

Article

# Large Aerial Image Tie Point Matching in Real and Difficult Survey Areas via Deep Learning Method

Xiuli Yuan , Xiuxiao Yuan \*, Jun Chen  and Xunping Wang

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

\* Correspondence: yuanxx@whu.edu.cn

**Abstract:** Image tie point matching is an essential task in real aerial photogrammetry, especially for model tie points. In current photogrammetry production, SIFT is still the main matching algorithm because of the high robustness for most aerial image tie points matching. However, when there is a certain number of weak texture images in a surveying area (mountain, grassland, woodland, etc.), these models often lack tie points, resulting in the failure of building an airline network. Some studies have shown that the image matching method based on deep learning is better than the SIFT method and other traditional methods to some extent (even for weak texture images). Unfortunately, these methods are often only used in small images, and they cannot be directly applied to large image tie point matching in real photogrammetry. Considering the actual photogrammetry needs and motivated by the Block-SIFT and SuperGlue, this paper proposes a SuperGlue-based LR-Superglue matching method for large aerial image tie points matching, which makes learned image matching possible in photogrammetry application and promotes the photogrammetry towards artificial intelligence. Experiments on real and difficult aerial surveying areas show that LR-Superglue obtains more model tie points in forward direction (on average, there are 60 more model points in each model) and more image tie points between airline (on average, there are 36 more model points in each adjacent images). Most importantly, the LR-Superglue method requires a certain number of points between each adjacent model, while the Block-SIFT method made a few models have no tie points. At the same time, the relative orientation accuracy of the image tie points matched by the proposed method is significantly better than block-SIFT, which reduced from 3.64  $\mu\text{m}$  to 2.85  $\mu\text{m}$  on average in each model (the camera pixel is 4.6  $\mu\text{m}$ ).

**Keywords:** image matching; model tie; deep learning; difficult survey area; special texture



**Citation:** Yuan, X.; Yuan, X.; Chen, J.; Wang, X. Large Aerial Image Tie Point Matching in Real and Difficult Survey Areas via Deep Learning Method. *Remote Sens.* **2022**, *14*, 3907. <https://doi.org/10.3390/rs14163907>

Academic Editors: Thomas Blaschke, Omid Rahmati and Omid Ghorbanzadeh

Received: 7 July 2022

Accepted: 8 August 2022

Published: 12 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

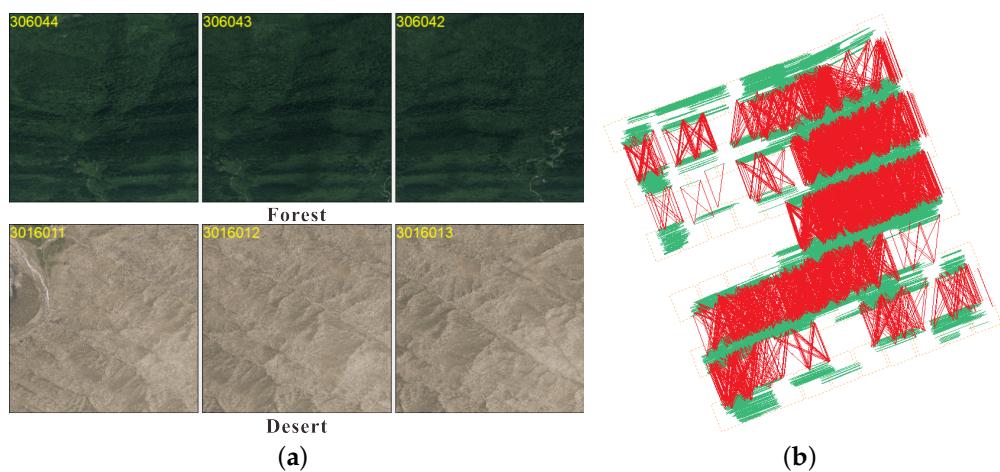
Images tie point matching plays a crucial role in the field of aerial photogrammetry, especially image point coordinates measurement, which has promoted the development of photogrammetry automation since the image correlation technology proposed, including NCC [1] (normalized correlation coefficient) and LSM [2] (least squares matching), etc. Afterwards, many corner point extraction methods, such as Moravec [3], Förstner [4] and Harris [5] were used to match aerial images for image tie points with NCC or LSM. Although these methods help with image tie matching, there are variants to transformation, scale and illumination changes. To extract local features and match them from two images taken from different viewpoints and under different illumination conditions, the SIFT [6] (scale-invariant feature transform) was first proposed. Then, numerous work, namely SURF [7] (speeded up robust features), FAST [8] (Machine Learning for High-Speed Corner Detection), ORB [9] (fast binary descriptor based on BRIEF [10] (Binary robust independent elementary features)), BRISK [11] (Binary Robust Invariant Scalable Keypoints), HOSS [12] (histogram of oriented self-similarity) were proposed inspired by SIFT and PCA-SIFT [13] (Principal Component Analysis SIFT), ASIFT [14] (Affine-SIFT), ABSIFT [15] (Adaptive Binning SIFT), L2-SIFT [16] (SIFT matching for large image and

