu Bay in Yantai (b) Jiayuguan in Gansu (c) Leibo County in Sichuan. From left to right are FS, FP, CMR, and RMSE.

Table 2. Statistical details of the discriminatory indicators of the transformation model. (Unit: Pixels).

Area	Model	FS	FP	CMR	RMSE
Zhifu Bay in Yantai	Affine	16,146	6079	37.65019	0.13873
	Perspective	16,104	6056	37.60556	0.13641
	Similarity	6747	653	9.67838	0.14002
Jiayuguan in Gansu	Affine	15,084	5825	38.61708	0.13385
	Perspective	15,054	5843	38.8136	0.13379
	Similarity	4890	526	10.75665	0.14045
Leibo County in Sichuan	Affine	4485	602	13.42252	0.13864
	Perspective	3975	545	13.71069	0.13436
	Similarity	761	76	9.98686	0.12831

From Figure 5 and Table 2, it can be seen that the similarity transformation model performs poorly in all four discrimination indicators for the three land cover types, with fewer matching points, lower screening accuracy, and lower matching accuracy compared to the other two transformation models. The affine transformation and perspective transformation models perform well in all four discrimination indicators, with only small differences in matching points, screening accuracy, and RMSE. However, the perspective transformation model performs more evenly than the affine transformation model under the three land cover types.

The perspective transformation model is more suitable as the transformation model for satellite video stabilization.

3.4. Experimental Precision Evaluation Methods

(1) Inter-frame video stabilization precision evaluation

The satellite video stabilization process, whether the fixed frame, frame-by-frame, or the setting the main frame at intervals method, requires the evaluation of the matching accuracy between the main frame and the auxiliary frame. The accuracy of each frame match is calculated and charted to assess the level of stabilization achieved. Then, the average value of the matching accuracy between all frames is used as a benchmark, and the difference between the average value of the matching accuracy of each frame is calculated to determine the fluctuation of the stabilization accuracy of the image stabilization method.

(2) Overall video stabilization precision evaluation

The inter-frame stabilization accuracy does not represent the real accuracy of the stabilization method, and sometimes the phenomenon of error propagation may occur. Therefore, it is necessary to perform overall accuracy verification of the output image sequence. The first frame in the output image sequence is used as the reference frame, and the image matching method is used to verify the accuracy with the image of each interval of 10 frames to see whether there is error accumulation. The average of the inter-frame validation matching accuracy is used as the true stabilization accuracy of the satellite video stabilization method.

3.5. Experimental Results and Analysis

According to the video stabilization method introduced in Section 2, the stabilization experiments were conducted on the satellite video image data of the three land cover type conditions in Section 3.1 as follows:

- (1) To verify the stability of the proposed method, the average frame-to-frame stabilization precision of each experimental data point was used as the reference. The difference between the stabilization precision of each image frame and its average value was calculated to study the fluctuation of the frame-to-frame stabilization precision.
- (2) The RANSAC, LO-RANSAC, 3sigma, and ED-RANSAC algorithm were used to conduct video stabilization experiments on the three types of land cover data to compare

the improvement of the stabilization precision before and after the improvement of the RANSAC algorithm. In this paper, we first perform a preliminary screening of homonymous points, and the number and content of input homonymous points are the same for the four algorithms.

- (3) The overall stabilization precision of the output image sequence was verified. The first frame of the output image sequence was used as the reference frame, and the image matching method was used to verify the precision of every 10 frames to check for error accumulation. The matching precision of the first and last frames was used to verify the true stabilization precision of the satellite video stabilization method.

3.5.1. Inter-Frame Video Stabilization Precision Evaluation

From Figure 6, the method in this paper has good performance in video steadiness accuracy under various land cover types. Among them, the steady image accuracy of Yantai Zhifu Bay and Sichuan Leibo County fluctuates within ± 0.01 pixels. The steady image accuracy of Gansu Jiayuguan fluctuates slightly in ± 0.02 pixels, and the steady image accuracy of the data in the three land cover types as a whole does not exceed ± 0.02 pixels, which fully illustrates the stability of the method in this paper.

