remote sensing

MDPI

Article

UAV Low-Altitude Aerial Image Stitching Based on Semantic Segmentation and ORB Algorithm for Urban Traffic

Gengxin Zhang [1], Danyang Qin [1,*], Jiaqiang Yang [1], Mengying Yan [1], Huapeng Tang [1], Haoze Bie [1] and Lin Ma [2]

[1] Department of Electronic and Communication Engineering, Heilongjiang University, Harbin 150080, China
[2] Department of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150080, China
* Correspondence: qindanyang@hlju.edu.cn

Abstract: UAVs are flexible in action, changeable in shooting angles, and complex and changeable in the shooting environment. Most of the existing stitching algorithms are suitable for images collected by UAVs in static environments, but the images are in fact being captured dynamically, especially in low-altitude flights. Considering that the great changes of the object position may cause the low-altitude aerial images to be affected by the moving foreground during stitching, so as to result in quality problems, such as splicing misalignment and tearing, a UAV aerial image stitching algorithm is proposed based on semantic segmentation and ORB. In the image registration, the algorithm introduces a semantic segmentation network to separate the foreground and background of the image and obtains the foreground semantic information. At the same time, it uses the quadtree decomposition idea and the classical ORB algorithm to extract feature points. By comparing the feature point information with the foreground semantic information, the foreground feature points can be deleted to realize feature point matching. Based on the accurate image registration, the image stitching and fusion will be achieved by the homography matrix and the weighted fusion algorithm. The proposed algorithm not only preserves the details of the original image, but also improves the four objective data points of information entropy, average gradient, peak signal-to-noise ratio and root mean square error. It can solve the problem of splicing misalignment tearing during background stitching caused by dynamic foreground and improves the stitching quality of UAV low-altitude aerial images.

Keywords: drone aerial; image stitching; semantic segmentation; ORB algorithm

Citation: Zhang, G.; Qin, D.; Yang, J.; Yan, M.; Tang, H.; Bie, H.; Ma, L. UAV Low-Altitude Aerial Image Stitching Based on Semantic Segmentation and ORB Algorithm for Urban Traffic. *Remote Sens.* **2022**, *14*, 6013. https://doi.org/10.3390/rs14236013

Academic Editors: Ming Li and Zhaohui Li

Received: 23 October 2022
Accepted: 24 November 2022
Published: 27 November 2022

1. Introduction

As a new type of industrial technology, UAV aerial photography is born in response to demand, develops rapidly, and is widely used in various fields. Due to their small size and flexible action, UAVs can survey and take pictures of areas that are inconvenient for humans to set foot in. For example, detecting and tracking enemy targets through UAV aerial images can significantly support military command in military surveillance. In earthquakes, plane crashes, and other disasters, the aerial images of drones are stitched into a real-time panorama, which can help rescuers quickly understand the disaster scene, search for trapped people, and reasonably plan rescue routes. In addition, in environmental governance, drone photography can be used to obtain images of the surrounding environment of the factory, to determine the source of pollution quickly, and then provide adequate assistance for law enforcement departments to investigate and visit. However, UAVs are limited by many factors, such as shooting height, angle, and UAV performance when shooting images, resulting in incomplete coverage of a single aerial image. It can be seen that the research on UAV aerial images is of great significance.

At present, among the many methods of aerial image stitching, the method based on image feature stitching occupies an important position. For the stitching of aerial images based on features, image registration and image fusion are two crucial steps, and the

accuracy of registration will directly affect the quality of stitching. There are many methods for extracting image feature points. The Harris corner detection algorithm [1] essentially uses a local window to detect grayscale information on the image. If the grayscale value changes greatly, there are corners in the window. This algorithm improves the accuracy and speed of the feature extraction algorithm, but the accuracy is very low when the image scale changes. The SIFT (Scale-Invariant Feature) algorithm [2] is different from it. First, the feature points are detected in the constructed Gaussian difference image pyramid, and then the corresponding feature descriptors containing 128 dimensions are extracted. The algorithm solves the problem of scale invariance, but the overall time-consumption is still relatively long, and the implementation process is relatively complex. To solve the time-consuming problem of extracting feature points, Bay et al. proposed the SURF (Speeded Up Robust Features) algorithm [3]; the algorithm detects feature points in the constructed Hessian matrix and reduces the 128-dimensional feature descriptor to a 64-dimensional feature descriptor, so the feature point extraction speed has been greatly improved. To further improve the speed of the feature extraction algorithm, Rublee et al. [4] retained the advantages of the FAST feature point detection algorithm and the Brief feature point description algorithm, and on this basis comprehensively improved and proposed the ORB (Oriented Fast and Rotated Brief) algorithm. The algorithm outperforms the SURF algorithm in speed and is robust to illumination and scale. The ORB feature is a very representative real-time image feature at present. The algorithm based on the ORB feature is superior to several other popular algorithms in terms of speed while ensuring the feature points have rotation and scale invariance. The UAV image has the characteristics of a large amount of data, so the speed of the image processing algorithm is very strict [5]. Considering the processing speed, this paper selects the ORB algorithm to extract the feature points of the aerial image.

Researchers have been making continuous efforts to optimize the stitching quality of aerial images by improving the accuracy of image matching, solving the problem of uneven distribution of feature points, and optimizing the quality of image stitching. The quadtree algorithm proposed by Mur-Artal et al. [6,7] focuses on the extraction and screening of feature points. By eliminating redundant feature points, the feature points of the entire aerial image are uniformly distributed; in 2017, Li et al. [8] proposed a new dynamic programming algorithm for aerial image fusion. The algorithm first corrected the images so that the aerial images were located in the same coordinate system, thereby improving the quality of image fusion and solving the problem. For the spatial dislocation problem in the stitching of UAV images, in 2018, Zhang et al. [9] designed and completed the overall work of UAV remote sensing image stitching based on the global stitching strategy optimized by Levenberg–Marquardt (L-M). Compared with the previous method, the drift problem of multi-frame image stitching is effectively improved; in 2019, Zhao et al. [10] optimized the number of matchable feature points by dynamically setting the contrast threshold in the Gaussian scale space, but the algorithm left the problem of poor real-time performance. Based on the above basis, Xie et al. [11] proposed a new projection plane selection strategy, which improved the influence of UAV flight posture changes on image stitching. In 2020, Xu et al. [12] proposed a stitching method based on multi-region guided local projection deformation, which reduces ghosting caused by projection changes with viewpoint and parallax. In 2021, Fan et al. [13] proposed a fast adaptive stitching algorithm for processing a large number of aerial images, by reducing the estimated reprojection error and the number of observed visual distortions, culling densely overlapping images from all images, improving the quality of the stitched image. Yuan et al. [14] proposed a superpixel-based slitting strategy, Xue et al. [15] proposed to identify and eliminate ghosting through multi-component collaboration, and Chen et al. [16] proposed a stitching strategy based on optimal seams and half-projection warping in 2022; they all optimize the stitching quality of the images.

UAV aerial image stitching has become a key research issue. However, in previous research on aerial image stitching, most of them focus on the generation of panoramic

images in static environments, and more attention was paid to improving the stitching speed. In low-altitude environments, the generation of panoramas with dynamic objects is a difficult problem. The traditional single static panorama generation method cannot effectively solve this problem [17]. In dynamic panoramas, the movement of foreground objects will cause the feature points to be mismatched, which will lead to splicing misalignment and tears in the generated background image. In view of this, this paper uses the semantic segmentation method to process the image, and separates the dynamic foreground from the static background, so as to retain the semantic information of the foreground.

Semantic segmentation is one of the cornerstone technologies in image processing. Semantic segmentation achieves pixel-level classification by assigning labels to each pixel in the image. This label is predefined and represents the semantic category of things in the image. Generally speaking, the semantic segmentation of images is to distinguish the specific category and position of things in the image by pixel and give the same color to the things of the same category for representation. Since the 1980s and 1990s, researchers have manually segmented original images based on features, such as color, gray level, texture and shape. The traditional segmentation methods mainly include threshold segmentation, edge segmentation and superpixel-based segmentation. Although the traditional segmentation method has achieved good results, there are many shortcomings, such as time consumption, poor adaptability and so on, thus it has been difficult to make a breakthrough in the segmentation results.

In recent years, with the rapid development of artificial intelligence, deep learning has brought earth-shaking changes to many fields, and computer vision is no exception to introduce it into image semantic segmentation to obtain accurate prediction results. In 2006, Hinton et al. put forward the idea of deep learning. By exploring the rules and features of data, the combination between low levels is used to represent the feature categories of attributes, so that the machine has the ability to classify and learn text, video, image and other content. Deep learning methods include convolutional neural network (CNN) [18] and recurrent neural network (RNN) [19] and generative adversarial network (GAN) [20]. Among them, the most typical convolutional neural network (CNN) has achieved far higher accuracy than the traditional method in international competitions.

The Lenet-5 [18] network proposed by Le Cun et al., using only a simple five-layer network, pioneered convolutional neural networks and provided a new idea for future generations to innovate. In 2014, Simonyan K et al., the vision research group of Oxford University, proposed VGGNet [21] network, which used 3×3 small convolution mode instead of large convolution mode, and divided into two essentially identical versions, VGG16 and VGG19, according to different network layers. In 2015, Google released 22 layers of GoogleNet [22] to integrate modules and modify the network. The increase in network depth was obviously the main research direction of that era, but as the number of network layers reached the bottleneck, scientists found that only increasing the number of network layers would not make the correct rate stack, but also cause the problem of complex model convergence and so on. In 2016, He Kaiwen et al., proposed deep residual network ResNet [23] based on this problem, designed residual blocks to link high–low input signals to solve network degradation and gradient dispersion problems. Compared with VGG, residual network has smoother forward and back propagation modes, so it gradually replaces VGG as the more commonly used backbone network. In 2015, Long et al. proposed the Fully Convolutional neural Network (FCN) in the International Conference on Computer Vision and Pattern Recognition [24]. Compared with the traditional algorithm, it has achieved breakthrough results and opened the door based on deep learning. Since then, more and more excellent semantic segmentation algorithms have been emerging. In the FCN network, based on the VGG-16 network, codec network structure is adopted [25–27]. The original fully connected layer of VGG network is discarded in encoder, and the upsampling classification network constructed by convolution layer is proposed. Meanwhile, deconvolution operation is adopted in decoder to upsample the feature map. Through the cross-layer connection strategy, the

low-level rich location features are combined with high-level semantic features, and then end-to-end image semantic segmentation is realized. Although the FCN will lose part of the location information when the decoder is pooled, compared with the traditional method, it proves that the convolutional neural network can effectively segment the image, which points out the way for subsequent research of experts and scholars, so more and more researchers are trying new design ideas. This paper uses the classical FCN semantic segmentation network.