large-scale aerial photogrammetry), UC-SIFT [17] (Uniform competency SIFT) based on SIFT in recently years.

All of the above local feature matching methods [12,18,19] can work well under certain conditions and test data, but they are rarely used in real applications of remote sensing, and much less in aerial image tie point matching in real surveying area, except for SIFT. So far, SIFT and LSM have become the main application methods in real aerial photogrammetry for point coordinates measurement automatically. However, the majority of the existing handcrafted matching methods [20,21] show poor performance (Figure 1) under the condition of significant illumination variations, especially, weak texture images (e.g., mountainous area) for the model tie-in forward overlap or lateral overlap, resulting in the airline network not being constructed well because of some multi-image tie points lacking (Figure 2). Inspired by the deep learning approaches [22,23], in particular, CNN (convolutional neural network) [24], more and more works of image matching based on deep learning were presented [25,26], including MatchNet [27], LIFT [28], SuperGlue [29], LoFTR [30], etc., and proven to be better than Handcrafted matching methods, in which SuperGlue [29] performed better in matching effect and efficiency.



**Figure 1.** Weak texture image matching case (colored lines represent corresponding matching points). (a) There is no match point by SIFT method; (b) There are many match points by SuperGlue method.



**Figure 2.** Special texture images and image tie points by Block-SIFT. (a) Special texture images; (b) Image tie points by Block-SIFT. Green lines are forward image tie lines (no model tie points between a few models), red lines are lateral image tie lines (some images have no image tie points between airlines).

Motivated by the SuperGlue [29] and the artificial intelligence photogrammetry development needs, we propose LR-SuperGlue, a novel approach to large image tie point matching in real surveying area based on deep learning. Although the deep learning-based image processing method outperforms the traditional method in the field of intelligent interpretation and change detection of remote sensing images. At present, the deep learning-based image matching method can only process small images and requires high computer

performance, less in aerial image tie points matching. Hence, inspired by L2-SIFT [16], the image blocking strategy is used for large aerial image matching. At the same time, image overlap was calculated via a pyramid matching strategy for large image blocking, which helps to obtain much more image tie points. We evaluate the proposed method on a real surveying area with the measurement of the model tie point number and relative orientation compared with Block-SIFT (similar to L2-SIFT [16]). The experiments show that LR-SuperGlue could obtain much more model tie points and construct airline networks in mountainous area images, but Block-SIFT could not. Meanwhile, the accuracy of the relative orientation of LR-SuperGlue was acceptable. The rest of the paper is structured as follows: A brief review of some related work is presented below, followed by the paper's contributions. The details of the proposed method are described in Section 2. Section 3 represents the experiments and results, and finally, conclusions are described in Section 4.

### 1.1. Related Work

In this section, we briefly review the method of image matching, especially, methods used for real aerial photogrammetry. According to the learning or not, we roughly classify these methods into two groups, i.e., methods for hand-crafted image matching and methods for learned image matching.

#### 1.1.1. Hand-Crafted Image Matching

Since the appearance of SIFT [6], local features have played a crucial role in computer vision, becoming the standard for image matching. It used a Dog (Difference-of-Gaussian) instead of LoG (scale-normalized Laplacian of Gaussian) [31]. Moreover, the local gradient orientation and magnitudes were computed so that reliable features may be extracted even when there are rotation and translation between two images. In the matching procedure, a 128-dimension descriptor was utilized to match reliable features.