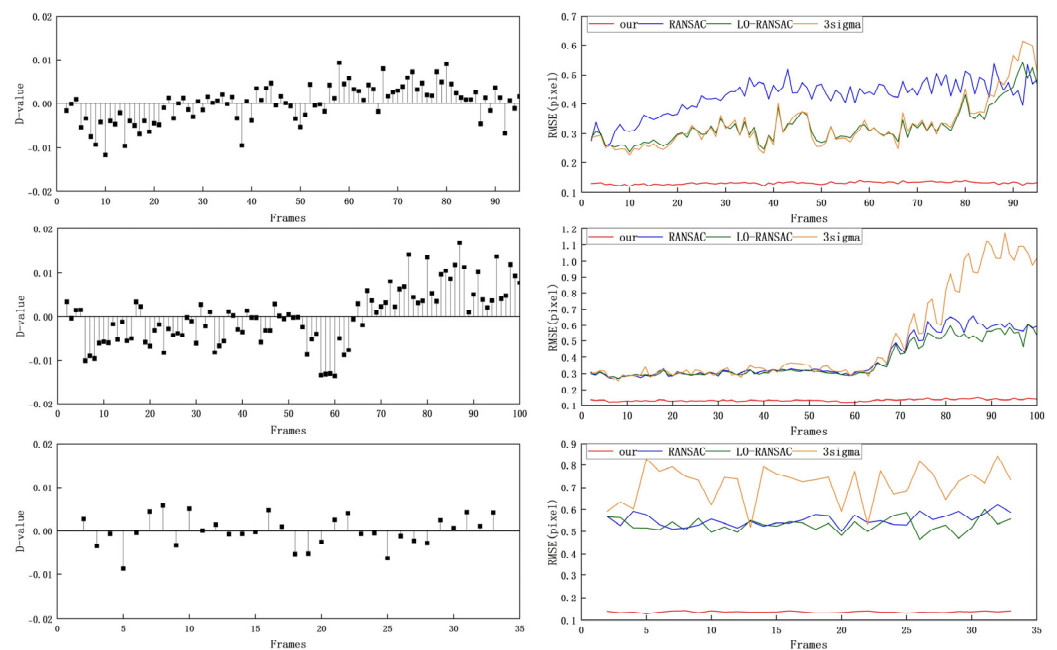


Figure 6. ED-RANSAC algorithm stability analysis (left) and comparison with RANSAC, LO-RANSAC, and 3sigma stability image results (right). From top to bottom: Zhifu Bay in Yantai, Jiayuguan in Gansu, and Leibo County in Sichuan. D-Value is the difference between the inter-frame RMSE and the average of the total RMSE.

The proposed ED-RANSAC algorithm has greatly improved performance compared to the original RANSAC, LO-RANSAC, and 3sigma algorithm. From the line chart on the right of Figure 6, it can be seen that the performance of the method proposed in this paper is the best. The stabilization accuracy in the marine area (Zhifu Bay in Yantai) has been improved to better than 0.15 pixels. The accuracy of the desert area (Jiayuguan in Gansu) is improved to better than 0.15 pixels. The stabilization accuracy in the mountainous area (Leibo County in Sichuan) has also been improved to better than 0.15 pixels, meeting the requirements for smooth video applications. This proves that the method proposed in this paper can eliminate the influence of terrain factors. The quantitative analysis of the stabilization accuracy is shown in Table 3.

Table 3. Video stabilization precision (RMSE) evaluation (Unit: Pixels).

Area	Methods	Max	Min	Median
Zhifu Bay in Yantai	RANSAC	0.5364	0.2576	0.4349
	LO-RANSAC	0.5408	0.2378	0.3081
	3sigma	0.6129	0.2277	0.3073
	Ours	0.1413	0.1202	0.1326
Jiayuguan in Gansu	RANSAC	0.6569	0.2631	0.3176
	LO-RANSAC	0.6065	0.2716	0.3129
	3sigma	1.1728	0.2508	0.3337
	Ours	0.1484	0.1181	0.1314
Leibo County in Sichuan	RANSAC	0.6217	0.4994	0.5485
	LO-RANSAC	0.6002	0.4648	0.5287
	3sigma	0.8417	0.5183	0.7359
	Ours	0.1421	0.1277	0.1359