This paper proposes an aerial image stitching algorithm based on semantic segmentation and ORB. Firstly, the images to be spliced are semantically segmented through the FCN network to extract foreground objects. Meanwhile, the ORB algorithm is used to extract the feature points of the two images. If the feature points are in the foreground range, the culling operation is performed, and only the feature points in the background are retained to reduce the occurrence of mismatched pairs, thereby improving the quality of aerial image stitching.

2. Materials and Methods

The traditional ORB algorithm is mainly composed of two parts: the FAST feature point detection algorithm [28] and the BRIEF feature point description algorithm [29], but the feature points extracted by the traditional ORB algorithm will be unevenly distributed. As a result, the feature points in some areas are dense, while the feature points in some areas are extremely sparse, which affects the quality of image stitching. This paper solves the problem of uneven distribution of feature points by introducing the idea of quadtree decomposition and realizes the uniform distribution of feature points in the image. Furthermore, it is considered that the wrong matching of the feature points located on the dynamic objects in the image will lead to misalignment in the static background stitching. This paper first performs semantic segmentation on the images to be stitched, extracts the moving foreground objects, compares the foreground pixel coordinate values with the feature point coordinate values, removes the feature points distributed in the foreground, and then performs matching. This greatly reduces the probability of misalignment in the background stitching of aerial images caused by dynamic foregrounds and improves the quality of stitching.

The process of this paper is shown in Figure 1. Figure 1 mainly consists of the first set of experimental images and their stitching results, as well as Figures 2, 3, 5, 6, 8, 12 and 13 in this paper. As indicated by the arrow in Figure 1, the UAV is used to shoot the road with vehicles at low altitudes, and then the ORB algorithm based on quadtree is used to extract the feature points of the image. At the same time, semantic segmentation is performed on the image to extract the moving vehicle. Next, the semantic information on the vehicle is compared with the feature point information, and the feature points on the vehicle are eliminated, the result is shown in the green marked box in the upper right corner of Figure 1. Finally, the images with the feature points on dynamic objects, such as vehicles, removed are spliced, and the splicing map and the contrast effect are shown in the red marked box.

2.1. Improved ORB Algorithm to Extract Feature Points

2.1.1. FAST Detection Algorithm

The original FAST feature point calculation is very fast, and its core idea is to find the distinctive point by comparing the difference between the pixel brightness. The FAST detection algorithm will first randomly select a pixel p in the detection image and assume that its brightness value is I_p. Then, a circle with radius of 3 will be drawn with this pixel p as the center of the circle, on which 16 consecutive pixel points can be obtained. When the circumference radius is 3, the accuracy and efficiency of detection can be taken into account at the same time, and the detection effect is the best. Set the threshold T at the same time. The upper limit is $I_p + T$, and the lower limit of brightness is $I_p - T$. If the brightness of N

consecutive points exceeds this range, then p is defined as a feature point. The calculation formula is as follows:

$$N = \sum_{x \subset (circle(p))} |I(x) - I(p)| \geq T \quad (1)$$

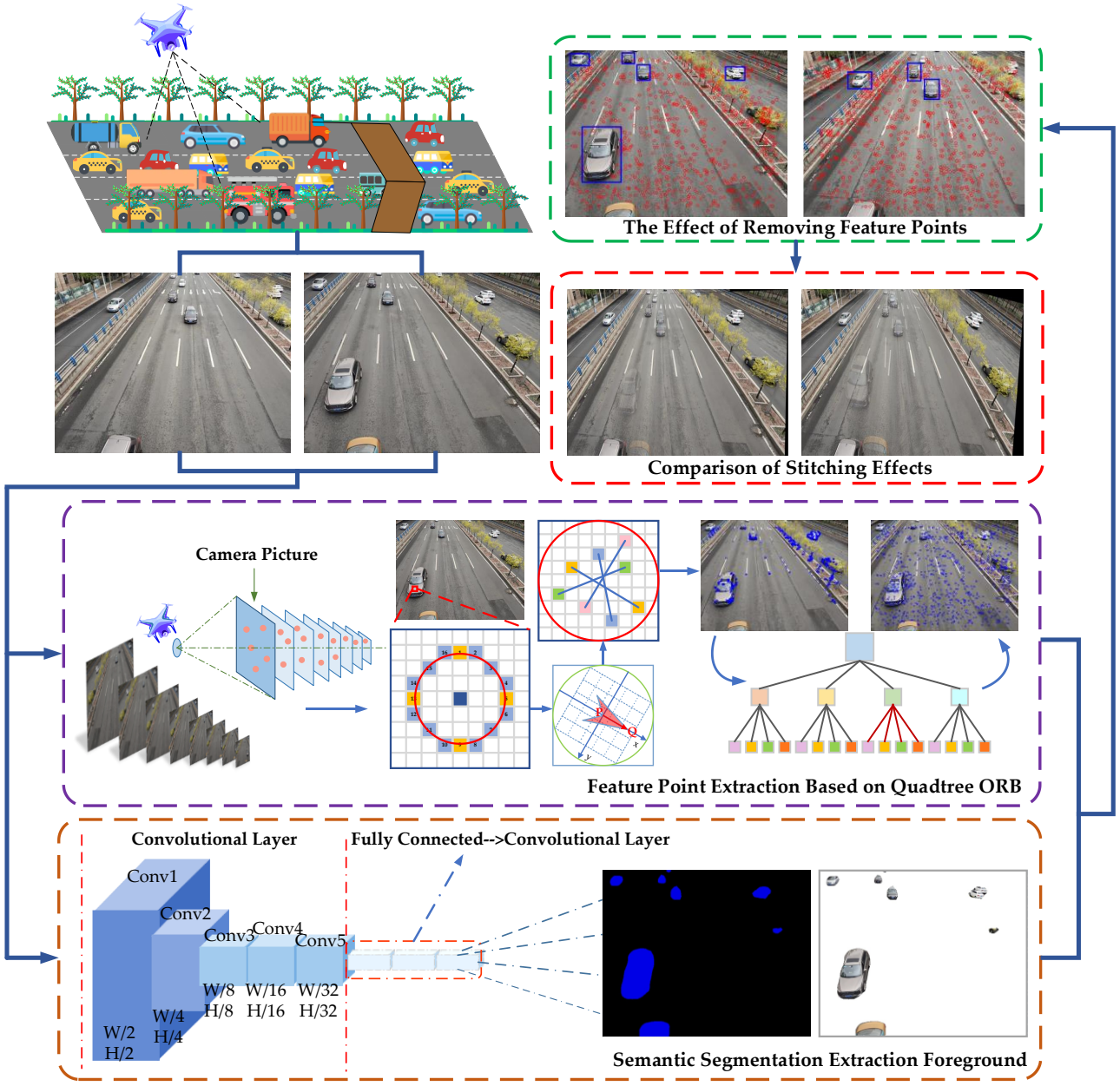


Figure 1. The flow chart of the algorithm in this paper.

In actual detection, if each feature point is detected, the difference between the 16 pixels on the circle and the center point p must be calculated, which will affect the detection efficiency. To be more efficient, FAST adds a pre-detection operation. As shown in Figure 2, take any pixel on the aerial image, and only detect the pixel located at the yellow square position on the circle. When three of these four pixels are out of range at the same time, the current pixel is the candidate feature point of the image, otherwise, it is directly excluded. This method greatly improves detection efficiency.

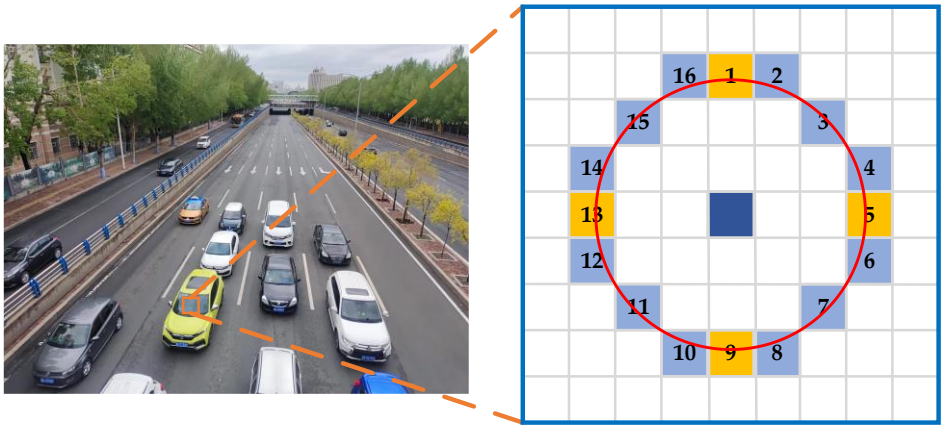


Figure 2. FAST detection feature point location map.

The feature points detected by FAST do not have rotation invariance. In order to make up for the deficiency, the gray centroid in the neighborhood of each feature point is calculated, then take the feature points as the starting point and the solved gray centroid as the endpoint to construct a vector, and this vector direction is the main direction of the FAST feature point. The grayscale centroid is shown in Figure 3. Let $I(x, y)$ represent the gray value of each pixel point (x, y) in the image to be detected, then the neighborhood moment of the feature point is:

$$m_{pq} = \sum_{x,y \in r} x^p y^q I(x,y), p, q = \{0, 1\} \tag{2}$$

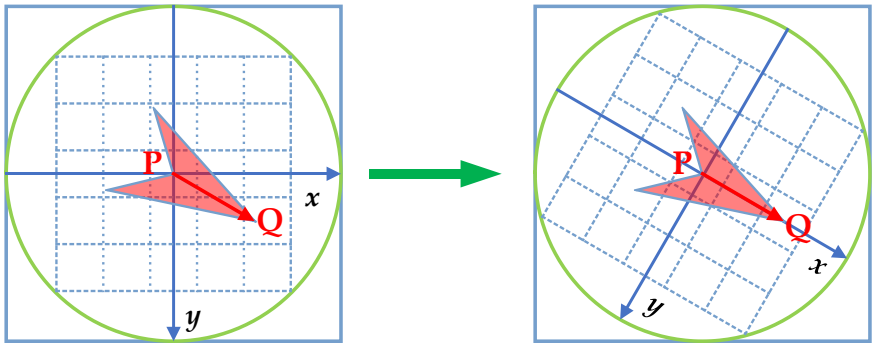


Figure 3. Grayscale centroid map.

The centroid of the image moment can be found by the neighborhood moment:

$$Q = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \tag{3}$$

Then the main direction of the FAST feature point is:

$$\theta = \arctan\left(\frac{m_{01}}{m_{10}} \right) \tag{4}$$

2.1.2. Image Pyramid

The image pyramid is a multi-scale representation of the image, and the image pyramid can be obtained by sampling the original image with different sampling rates. Therefore, an image pyramid is constructed for the image to be detected, which can achieve scale invariance. Set the scaling factor of the pyramid to 1.2, and downsample the image to be detected at a scaling ratio of 1/1.2 to obtain 8 pictures, and then perform feature extraction on each picture and record the number of layers of the pyramid where the feature points are located, to obtain the image feature points. As shown on the left side of Figure 4, this

paper constructs an eight-layer pyramid based on the above principles to be stitched, which solves the problem that the feature points do not have scale invariance.

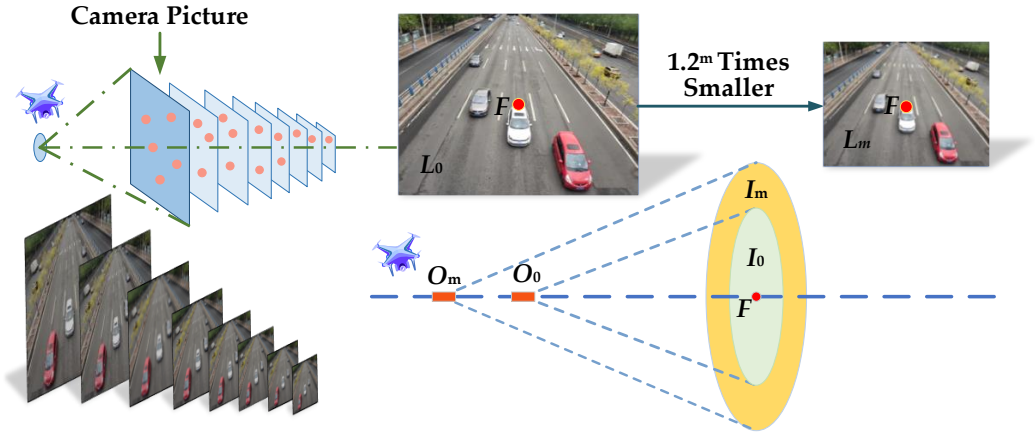


Figure 4. Image pyramid and distance relationship.

In the process of constructing a pyramid of an image to extract feature points, the image scaling does not cause the position of the feature points to move vertically. Suppose there is a feature point F_m on the mth layer, and its spatial position and the position of the camera center during aerial photography are d_m. According to the imaging characteristics of a "small objective lens", if the patch corresponding to the feature point of the mth layer is of the same size as the patch corresponding to the 0th layer of the pyramid, the maximum object distance that is constant in scale is $d_{\max} = d_m \times 1.2^{\mathrm{m}}$. According to the imaging characteristics of "close-up image is large", if you want to move the feature points of the current mth layer to the seventh layer, the real camera imaging image must be magnified by the original $1.2^{7-\mathrm{m}}$ times, so it corresponds to the minimum object distance $d_{\min} = d_m \times 1.2^{\mathrm{m}-7}$. The right side of Figure 4 shows the corresponding relationship between the feature point patches obtained after scaling the original image by $1/1.2^{\mathrm{m}}$ times, and the relationship between the imaging distance of the aerial camera and the scene distance.

2.1.3. Brief Feature Description

The detected ORB feature points need to be characterized by the Brief algorithm. BRIEF is a binary descriptor whose description vector consists of 0s and 1s. The core idea is to select n point pairs in a specific pattern around the feature points and combine the comparison results of the point pairs as a descriptor. The basic description idea is shown in Figure 5 below:

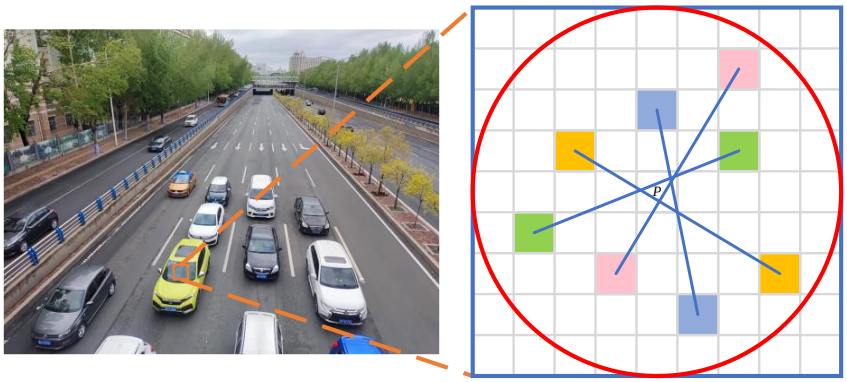


Figure 5. Descriptor formation diagram.

First, the image is processed by Gaussian filtering to reduce noise interference. In the $S \times S$ pixel area centered on the feature point p, n point pairs are randomly selected for gray value comparison, and the following binary assignments are performed:

$$\tau(p;x,y) = \begin{cases} 1, p(x) < p(y) \\ 0, p(x) \geq p(y) \end{cases} \tag{5}$$

Through comparison, the n-dimensional binary feature descriptor can be obtained as:

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p;x_i,y_i) \tag{6}$$

Next, define a matrix Q of order $2 \times n$, which is expressed as:

$$Q = \begin{bmatrix} x_1, x_2 \cdots, x_n \\ y_1, y_2, \cdots, y_n \end{bmatrix} \tag{7}$$

Rotate the matrix θ by the rotation matrix R_θ corresponding to the main direction Q of the feature point to obtain a new matrix Q_θ:

$$Q_\theta = R_\theta Q = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1, x_2, \cdots, x_n \\ y_1, y_2, \cdots, y_n \end{bmatrix} \tag{8}$$

Finally, the descriptor with direction can be expressed as:

$$g_n(p,\theta) = f_n(p) | (x_i, y_i) \in Q_\theta \tag{9}$$

2.1.4. Uniform Feature Points Based on Quadtree

In the original algorithm, the FAST detection speed is breakneck. However, the extracted feature points are often too concentrated in the area with pronounced texture, while it is difficult to extract feature points in the area with weak texture. As a result, the uneven distribution of feature points occurs, leading to decreased matching accuracy. According to the idea of the quadtree, this paper removes redundant and invalid feature points in the picture so that the feature points can be distributed more evenly [11]. Figure 6 shows the division idea of the quadtree. In simple terms, the concept of a quadtree is to divide each space into four small areas equally, and so on recursively, stop dividing after reaching a certain depth or meeting specific requirements.

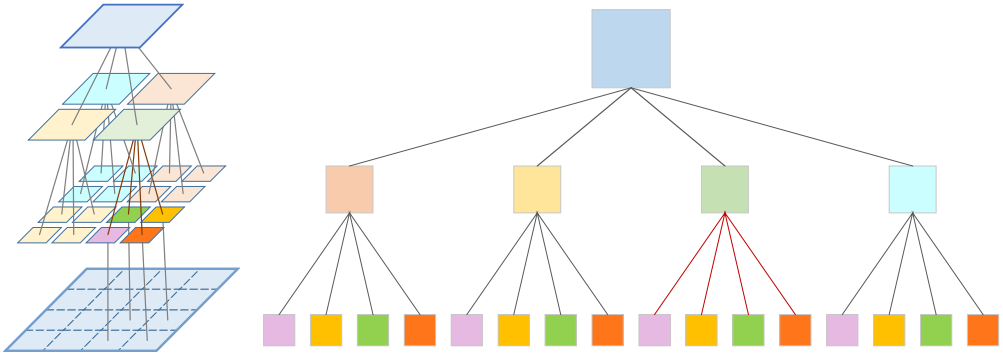


Figure 6. Quadtree partition idea.

As shown in Figure 7, the distribution of feature points extracted from aerial images by the traditional ORB algorithm is not uniform. Therefore, based on the quad-tree idea, this paper again screened the feature points extracted by FAST on each layer of the pyramid. First, calculate the initial node to obtain the initial quad-tree structure, represented by four quadrants (UL, UR, BL, BR), and then map the feature points to the initialized quad-tree. The number of feature points in the initialized node is not 1. On this basis, the image is divided into four nodes, n1, n2, n3, and n4, which are mapped to the child nodes according to the position of the feature points. If the number of feature points in the node is greater than 1, each node continues to be split into four. When the number of feature points in the node is greater than the expected number of feature points, the splitting is stopped. Finally, each nodes' feature points with the best quality are extracted. It can be seen from Figure 7 that after processing, the feature points are evenly distributed on the aerial image.

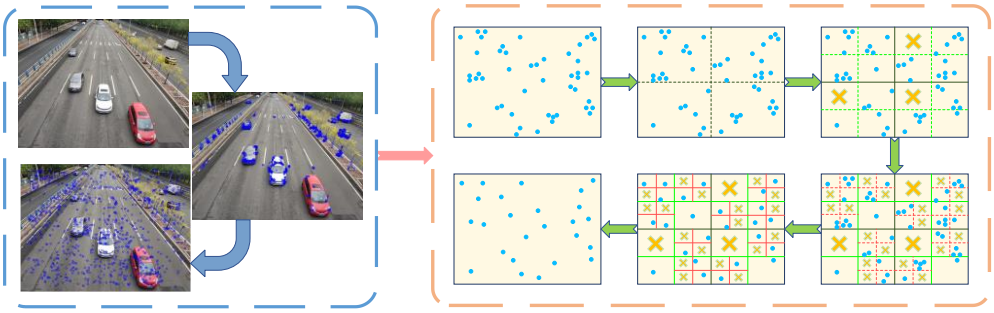


Figure 7. Quadtree uniform feature points. The blue dots indicate the feature points extracted from the image. Orange x is indicate the absence of feature points in the divided region.