Table 3 summarizes the RMSE of all frames obtained from the video stabilization experiments of the four algorithms on three different datasets. The maximum, minimum, and median RMSE values of all frames for all algorithms under the three datasets are recorded to quantitatively analyze the improvement in stabilization accuracy after the improvement of the RANSAC algorithm, compared with the other three algorithms. 3sigma has the worst results, with a maximum RMSE of more than 1.0 pixel in the desert region. LO-RANSAC and the original RANSAC algorithm show significant fluctuations in stabilization accuracy under the three different land cover types. The largest difference between the maximum and minimum RMSE values is around 0.4 pixels in the Jiayuguan in Gansu dataset, and the median RMSE is also above 0.3 pixels. The improved ED-RANSAC algorithm greatly improves the image stabilization accuracy so that the image stabilization accuracy under the conditions of three types of ground objects is better than 0.15 pixels, and the fluctuation of image stabilization accuracy is also greatly reduced. The overall image stabilization accuracy fluctuates around 0.03 pixels. The high accuracy and stability of the method in this paper are well-proven.

3.5.2. Overall Video Stabilization Precision Evaluation

By correcting the satellite video frame images, the geometric correspondence between the video frames can be restored. The experimental data were stabilized using the stabilization method proposed in this paper to obtain a stabilized video sequence. The first and last frame images of the stabilized sequence are shown in Figure 7. As can be seen from Figure 7, the effective coverage range of the two images on the ground differs significantly due to the influence of satellite platform shake and differences in the satellite video imaging angle.

To validate the effectiveness of the proposed method, image matching was performed on every 10 frames of the stabilized video sequence, using the first stabilized frame as the reference. The RMSE between corresponding points was used as the metric to measure the inter-frame matching accuracy. The results are shown in Table 4. It can be seen that the video stabilization accuracy obtained by this method is better than 0.15 pixels, which is consistent with the geometric accuracy between video frames in Table 3 and can meet the application requirements of high-precision satellite video stabilization. However, the Table 4 also shows that as the difference in frame numbers between the compared frames increases, the number of checkpoints decreases. Especially for the mountainous area data, the overall stabilization accuracy is not strictly meaningful due to the large difference in frame numbers of the data itself and the insufficient number of verification checkpoints, which is a problem that needs to be addressed in future work.

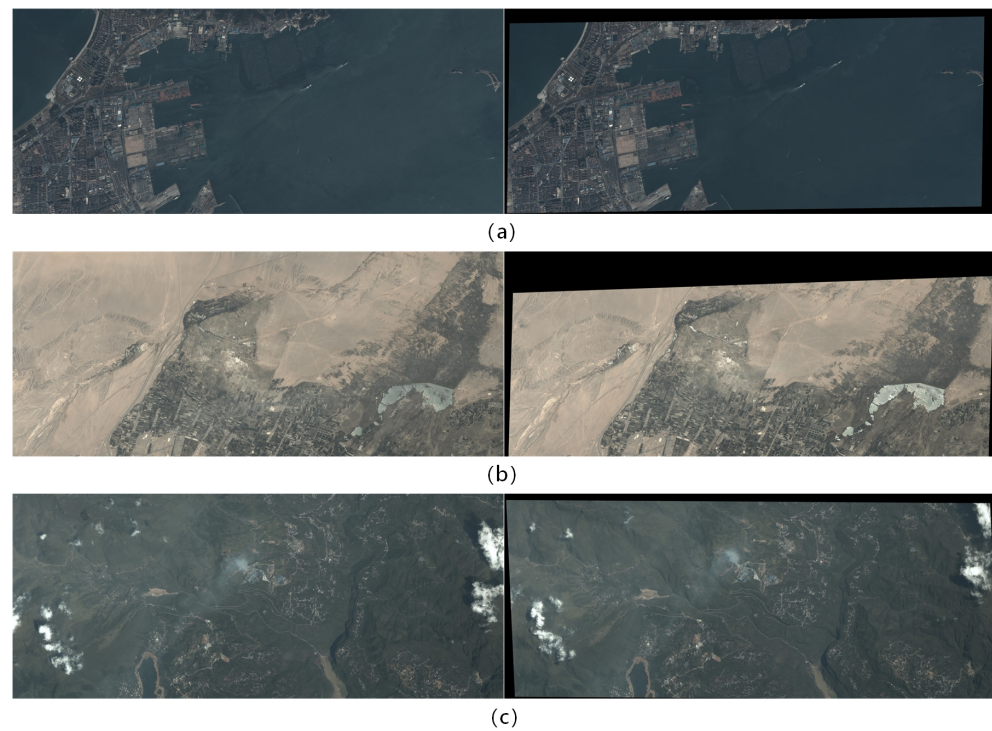


Figure 7. First (left) and last (right) steady images. (a) Zhifu Bay in Yantai. (b) Jiayuguan in Gansu. (c) Leibo County in Sichuan.