2.2. Semantic Segmentation Extraction Foreground

To reduce the quality problem of aerial image stitching caused by the feature points of dynamic foreground, this paper separates the foreground and background of the image to be stitched through semantic segmentation to prepare for the removal of foreground feature points. This paper uses the classic semantic segmentation network FCN [24]. FCN classifies the objects in the image by pixel-level classification of the image. Unlike the classic CNN, FCN has no specific requirement on the input image size. FCN replaces the final fully connected layer of CNN with a convolutional layer, uses the deconvolution layer to upsample the feature map of the last convolutional layer, and restores it to the same size as the input image. While preserving the spatial information of the original input image, a prediction is generated for each pixel. Finally, each pixel will be classified on the feature map of the upper sampling. As shown in Figure 8, an urban street image taken by a drone at a low altitude is randomly selected. After the semantic segmentation network processes the image, the foreground object vehicle is separated and marked as gray, and the static background in the image is marked as black. A relatively perfect separation of foreground and background is achieved.

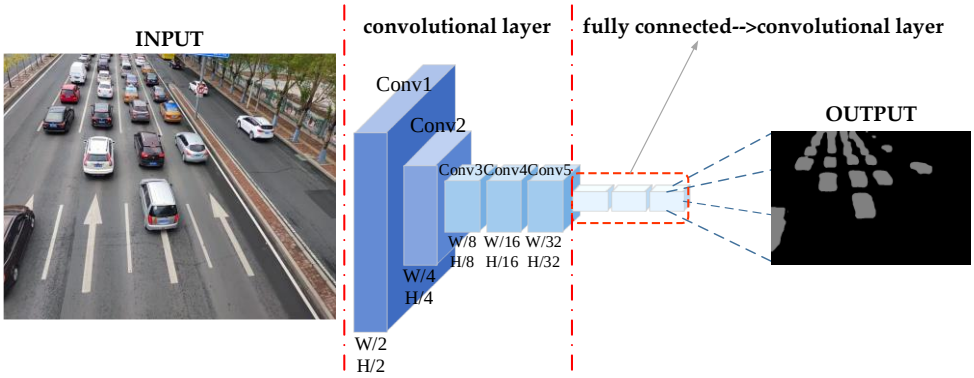


Figure 8. FCN segmentation image process diagram.

The feature map of the last layer is often too small, and many details are lost. If only the feature map of the previous layer is upsampled, the segmentation effect is not ideal. Therefore, the skip structure [30] is one of the characteristics of FCN. Through this structure, the predictions of the last layer and the shallower layers can be combined, and global forecasts and local predictions can be performed simultaneously. As shown in Figure 9 below, FCN-32s is a segmented image obtained directly after five convolution pooling and then through the several successive layers of convolution. At the same time, FCN-16s fuse the information of the underlying feature map based on FCN-32s, which can restore some details. The difference between FCN-8s and FCN-16s is that the underlying feature map information is fused twice. If the images of the three deconvolution results are combined, the accuracy can be improved to a certain extent.

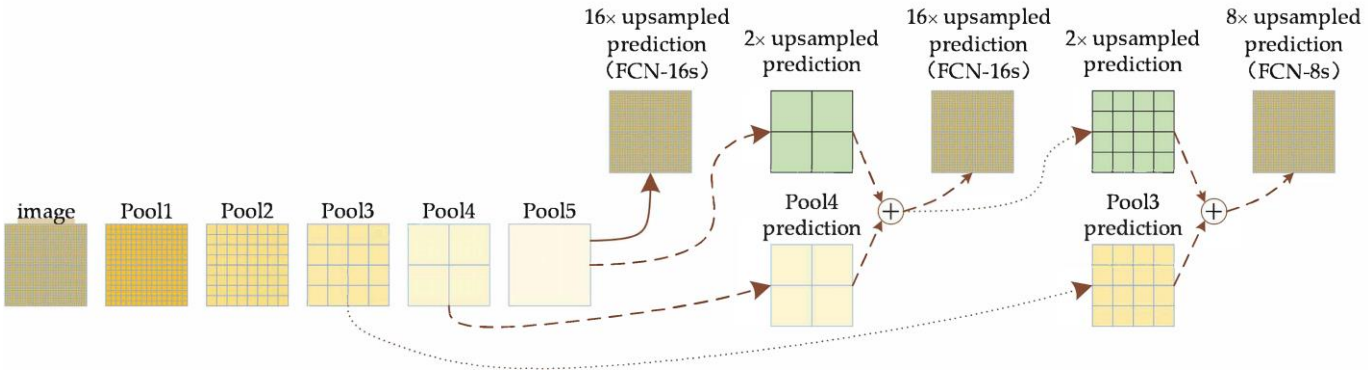


Figure 9. Jump structure.

2.3. Eliminate Foreground Feature Points and Delete False Matching Pairs

There are usually architectural backgrounds, fast-moving vehicles, ships, and other objects in the low-altitude aerial images of drones. While the architectural environment is static, things such as cars and boats in the foreground are often moving. If the image is not subjected to the foreground and background separation processing, the feature points are directly extracted and matched, and the homography matrix is used for splicing. Due to the motion of the foreground object, false matching pairs will appear, resulting in tearing and splicing misalignments in the spliced image. Therefore, in this paper, the feature points existing in the dynamic foreground are deleted to reduce the occurrence of mismatching to a certain extent, and then the mismatching pair is deleted.

This paper uses the RANSAC algorithm [31] to remove false matching pairs. When the error rate of matching pairs exceeds 50%, the algorithm can terminate false matching pairs well, and the robustness is ideal. RANSAC is essentially an iterative process. First, four matching point pairs are randomly selected, in the obtained matching points, and the other pair are outer points. Then, the homography matrix is calculated according to the four pairs of inner points, tested other matching points according to this matrix, and set the threshold. If it exceeds the threshold, it is considered an outer point. If it is less than the threshold, it is regarded as an inner point, and the value of the inner point is recorded. Statistically update all interior points, and on this basis, repeat the above process, iterate M times, and finally obtain the model with the most interior points. At this time, the model is the optimal model. In this paper, the feature point pairs that cannot be matched correctly are used as outliers, the interior points are used to update the initial assumed model, and the initial number of iterations is set to 10,000 times. Finally, a model that meets the expectations is obtained, and the mismatches that do not fit the model are eliminated.

2.4. Image Fusion and Stitching

After completing the image registration, it is necessary to stitch the two images together to output a complete image [32]. In this paper, based on the best matching feature point pair obtained by image registration, the homography matrix is calculated, and the corresponding relationship and spatial transformation model of the overlapping parts of the images to be spliced are determined. After calculating the size of the entire image, the two images are geometrically transformed and stitched together. However, there will be a noticeable gap after splicing, so image fusion is essential. Through image fusion processing, the crack of splicing can be eliminated. In this paper, the stitched images are fused by the weighted average method, and the stitching of the aerial images is completed.

In the fusion process, if I_1 and I_2 represent the two images to be spliced, and I represent the spliced image, the image fused by this method can be described as follows:

$$\begin{cases} I(x,y) = I_1(x,y) & (x,y) \in I_1 \\ I(x,y) = w_1 I_1(x,y) + w_2 I_2(x,y) & ,(x,y) \in (I_1 \cap I_2) \\ I(x,y) = I_2(x,y) & (x,y) \in I_2 \end{cases} \tag{10}$$

In the formula, w_1 and w_2 represent the weight of the overlapping area, and $0 < w_1$, $w_2 < 1$, they satisfy the following relationship with the width of the overlapping area:

$$\begin{cases} w_1 = \frac{1}{width} \\ w_2 = 1 - w_1 \end{cases} \tag{11}$$

In the overlapping area, w_1 gradually changes from 1 to 0 and w_2 from 0 to 1 so that the overlapping area I_1 can be smoothly transitioned to I_2 during the image fusion process. The images stitched by this method can effectively eliminate stitching gaps and achieve seamless stitching.

3. Result

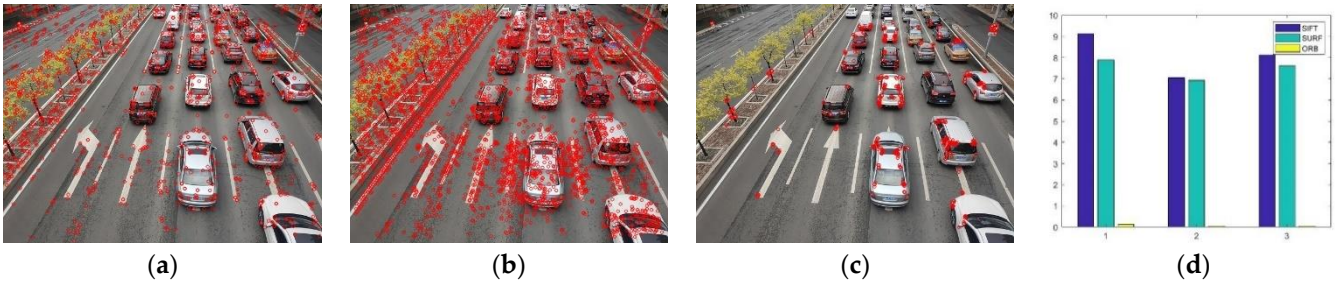
To verify the superiority of the algorithm in this paper, we used authentic images obtained by the DJI MATRICE200 V2 UAV flying at low altitudes and speeds. The drone is equipped with a DJI Zenmuse X5S gimbal camera. Its most significant feature is that it supports eight standard M4/3 lenses, including zoom lenses, and the focal length covers 9–45 mm. It features 20.8 million pixels with DNG infinite continuous shooting technology at 20 frames per second. The inspiration for the algorithm in this paper comes from solving the engineering application problems of road network traffic, so all the images in our research paper experiments are obtained from surveying and mapping road streets in Harbin. The following explains why we choose the ORB algorithm and shows the effect of separating foreground and background, removing foreground feature points, and stitching.