Table 4. Overall steady image accuracy verification (Optical, Unit: Pixels).

Frames	Zhifu Bay in Yantai		Jiayuguan in Gansu		Leibo County in Sichuan	
	Points	RMSE	Points	RMSE	Points	RMSE
1–10	6546	0.1330	4060	0.1363	54	0.1289
1–20	4636	0.1329	4592	0.1368	12	0.0990
1–30	2612	0.1371	3442	0.1380	6	0.1130
1–40/33	1557	0.1409	2921	0.1392	8	0.1218
1–50	1917	0.1374	2144	0.1396		
1–60	2525	0.1356	1783	0.1399		
1–70	1831	0.1368	1287	0.1374		
1–80	1406	0.1382	640	0.1364		
1–90	560	0.1398	509	0.1402		
1–100/94	871	0.1408	370	0.1379		
Average		0.1372		0.1381		0.1157

To visualize the video stabilization effect achieved by the method in this paper. In the obtained image sequence, the first frame and the last frame (combined with Table 4, the first frame and the tenth frame are selected in mountainous areas) are selected to show the local image edge map, and the local image edge map of the first frame and the last frame of the original image is listed below as a comparative display. For enhanced display, color processing was applied to one of the images. It can be seen from Figure 8 that the video stabilization effect between the two frames is excellent. There is no misalignment in areas such as water boundaries, buildings, roads, and farmland.

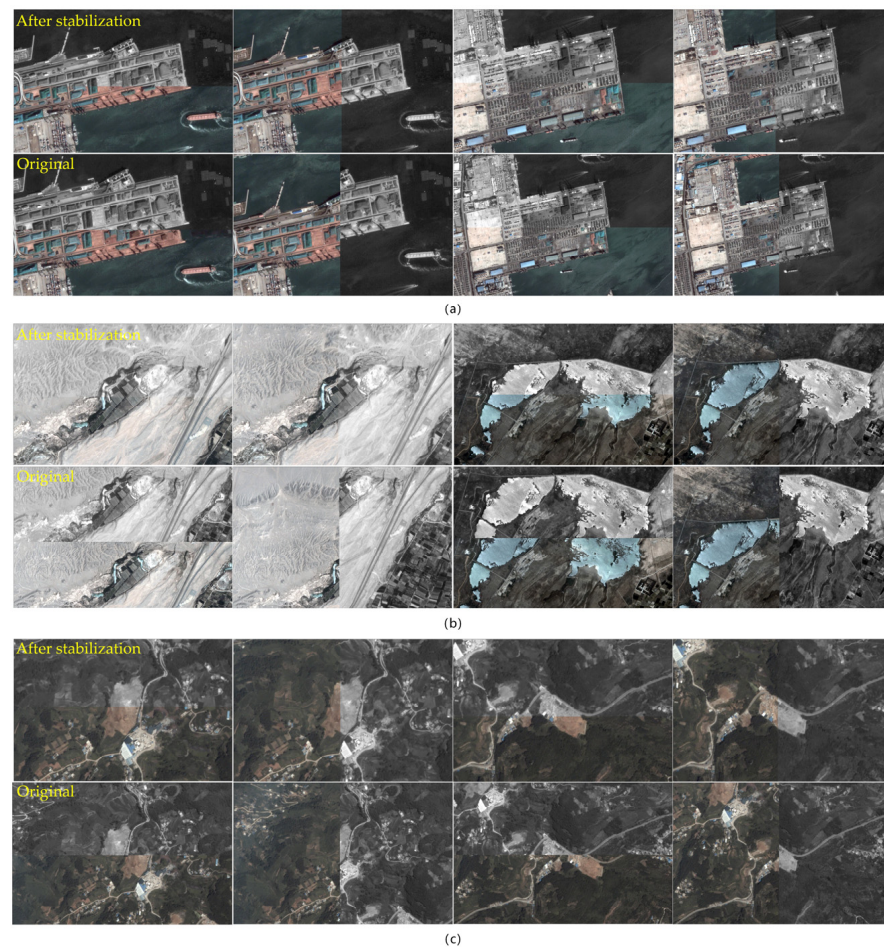


Figure 8. Diagram of adjacent image edges after resampling. (a) Zhifu Bay in Yantai. (b) Jiayuguan in Gansu. (c) Leibo County in Sichuan. The upper part is the corrected first frame and the last frame edge map, the under part is the original first frame and the last frame edge graph.