3.1. Time-Consuming Comparison of SIFT, SURF and ORB

At present, the popular feature point extraction methods mainly include SIFT, SURF and ORB. SIFT feature is a local feature of image. It has good invariance to translation, rotation, scaling, brightness change, occlusion and noise, etc. It also maintains a certain degree of stability to visual change and affine transformation. The biggest disadvantage of SIFT is that it is difficult to achieve real-time without the help of hardware or special image processors. SURF is a feature point detection and description algorithm similar to SIFT. The advantages of the SURF algorithm are that it is much faster than SIFT and has good stability. SURF has good robustness, and the feature point recognition rate is higher than SIFT. In the case of angle of view, illumination and scale change, SURF is generally better than SIFT. The ORB feature combines the detection method of FAST feature points with Brief feature descriptors and improves and optimizes the original feature points. SIFT features have high accuracy, SURF invariance is strong, and ORB features are fast. SIFT and SURF are superior to the ORB algorithm in the accuracy and invariants of extracting feature points, but the ORB algorithm is far superior to the SIFT and SURF algorithms in speed [33]. The following figure shows the experimental results of real-time testing of the three algorithms.

We randomly selected three aerial images, used SIFT, SURF, and ORB to extract feature points, respectively, and compared the time taken. Figure 10 shows the extraction effect of one of the images and the histogram of the time-consuming comparison of the three images extracted by three algorithms respectively. In Figure 10a–c are renderings of feature points extracted by SIFT, SURF and ORB. SIFT extracted feature points with 128 dimensions, SURF with 64 dimensions, and ORB with 32 dimensions. As can be seen from the figure, SIFT has the best effect although it extracts the least feature points. SURF extracted the most feature points. ORB extracted the feature points more than SIFT, less than SURF, but has the problem of uneven distribution of sparse. Therefore, this paper introduces the quadtree idea to improve the classical ORB algorithm to improve this problem. The relevant content has been described in detail previously. However, in terms of speed, the ORB algorithm has a big advantage. As shown in Figure 10d, blue means SIFT time-consuming, green means SURF time-consuming, and yellow means ORB time-consuming. It can be seen from the experimental results that the ORB extraction speed is much better than the other two algorithms. Due to a large amount of UAV aerial image data, the time-consuming problem

must be considered when processing aerial images. Therefore, the ORB algorithm is more suitable for extracting feature points of aerial images than the other two algorithms.



(**a**) (**b**) (**c**) (**d**)

Figure 10. SIFT, SURF, ORB extraction feature point renderings and time-consuming comparison: (**a**) is the effect of SIFT extraction of feature points; (**b**) is the effect of SURF extraction of feature points; (**c**) is the effect of ORB extraction of feature points; (**d**) is a time-consuming comparison histogram of the three algorithms for extracting feature points.

3.2. Separate Foreground and Background

We randomly selected a city street image taken by a drone at a low altitude and performed segmentation experiments on the PyCharm Community Edition 2021.2.2 platform. There are various moving vehicles in the city streets, we call it the foreground; in contrast, we call the static road and the buildings on both sides of the road the background. After semantic segmentation of the image, adjust the RGB value to change the black background to a white background. Then, use the additional function in the OpenCV library to add the segmented foreground image to the original image. In the end, a relatively clear and complete dynamic foreground image was obtained. The results are as follows: Figure 11a is the street image taken by the drone at a low altitude, after semantic segmentation processing, as shown in Figure 11b, the moving vehicle and the static background are segmented, Figure 11c is the final extracted foreground vehicle image.

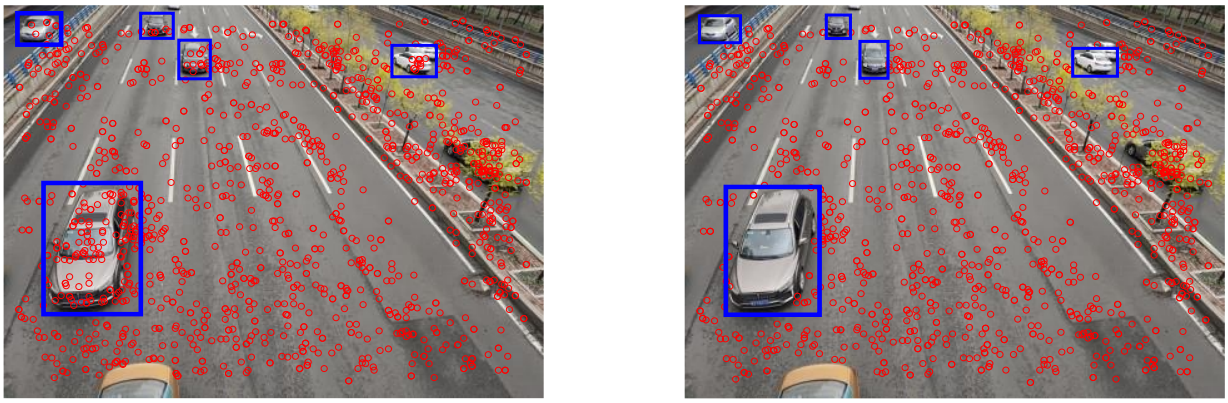
(**a**) (**b**) (**c**)

Figure 11. Semantic segmentation extraction foreground: (**a**) is a low-altitude aerial image; (**b**) is the foreground and background separation image after semantic segmentation; (**c**) is the final extracted foreground image.

3.3. Eliminate Foreground Feature Points

In this paper, the ORB algorithm based on quadtree is used to extract the feature points of the image to be stitched, and the coordinates of the detected feature points are output and saved. While extracting feature points, semantic segmentation is performed on the image to be spliced, and after a clear and complete foreground image is extracted, the pixel coordinates of the foreground image are output. Then, the coordinates of the foreground pixel points are compared with the coordinates of the feature points of the image. If the feature point is located in the foreground, it will be deleted, and if it is not in

the foreground, it will be retained. This paper randomly selects a low-altitude aerial photo for the experiment, after processing, as shown in the experimental results in Figure 12. Figure 12a shows the effect of extracting feature points through the orb algorithm based on the idea of the quadtree. As seen in Figure 12b, the feature points on the moving vehicles on the city streets in the image have been eliminated, and do not affect the position of the feature points on the static background. It can be seen from the experimental results that the removal effect is perfect.



(**a**) (**b**)

Figure 12. The effect of removing foreground feature points: (**a**) is the rendering of feature points extracted by ORB based on quadtree; (**b**) is the rendering after removing the feature points on the foreground of the vehicle.

3.4. Stitching Result

We selected three groups of low-altitude aerial photography images of UAVs and conducted experimental tests on the Visual Studio 2017 platform. We experimented with two algorithms. The first algorithm uses the ORB algorithm based on the quadtree idea to extract feature points and then perform image stitching experiments. The second algorithm uses the algorithm proposed in this paper to conduct experiments, perform semantic segmentation processing on the image, extract feature points based on the ORB algorithm based on the quadtree idea, delete the feature points of the dynamic foreground, and then perform image stitch stitching. The splicing results of the two algorithms are compared and analyzed from two aspects: subjective evaluation and objective evaluation.

3.4.1. Visual Subjective Evaluation

Figure 13 is the first group of experimental pictures and results in comparison charts. Figure 13a,d is the image to be stitched, and we can see the difference between the two images. There are clearly more trees on the right-hand side of the road in diagram (a) than in diagram (d), while the field of view on the left-hand side of diagram (a) is significantly less than on the left-hand side of diagram (d). Both the stitched image of the first algorithm and the stitched image of the algorithm in this paper contains all the details of the original image. Figure 13b is the splicing result of the first algorithm. It can be clearly seen that the indication line on the road is obviously dislocated, and the vehicle above the image has collided. There is also a dislocation phenomenon in the spliced road. Figure 13e is the splicing result of the algorithm in this paper. The indicator line, the vehicle behind, and the side of the road are all spliced perfectly, and there is no dislocation phenomenon. Figure 13b,e both contain the details of the images to be stitched, but the stitching quality of Figure 13e is significantly better than that of Figure 13b.

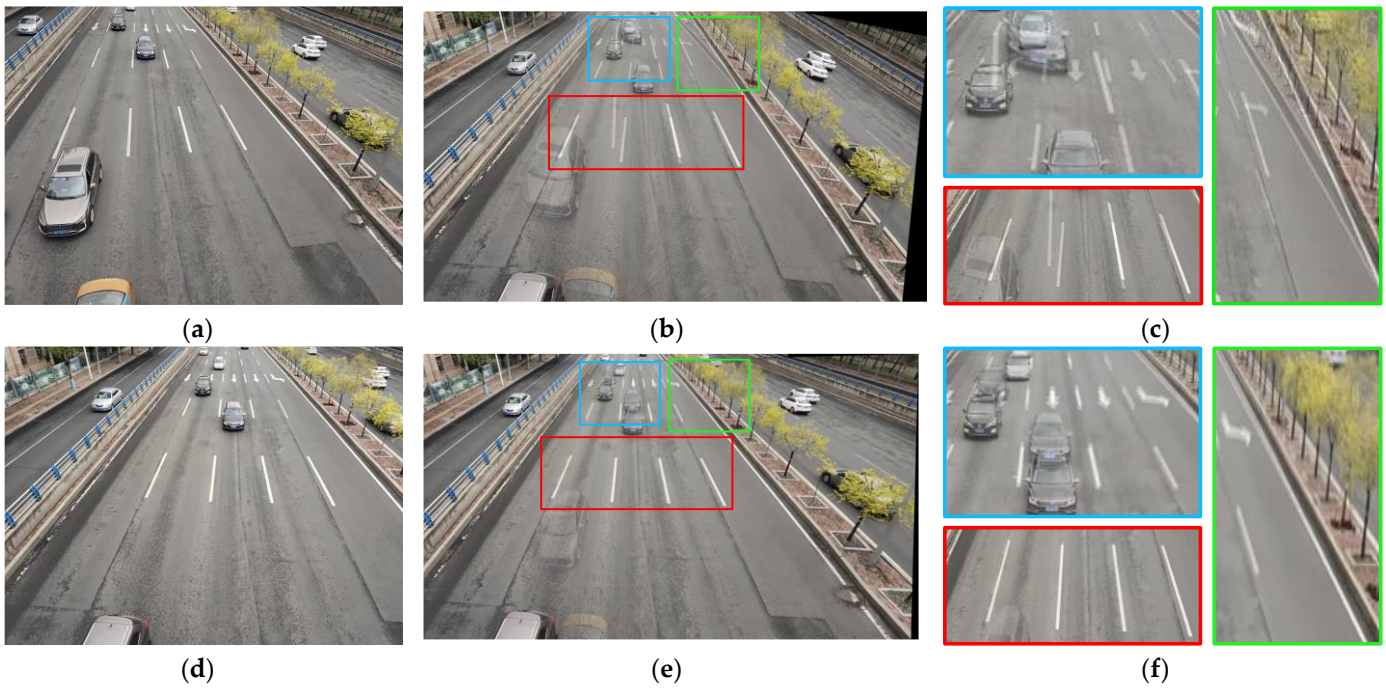


Figure 13. The first set of experimental pictures and results in the comparison chart: (**a**,**d**) are the first group of images to be stitched, (**b**) is the first algorithm stitching effect, (**e**) is the stitching effect of the algorithm in this paper. The comparison of (**b**,**e**) highlights the superiority of the algorithm in this paper. Boxes of the same color indicate the corresponding identical areas in (**b**,**e**). Figure (**c**,**f**) shows a detailed demonstration of the misalignment phenomenon in Figure (**b**,**e**) respectively.

Figure 14 is the second group of low-altitude aerial images to be spliced and the comparison of the results. Figure 14a,d are images to be spliced, and it can be seen that the fields of view of the two images are different. Experiments were carried out based on the above two methods, and the stitching results did not lose the details of the original images. Figure 14b is the splicing result of the first method, and the influence of the feature points on the dynamic vehicle is not considered. It can be seen that the splicing effect is not ideal, the road indication line is seriously misaligned, and there is also a situation where the vehicle splicing is incomplete. Figure 14e is the effect of splicing based on the algorithm in this paper. There is no misalignment of the background highway indicator line, and the vehicle is not spliced completely. From the comparison of the stitching results of the two groups, it can be seen that the background stitching effect of Figure 14e is better than Figure 14b. The algorithm in this paper greatly reduces the misalignment of background stitching caused by moving vehicles.

Figure 15 is a comparison of the third set of experimental images and splicing results collected by UAV low-altitude flight. Figure 15a,d are images to be spliced. Comparing the details of the two images, it can be seen that the field of view is different. We stitch the third set of images using the two algorithms mentioned previously. Figure 15b is obtained by using the first stitching algorithm. Since the influence of dynamic vehicles is not considered, the static background in the stitched image has a lot of misalignments, such as the misalignment of the traffic indication lines marked in the figure. Figure 15e is the result obtained by splicing the algorithm proposed in this paper. It can be seen that the splicing map of the algorithm in this paper does not appear in the misalignment of the traffic line. The comparison shows that the stitching effect of Figure 15e is much better than that of Figure 15b. The algorithm in this paper reduces the phenomenon of static background stitching dislocation.

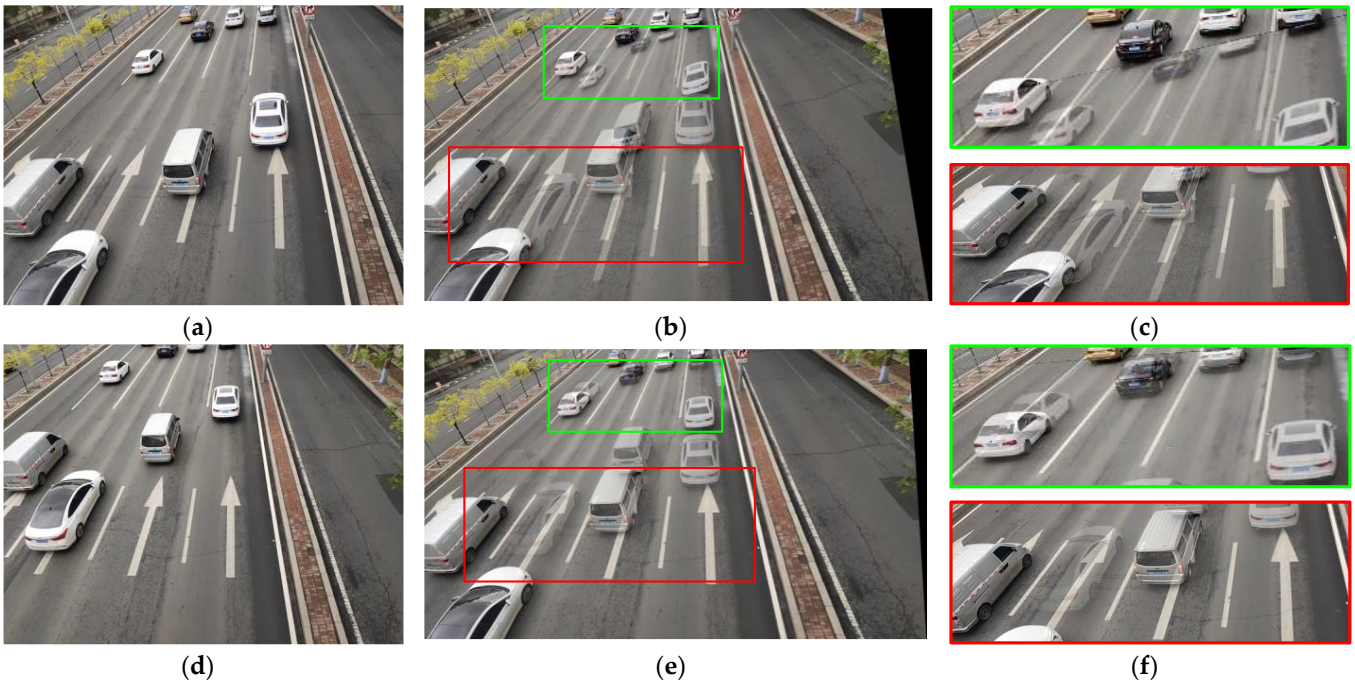


Figure 14. The second group of splicing renderings: (**a**,**d**) are the second group of images to be stitched; (**b**) is the first algorithm stitching effect; (**e**) is the stitching effect of the algorithm in this paper. The comparison of (**b**,**e**) highlights the superiority of the algorithm in this paper. Boxes of the same color indicate the corresponding identical areas in (**b**,**e**). Figure (**c**,**f**) shows a detailed demonstration of the misalignment phenomenon in Figure (**b**,**e**) respectively.

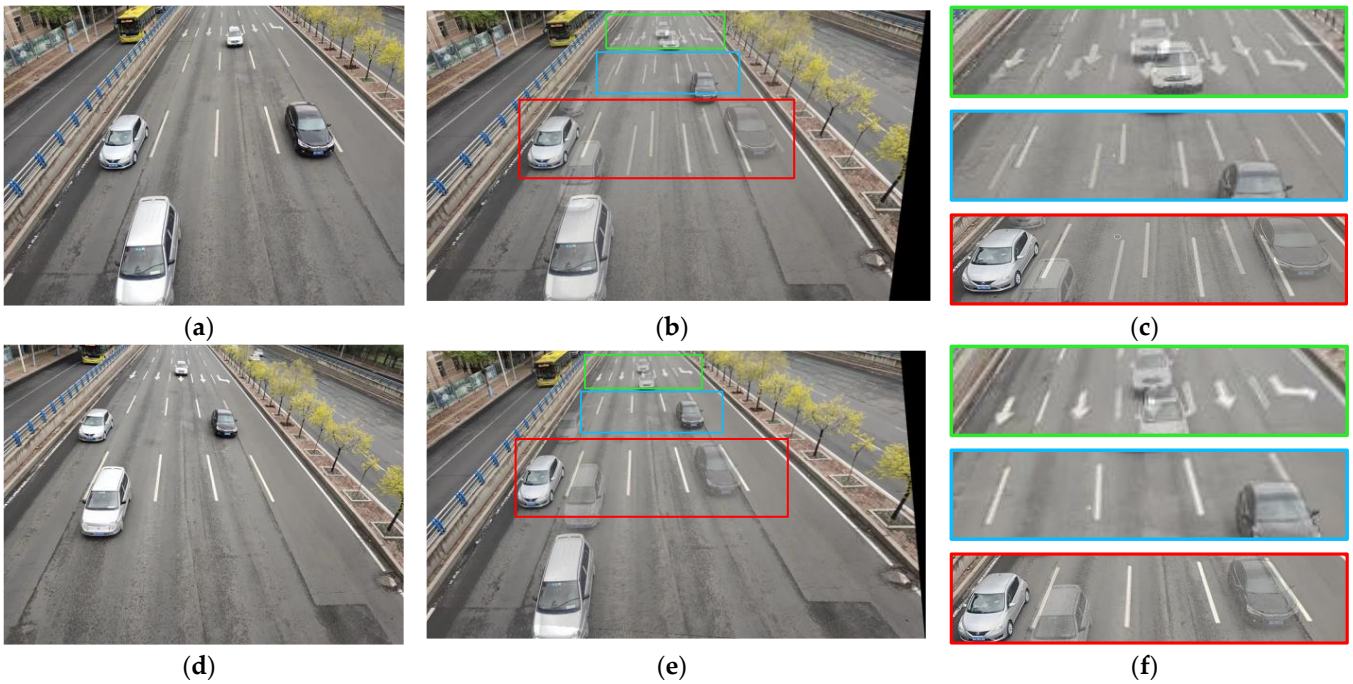


Figure 15. The third group of splicing renderings: (**a**,**d**) are the third group of images to be stitched; (**b**) is the first algorithm stitching effect; (**e**) is the stitching effect of the algorithm in this paper. The comparison of (**b**,**e**) highlights the superiority of the algorithm in this paper. Boxes of the same color indicate the corresponding identical areas in (**b**,**e**). Figure (**c**,**f**) shows a detailed demonstration of the misalignment phenomenon in Figure (**b**,**e**) respectively.

3.4.2. Quantitative and Objective Evaluation of Fusion Effect

To more accurately measure the splicing quality of the above experimental results, this paper selects information entropy, average gradient, peak signal-to-noise ratio, and root mean square error as objective evaluation indicators.

Information entropy (E) is mainly used to measure the information content of the image. The larger the measurement value, the more information the image has. The calculation formula is as follows:

$$H(x) = -\sum_{i=1}^{n} p(x_i) \log(p(x_i)) \tag{12}$$

where n is the gray level of the image, and $p(x_i)$ is the distribution density of x_i.