3.6. Application in SAR Video

3.6.1. Experimental Data

A SAR video released by Sandia National Laboratories is used as the experimental data in this section. The video size is 657×720 pixels, with a total of 150 frames. The video is shot in the “circular trajectory” mode [31,32], and the large displacement and angle change increase the difficulty of video stabilization. Figure 9 shows the experimental data of SAR video images in this paper.

3.6.2. Experimental Results and Analysis

Since there are significant differences between optical images and SAR images due to different imaging methods, the traditional SIFT algorithm cannot effectively detect the homologous features on SAR images. This paper adopts the SAR-SIFT algorithm instead of the SIFT algorithm for the detection of homologous points in the experimental process. This section conducts video stabilization experiments on SAR video data using the experimental process, methods, and evaluation indexes mentioned in the previous section. The method’s steady image stability verification results in the SAR video in this paper are shown in Figure 10. The comparison verification results of the ED-RANSAC RANSAC, LO-RANSAC and 3sigma algorithm are shown in Figure 11. The quantitative analysis of video stabilization accuracy is shown in Table 5.



Figure 9. SAR video image diagram.

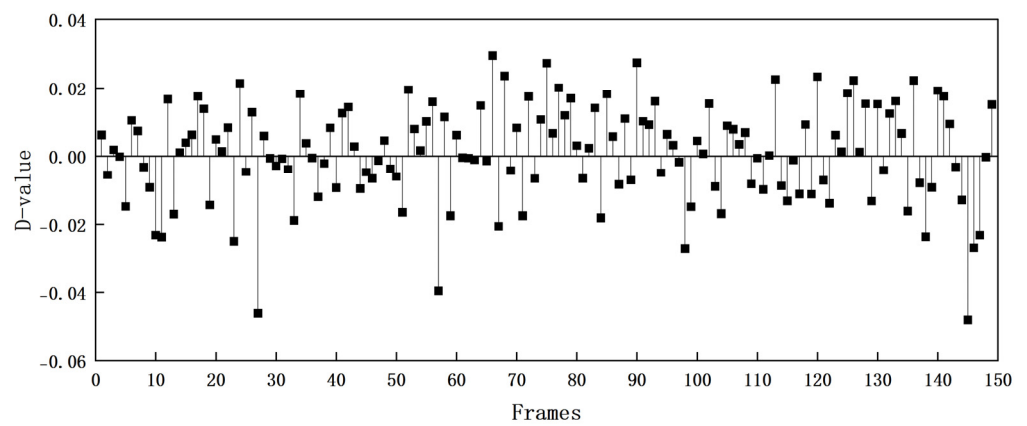


Figure 10. ED-RANSAC algorithm stability analysis. D-Value is the difference between the inter-frame RMSE and the average of the total RMSE.

Table 5. Steady image precision (RMSE) evaluation (Unit: Pixels).

Data Source	Method	Max	Min	Median
SAR Video	RANSAC	0.6944	0.5643	0.6182
	LO-RANSAC	0.4687	0.2676	0.3879
	3sigma	0.5918	0.2550	0.4012
	Our	0.2796	0.1469	0.2513