The average gradient (AG) reflects the sharpness of the image. The larger its value, the more precise the image. The average gradient (AG) can be calculated by Equation (13):

$$G = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\frac{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}{2}} \tag{13}$$

where $M \times N$ represents the size of the image, $\frac{\partial f}{\partial x}$ represents the horizontal gradient, and $\frac{\partial f}{\partial y}$ represents the vertical gradient.

The peak signal-to-noise ratio (PSNR) is used to measure the adequate information and noise ratio in the image. It can quantitatively express the distortion degree of the stitched image. The larger the value, the smaller the noise of the image. The peak signal-to-noise ratio (PSNR) is calculated by Equation (15):

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (I_H(i,j) - I_W(i,j))^2 \tag{14}$$

$$PSNR = 10 \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right) \tag{15}$$

where $2^n - 1$ is the maximum pixel value of the input image.

The root mean square error (RMSE) is mainly used to calculate the difference between the stitched image and the original image, and the smaller the value, the better the stitching effect. The calculation formula is given by Equation (16):

$$RMSE = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (I_H(i,j) - I_W(i,j))^2} \tag{16}$$

where I_H (i, j) is the gray value of pixels located in row i and column j of the spliced image, I_W (i, j) is the gray value of pixels located in row i and column j of the original image.

Table 1 shows the evaluation index values corresponding to the three groups of experimental results, where (**i**) represents the first algorithm, and (**ii**) represents the algorithm in this paper. Looking at the data in the table, in the four objective evaluation indicators of information entropy, average gradient, peak signal-to-noise ratio, and root mean square error, it can be seen that the index values of the algorithm in this paper are generally better than the index values of the first algorithm. It reflects the superiority of the algorithm in this paper. Therefore, from the visual subjective evaluation and quantitative objective evaluation, it can be seen that the algorithm in this paper well preserves the details of the original image and can effectively reduce the splicing dislocation and tearing phenomena in the background splicing caused by dynamic foreground objects.

Table 1. Objective evaluation index comparison.

Experimental Group Number	Algorithm	E	AG	PSNR (dB)	RMSE
Group 1	**(i)**	7.0670	12.6791	14.5035	48.0140
	(ii)	7.1066	13.3311	15.1251	44.1360
Group 2	**(i)**	7.2703	10.7097	12.9428	56.4684
	(ii)	7.3453	10.8163	13.2839	54.2050
Group 3	**(i)**	7.1369	12.9058	13.4203	54.3907
	(ii)	7.1557	12.9119	13.8843	51.5612

3.5. Comparison of Stitching Results

In order to verify the superiority of the algorithm in this paper, we use the first algorithm mentioned above, the APAP algorithm, the AANAP algorithm, and the algorithm in this paper to stitch the experimental images.

The original experimental image and the comparison of experimental results are shown in Figure 16. Figure 16a is the image to be stitched. Figure 16b is the result obtained by using the first algorithm mentioned above. Since the influence of the dynamic foreground is not considered, in the splicing result, there is a problem of dislocation and tearing of the traffic line in the static background. Figure 16c is the splicing effect of the AANAP algorithm. There are obvious splicing gaps in the figure, and as shown in the green indicator box, the splicing effect of the traffic indicator lines is not perfect. Figure 16d is the result obtained by APAP algorithm splicing, and there are also obvious splicing traces. The splicing effect of the traffic lines in the background is better than that of AANAP, but there is still room for improvement. Figure 16e is the effect image obtained by using the splicing algorithm in this paper. There is no splicing gap in the figure, the splicing of the traffic indicator lines in the static background is also very good, and there is no dislocation tearing phenomenon. Comparing the above four splicing results, all four algorithms can retain the rich details of the original image, but after the image is spliced by the previous three algorithms, there are obvious splicing gaps, splicing dislocation and tearing and other problems. The algorithm in this paper avoids the influence of dynamic foreground on static background stitching. While ensuring the rich details of the original image, there are no stitching gaps and no static background stitching dislocation phenomenon, which improves the stitching quality of the image. The AANAP and APAP algorithms are traditional stitching algorithms. They can meet the basic requirements for the processing of static aerial images at medium and high altitudes, but there are obvious shortcomings in processing images taken at low altitudes in complex environments containing moving objects. The algorithm in this paper takes into account the effects of dynamic foregrounds and proposes a solution. The effect of dynamic foregrounds on the stitching of static backgrounds is avoided. While ensuring the rich details of the original image, there are no stitching gaps and no static background stitching misalignment, which fully demonstrates that the algorithm of this paper has greatly improved the stitching quality of the image.

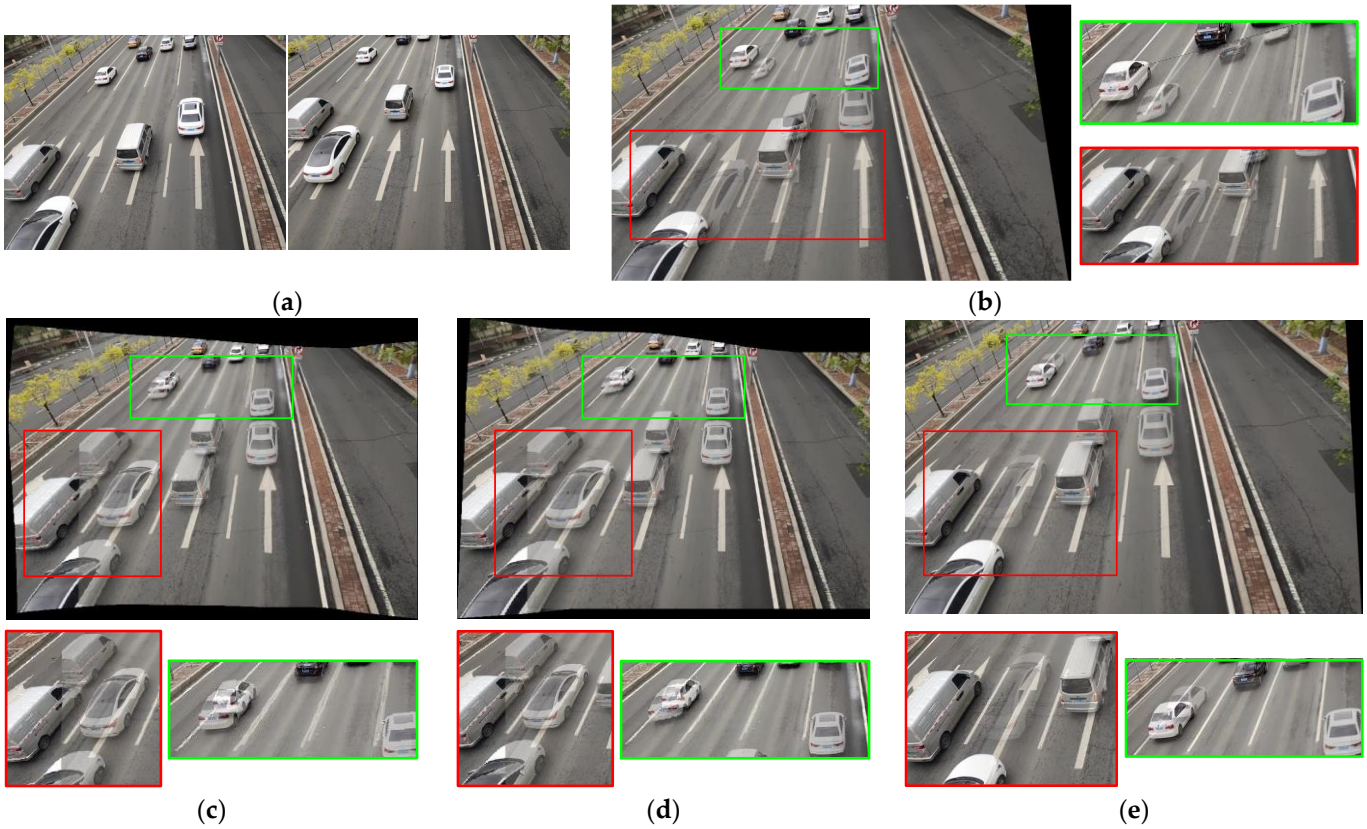


Figure 16. This image shows the splicing effect diagram of the first algorithm described in the text, the APAP algorithm, the AANAP algorithm and the algorithm in this paper: (**a**) is the experimental image to be stitched; (**b**) is the stitched image of the first algorithm; (**c**) is the splicing effect of the AANAP algorithm; (**d**) is the splicing effect diagram of the APAP algorithm; (**e**) is the stitching effect of the algorithm in this paper. Comparing the splicing renderings, it shows the superiority of the algorithm in this paper. The red box highlights the splice marks and the green box highlights the misplaced splices.

4. Discussion

The inspiration for the algorithm in this paper comes from practical engineering applications, mainly for the problems of traffic road network monitoring and urban traffic emergency management. Therefore, using drones to fly at low altitudes, equipped with cameras to take pictures of road conditions, and obtain real-time panoramic images after stitching processing, will be conducive to real-time supervision of urban traffic. Most traditional stitching algorithms are suitable for hollow or high-altitude remote sensing images. However, there is a big difference between hollow or high-altitude remote sensing images and low-altitude aerial photography images. Most hollow or high-altitude remote sensing images depict the general outline of the image and lack of details. As the vertical gradient decreases, the field of view of aerial images taken by drones becomes smaller, and the detailed information contained in aerial images becomes more and more abundant. The use of UAV low-altitude aerial photography to solve practical engineering problems will have a great application prospect. Using drones to supervise urban traffic problems belongs to the field of low-altitude remote sensing, so the algorithm in this paper is mainly aimed at the problem of image stitching in low-altitude environments.

Image registration is a crucial step in image stitching. Experiments have proved that when processing low-altitude aerial images, ORB feature extraction has obvious advantages of saving time compared with SIFT and SURF algorithms and can meet real-time performance requirements. It is fast and convenient for drones to obtain images. Still, the number of pictures is also large, and the processing time of the pictures must be

as short as possible, so the ORB algorithm is selected for feature extraction. The feature points extracted by the classical ORB algorithm meet the requirements in terms of quantity and time, but there are sparse and uneven phenomena. Therefore, the idea of quadtree is introduced in this paper, and the extraction points of ORB are divided and screened, so that the feature points can be evenly distributed in the image, and the drawbacks of the classical ORB algorithm are removed.