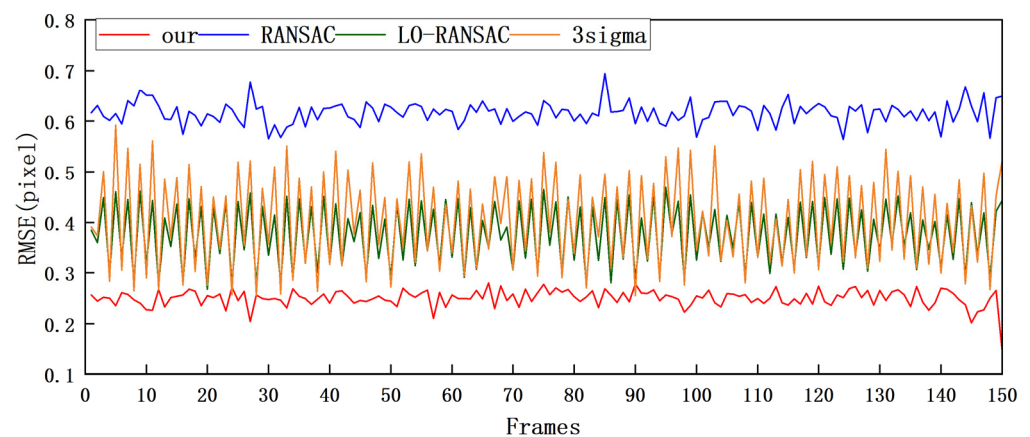


Figure 11. Our ED-RANSAC algorithm stability in comparison with RANSAC, LO-RANSAC, and 3sigma stability image results.

From Figure 10, it can be seen that the method in this paper also shows good stability in the stabilization of SAR video, and the stabilization accuracy does not exceed ± 0.05 pixels. It also proves the universality of the method in this paper again.

Figure 11 shows the comparison results between the algorithm in this paper and the other three algorithms. Table 5 shows the comparison results of the four algorithms quantitatively by the maximum, minimum, and median values of RMSE of all frames. This method improves the stabilization accuracy in SAR video stabilization from about 0.6 pixels before improving to about 0.25 pixels. The stabilization accuracy is improved significantly to meet the application requirements of high-precision satellite video stabilization.

The first frame of the output image sequence is used as the main frame to match with it every 10 frames in turn for overall accuracy verification, and the verification results are shown in Table 6. From Table 6, the overall accuracy of video steady is better than 0.3 pixels, which indicates that the method in this paper can also affect accuracy improvement for SAR video.

Table 6. Overall steady image accuracy verification (SAR, Unit: Pixels).

Frames	Checkpoints	RMSE	Frames	Checkpoints	RMSE
1–10	82	0.2387	1–90	18	0.2847
1–20	64	0.2008	1–100	11	0.2775
1–30	44	0.2697	1–110	17	0.2207
1–40	26	0.1061	1–120	19	0.1908
1–50	35	0.2881	1–130	14	0.2429
1–60	28	0.2720	1–140	16	0.1499
1–70	19	0.2463	1–150	13	0.1194
1–80	15	0.1971			
Average					0.2203

In order to visually interpret the video stabilization effect obtained by the method in this paper, the first frame and the last frame in the obtained stabilized video image sequence are selected to display their local image border maps. The partial image border maps of the first frame and the last frame of the original image are listed below as a comparison display. For enhance the display effect, one of the images is color processed. From Figure 12, this method has a good effect on stabilizing SAR video images, in which there is no misalignment of roads, flower beds, buildings, etc.

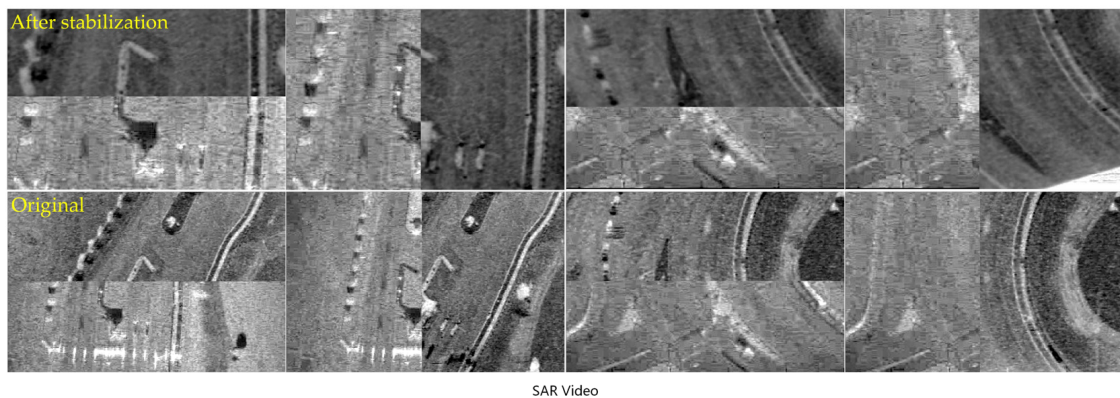


Figure 12. Diagram of adjacent image edges after resampling. The upper part is the corrected first frame and the last frame edge map, the under part is the original first frame and the last frame edge graph.

The above figures and tables indicate that the method of this article is not only applicable to different types of ground conditions in optical satellite video data, but also has good performance in the application of SAR video stabilization. The universality, high precision, and stability of the proposed method are fully proved.

4. Conclusions

This paper proposes a high-precision satellite video stabilization method based on the ED-RANSAC operator, which achieves superior stabilization results. The following conclusions are drawn:

- (1) The proposed ED-RANSAC method realizes high-precision feature extraction and matching of homologous points and provides a reliable data guarantee for video stabilization.
- (2) The method achieves a better stabilization accuracy than 0.15 pixels for various optical satellite video land cover types and better than 0.3 pixels for SAR video, meeting the high-accuracy requirements of satellite video stabilization.
- (3) The ED-RANSAC-based method in this paper accomplished high-precision stabilization of multi-source video loads without considering factors such as geometric models and terrain changes. Further research will be conducted to improve the accuracy of SAR video stabilization.

Author Contributions: Conceptualization, G.Z.; methodology, F.Z., X.L. and T.W.; validation, F.Z., J.H., Q.C. and T.D.; writing—original draft preparation, F.Z. and X.L.; writing—review and editing, F.Z., X.L. and T.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Civil Aerospace Technology Advance Research Project of National Defense Science and Engineering (No. D040107); Foundation Strengthening Fund Project, grant number 2021-JCJQ-JJ-0251.

Acknowledgments: We give thanks to the research team at Wuhan University for providing data support. Furthermore, the authors would like to thank the reviewers for their helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Man, Y.Y.; Li, H.C. Imaging Characteristics Analysis for LEO Video Satellite. *Spacecr. Eng.* **2015**, *24*, 52–57.
2. Zhang, G. Satellite Video Processing and Applications. *J. Appl. Sci.* **2016**, *34*, 361–370.
3. Yi, M. *Image Registration Based on Feature Point and Its application to Electronic Image Stabilization*; Xidian University: Xi'an, China, 2013.
4. Kim, T.; Lee, S.; Paik, J. Combined Shape and Feature-based Video Analysis and Its Application to Non-rigid Object Tracking. *IET Image Process.* **2011**, *5*, 87–100. [[CrossRef](#)]
5. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]