Compared with hollow and high-altitude remote sensing imagery, low-altitude remote sensing images are more detailed. The experimental results show that moving vehicles can cause stitching dislocation and tearing problems when stitching static backgrounds, such as traffic lines. To increase the number of matching pairs, Zheng et al. [34] have performed foreground and background separation experiments based on Grab Cut for indoor scene pictures. Still, the effect is completely different from this article. Different scenes have different definitions for foreground and background. For low-altitude aerial images, we define moving objects as foreground and static objects as background. In separation experiments based on semantic segmentation, dynamic vehicles are perfectly separated. According to the semantic information of the vehicle and the coordinate information on the image, all the feature points distributed on the vehicle are eliminated. From the comparison of (b) and (e), (c) and (f) in Figures 13–15, it can be seen that the algorithm in this paper has fully examined the dynamic factors of the foreground, removed the phenomenon of static background dislocation, and improved the stitching quality. The proposed algorithm greatly reduces the problem of background stitching caused by dynamic foreground.

The current data set of the algorithm in this paper is relatively simple, and the season, climate, and other factors are not fully considered. Due to the need for practical applications, to obtain as much picture detail information as possible, our UAV collects pictures at low altitudes and with a certain speed limit, which results in the difference between the pictures being not particularly large. At present, our experiments are mainly carried out on two images. Considering practical engineering problems, improving the quality of image stitching with large differences, and stitching multiple consecutive aerial images is one of the follow-up research directions.

5. Conclusions

The shooting environment of the drone and the height angle of the shooting are complex and changeable. Aerial drone photography covers a large field of view and can clearly depict the exact outline of things on the ground. However, in medium- to high-altitude environments, the aerial images contain a lack of detailed information about specific things. As the drone is lowered on the vertical gradient, the field of view becomes smaller and smaller, and the aerial images contain more and more detailed information about specific things on the ground, with dynamic objects clearly visible. In hollow and high-altitude environments, the details of the objects in the aerial image are not obvious. In contrast, in low-altitude settings, the dynamic objects in the image taken by the drone are visible. Therefore, the stitching of low-altitude aerial images will be affected by changes in the position of dynamic foreground objects, resulting in quality problems, such as stitching dislocation and tearing in the stitched image. Traditional algorithms cannot effectively solve this problem. A UAV aerial image stitching algorithm is proposed to solve this problem based on semantic segmentation and ORB. It combines the semantic information obtained by semantic segmentation with the traditional ORB stitching algorithm to achieve high-quality aerial image stitching.

This paper uses aerial images to perform the speed comparison test of SIFT, SURF, and ORB extraction features. The advantages and shortcomings of the three algorithms for extracting feature points are analyzed. Because the number and speed of feature points extracted by the ORB algorithm meet the experimental requirements, the ORB algorithm is selected for feature extraction of aerial images. However, in order to solve the problem of uneven distribution of feature points extracted by ORB, this paper introduces the idea of quadtree to filter the extracted feature points, so that the feature points are

evenly distributed in the image. Then, to reduce the impact of dynamic moving objects on static background splicing, during image registration, semantic segmentation is performed on the images to be spliced, the foreground and background are separated, the semantic information of the foreground is obtained and compared with the feature point information, and the feature points located on the dynamic foreground were removed. Finally, the image splicing and fusion are realized based on the homography matrix and the weighted fusion algorithm.

This paper conducts stitching experiments based on two algorithms. The first one does not consider dynamic foreground factors to perform image stitching directly, and the second one is based on the algorithm of this paper for stitching experiments. According to the analysis of the experimental comparison results, both from the subjective visual evaluation and the objective data, the superiority of the algorithm in this paper is strongly proved. In addition to this, experiments have been carried out with conventional stitching algorithms based on the experimental data in this paper, and the stitching results have been compared with those of this paper. It is well demonstrated that the algorithm in this paper effectively reduces the problems of background image stitching caused by dynamic foregrounds, such as background stitching dislocation, and improves the quality of image stitching. It provides technical support for the broad application of images captured by UAVs at various heights or in complex environments.

Author Contributions: Methodology, software, writing—original draft preparation and writing—review and editing, G.Z.; formal analysis, G.Z. and D.Q.; resources and project administration, D.Q.; supervision, D.Q., L.M., J.Y., H.T., M.Y. and H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Outstanding Youth Project of Provincial Natural Science Foundation of China in 2020 (YQ2020F012), National Natural Science Foundation of China (61771186) and Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (UNPYSCT2017125).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: We thank the reviewers for the thorough review and greatly appreciate the comments and suggestions, which significantly contributed to improving the quality of the publication.

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

References

1. Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In *Alvey Vision Conference*; The Plessey Company Plc: Manor, UK, 1988.
2. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
3. Bay, H.; Tuytelaars, T.; Gool, L.V. SURF: Speeded up robust features. In Proceedings of the 9th European conference on Computer Vision-Volume Part I, Graz, Austria, 7–13 May 2006.
4. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G.R. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 6–13.
5. Yeh, C.C.; Chang, Y.L.; Alkhaleefah, M.; Hsu, P.H.; Eng, W.; Koo, V.C.; Huang, B.R.; Chang, L.A. YOLOv3-Based Matching Approach for Roof Region Detection from Drone Images. *Remote Sens.* **2021**, *13*, 127. [CrossRef]
6. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [CrossRef]
7. Mur-Artal, R.; Tardós, J. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [CrossRef]
8. Li, M.; Chen, R.; Zhang, W.; Li, D.; Liao, X.; Wang, L.; Pan, Y.; Zhang, P. A Stereo Dual-Channel Dynamic Programming Algorithm for UAV Image Stitching. *Sensors* **2017**, *17*, 2060. [CrossRef] [PubMed]
9. Zhang, W.; Li, X.; Yu, J.; Kumar, M.; Mao, Y. Remote sensing image mosaic technology based on SURF algorithm in agriculture. *EURASIP J. Image Video Process.* **2018**, *2018*, 85. [CrossRef]
10. Zhao, J.; Zhang, X.; Gao, C.; Qiu, X.; Cao, W. Rapid Mosaicking of Unmanned Aerial Vehicle (UAV) Images for Crop Growth Monitoring Using the SIFT Algorithm. *Remote Sens.* **2019**, *11*, 1226. [CrossRef]

11. Xie, R.; Tu, J.; Yao, J.; Xia, M.; Li, S. A robust projection plane selection strategy for UAV image stitching. *Int. J. Remote Sens.* **2018**, *40*, 2019. [CrossRef]
12. Xu, Q.; Chen, J.; Luo, L.B.; Gong, W.P.; Wang, Y. UAV Image Mosaicking Based on Multiregion Guided Local Projection Deformation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3844–3855. [CrossRef]
13. Pham, N.T.; Park, S.; Park, C.S. Fast and Efficient Method for Large-Scale Aerial Image Stitching. *IEEE Access* **2021**, *9*, 127852–127865. [CrossRef]
14. Yuan, Y.T.; Fang, F.M.; Zhang, G.X. Superpixel-Based Seamless Image Stitching for UAV Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1565–1576. [CrossRef]
15. Xue, W.L.; Zhang, Z.; Chen, S.Y. Ghost Elimination via Multi-Component Collaboration for Unmanned Aerial Vehicle Remote Sensing Image Stitching. *Remote Sens.* **2021**, *13*, 1388. [CrossRef]
16. Chen, J.; Li, Z.X.; Peng, C.L.; Wang, Y.; Gong, W.P. UAV Image Stitching Based on Optimal Seam and Half-Projective Warp. *Remote Sens.* **2022**, *14*, 1068. [CrossRef]
17. Gao, J.; Kim, S.J.; Brown, M.S. Constructing image panoramas using dual-homography warping. In *CVPR 2011 IEEE*; IEEE: Piscataway, NJ, USA, 2011.
18. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
19. Pearlmutter, B.A. Gradient calculations for dynamic recurrent neural networks: A survey. *IEEE Trans. Neural Netw.* **1995**, *6*, 1212–1228. [CrossRef]
20. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2014**, *63*, 139–144. [CrossRef]
21. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:abs/1702.00307.
22. Szegedy, C.; Liu, W.; Jia, Y. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; 2016; pp. 770–778.
24. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]
25. Mesbah, R.; McCane, B.; Mills, S. Deep convolutional encoder-decoder for myelin and axon segmentation. In Proceedings of the 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ), Palmerston North, New Zealand, 21–22 November 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
26. Yasrab, R.; Gu, N.; Zhang, X. SCNet: A simplified encoder-decoder CNN for semantic segmentation. In Proceedings of the 2016 5th International Conference on Computer Science and Network Technology (ICCSNT), Changchun, China, 10–11 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 785–789.
27. Emeršič, Ž.; Gabriel, L.L.; Štruc, V.; Peer, P. Pixel-Wise Ear Detection with Convolutional Encoder-Decoder Networks. *arXiv* **2017**, arXiv:1409.1556.
28. Rosten, E. Machine learning for very high-speed corner detection. In *ECCV*; Springer: Berlin, Germany, 2006.
29. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. BRIEF: Binary Robust Independent Elementary Features. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2010.
30. Shi, Z.; Li, H.; Cao, Q.; Ren, H.; Fan, B. An Image Mosaic Method Based on Convolutional Neural Network Semantic Features Extraction. *J. Signal Process. Syst.* **2020**, *92*, 435–444. [CrossRef]
31. Barath, D.; Matas, J. Graph-Cut RANSAC. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Piscataway, NJ, USA, 2018.
32. Sun, Y.; Lv, Y.; Song, B.; Guo, Y.; Zhou, L. Image Stitching Method of Aerial Image Based on Feature Matching and Iterative Optimization. In Proceedings of the 40th Chinese Control Conference (CCC), Shanghai, China, 26–28 July 2021; IEEE: Piscataway, NJ, USA, 2021.
33. Wei, L.; Li, Q. Comparative Analysis of SIFT and SURF and ORB Algorithms Based on OpenC V Environment. *Control. Instrum. Chem. Ind.* **2018**, *45*, 714–716.
34. Zheng, P.; Qin, D.; Han, B.; Ma, L.; Berhane, T.M. Research on Feature Extraction Method of Indoor Visual Positioning Image Based on Area Division of Foreground and Background. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 402. [CrossRef]