6. Wang, T.; Li, X.; Zhang, G.; Lin, M.; Deng, M.; Cui, H.; Jiang, B.; Wang, Y.; Zhu, Y.; Wang, H.; et al. Large-Scale Orthorectification of GF-3 SAR Images Without Ground Control Points for China's Land Area. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5221617. [[CrossRef](#)]
7. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; IEEE: New York, NY, USA, 1999; Volume 2, pp. 1150–1157.
8. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-Like Algorithm for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 453–466. [[CrossRef](#)]
9. Tang, L.; Tang, W.; Qu, X.; Han, Y.; Wang, W.; Zhao, B. A Scale-Aware Pyramid Network for Multi-Scale Object Detection in SAR Images. *Remote Sens.* **2022**, *14*, 973. [[CrossRef](#)]
10. Han, Y.; Liu, H.; Wang, Y.; Liu, C. A Comprehensive Review for Typical Applications Based Upon Unmanned Aerial Vehicle Platform. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9654–9666. [[CrossRef](#)]
11. Li, X.; Wang, T.; Cui, H.; Zhang, G.; Cheng, Q.; Dong, T.; Jiang, B. SARPointNet: An Automated Feature Learning Framework for Spaceborne SAR Image Registration. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 6371–6381. [[CrossRef](#)]
12. Beijing Institute of Space Electromechanical Research. A DEM-Assisted Image Stabilization Method for Optical Video Satellites. CN201910964700.5, 28 February 2020. Available online: <https://www.cast.cn/english/channel/1808> (accessed on 8 May 2023).
13. Zhou, N.; Cao, J.S.; Xiao, L.; Cao, S. Object image stabilization method of optical video satellite with geocoding. *J. Wuhan Univ.* **2022**, *48*, 308–315.
14. Zhang, G.; Li, B.B.; Jiang, Y.H. Research on image stabilization of satellite-borne video with geometric model constraints. In Proceedings of the 2015 Small Satellite Technology Exchange Conference, Beijing, China, 1 June 2015; pp. 452–456.
15. Wang, X.; Zhang, G.; Shen, X.; Li, B.; Jiang, Y. Satellite video image stabilization considering image plane distortion. *J. Surv. Mapp.* **2016**, *45*, 194–198.
16. Li, F. *Research on the Electronic Image Stabilization and Target Tracking Algorithm for Space Infrared Earth Observation Video Camera*; Shanghai Institute of Technical Physics of CAS: Shanghai, China, 2018.
17. Kumar, S.; Azartash, H.; Biswas, M.; Nguyen, T. Real-Time Affine Global Motion Estimation Using Phase Correlation and Its Application for Digital Image Stabilization. *IEEE Trans. Image Process.* **2011**, *20*, 3406–3418. [[CrossRef](#)]
18. Zhang, M.L.; Chen, J.G.; Yuan, H.Y. Video Stabilization on a Six-rotor Aircraft Platform. *J. Tsinghua Univ.* **2014**, *54*, 1412–1416.
19. Murthy, K.; Shearn, M.; Smiley, B.D.; Chau, A.H.; Levine, J.; Robinson, M.D. SkySat-1: Very high-resolution imagery from a small satellite. In Proceedings of the Sensors, Systems, and Next-Generation Satellites XVIII, Amsterdam, The Netherlands, 22–25 September 2014; SPIE: Bellingham, WA, USA, 2014; Volume 9241, pp. 367–378.
20. Xing, H.; Yan, J.L.; Zhang, S.J. Digital Image Stabilization Using Kalman Filtering. *Acta Armamentarii* **2007**, *28*, 175–177.
21. Walha, A.; Wali, A.; Alimi, A.M. Video Stabilization with Moving Object Detecting and Tracking for Aerial Video Surveillance. *Multimed. Tools Appl.* **2015**, *74*, 6745–6767. [[CrossRef](#)]
22. Yan, S.; Li, Y.; Zhou, Z. Real-time motion compensation of an airborne UWB SAR. In Proceedings of the 2011 8th European Radar Conference, Manchester, UK, 12–14 October 2011; pp. 305–308.
23. Linnehan, R.; Miller, J.; Asadi, A. Map-drift autofocus and scene stabilization for video-SAR. In Proceedings of the 2018 IEEE Radar Conference (RadarConf18), Oklahoma City, OK, USA, 23–27 April 2018; IEEE: New York, NY, USA, 2018; pp. 1401–1405.
24. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
25. Chum, O.; Matas, J.; Kittler, J. Locally optimized RANSAC. In Proceedings of the Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, 10–12 September 2003; Proceedings 25. Springer: Berlin/Heidelberg, Germany, 2003; pp. 236–243.
26. Wang, Y.; Wang, T.; Zhang, G.; Cheng, Q.; Wu, J.Q. Small target tracking in satellite videos using background compensation. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7010–7021. [[CrossRef](#)]
27. Wang, T.; Hong, J.; Han, Y.; Zhang, G.; Chen, S.; Dong, T.; Yang, Y.; Ruan, H. AOSVSSNet: Attention-guided optical satellite video smoke segmentation network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8552–8566. [[CrossRef](#)]
28. Chen, S.; Wang, T.; Wang, H.; Wang, Y.; Hong, J.; Dong, T.; Li, Z. Vehicle Tracking on Satellite Video Based on Historical Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7784–7796. [[CrossRef](#)]
29. Wang, K.S.; Liu, M. Talking about the application of image affine transformation. *Inf. Technol. Informatiz.* **2015**, *1*, 155–156.
30. He, D.J.; Geng, N.; Long, M.S. *Digital Image Processing*; Xidian University Press: Xi'an, China, 2015.
31. Hudgens, J.J. *Advancements in Synthetic Aperture Radar (SAR) for Improved ISR*; Sandia National Lab (SNL): Albuquerque, NM, USA, 2014.
32. Wells, L.; Sorensen, K.; Doerry, A.; Remund, B. Developments in SAR and IFSAR systems and technologies at sandia national laboratories. *Proc. IEEE Aerosp. Conf.* **2005**, *2*, 1085–1095.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.