Many subsequent efforts focused on reducing its computational requirements. For instance, SURF [7] used Haar filters and integral images for fast keypoint detection and descriptor extraction. PCA-SIFT [13] adopted the dimensionality reduction strategy to express feature patches and projected high-dimensional samples into low-dimensional space. For different application requirements (different data), some works were proposed. For instance, ASIFT used six parameters of affine transform to obtain better results. L<sup>2</sup>-SIFT mainly adopts the image blocking strategy for large image matching in large-scale aerial photogrammetry. The UC-SIFT [32] approach obtains the best results compared to four other well-known detectors, including SURF and Hessian-affine for remote sensing image matching with rotation, scale and viewpoint transformations the literature on this topic is very extensive, more methods can refer these works [26,33–35]. Although the above methods have good experimental results in their respective papers, these methods still cannot match weak texture images well and have hardly been applied to the actual photogrammetry.

Because the local feature matching algorithm only reaches pixel accuracy and aerial photogrammetry needs sup-pixel accuracy, the LSM algorithm is often used to improve point accuracy. LSM is a procedure for digital image correlation based on least squares window matching, which added geometric and radiometric transformation parameters into window matching, improving the image matching accuracy.

#### 1.1.2. Learned Image Matching

While methods such as FAST [8] used machine learning techniques to extract key points, most early efforts in this area targeted descriptors. However, with the advent of deep learning, there has been a renewed push towards replacing all the components of the standard pipeline with convolutional neural networks. For example, MatchNet [27] trained a Siamese CNN for feature representation, followed by a fully-connected network to learn the comparison metric. DeepCompare [36] showed that a network that focuses on the center of the image can increase performance. While previous works [23,37,38] have successfully

tackled each one of these problems individually, LIFT [28] implemented the full feature point handling pipeline, that is, detection, orientation estimation, and feature description. To solve the problem of low execution efficiency of the deep learning matching method, the PN-NET [39] method adopts the image segmentation matching method and involves a feature descriptor network structure based on CNN, which improves the matching efficiency. L2Net [40] adopts a progressive sampling strategy and image segmentation method to learn feature descriptors, and the network model has good generalization ability. SuperPoint [41] method uses a self-supervised learning method to construct a fully convolutional network model for feature extraction and description. Different from the network architecture based on image segmentation, the network model is trained and tested on the full-size image, and the feature point positions and descriptors are calculated by a forward propagation.

With the development of deep learning-based image matching, graph neural network and attention [42] was adopted into image matching, including local feature extraction, feature description and description matching. Until now, SuperGlue is the one of best image matching methods (detects local features by SuperPoint) via deep learning, considering the matching effect and implementation in spite of the LoFTR [30] proposed recently. SuperGlue uses a graph neural network and attention to solve an assignment optimization problem, and handles partial point visibility and occlusion elegantly, producing a partial assignment. Due to the calculation on GPU, the matching speed is very fast.

Motivated by the deep learning-based image matching. In the field of remote sensing, some scholars have proposed matching methods based on deep learning in consideration of the characteristics of remote sensing images or different processing processes. Hao Zhu et al. [43] proposed network adopts a two-stage training mode to deal with the complex characteristics of RS images with better feasibility, robustness, and effectiveness. Shuang Wang et al. [44] proposed an effective deep neural network for remote sensing image registration, which optimized the whole processing through information feedback when training the network and improved the registration accuracy. Ye Yuanxin et al. [45] proposed the MU-Net for multimodal image registration, which achieved more comprehensive and accurate registration performance between these image pairs with geometric and radiometric distortions. In the last two years, the related research of image matching mainly focuses on multimodal images, such as [46–48]. Although the above methods have achieved good results in their own test data, the method cannot process the large and special texture image.

Although SuperGlue outperformed the traditional method in the field of image matching and got a very good matching result on different public data sets, these methods cannot be directly used because of the improper parameters and image size in this method for large aerial tie points matching. Then, in order to do some work promoting the intelligent development of photogrammetry; we must make changes and propose a novel method for large aerial image matching in real surveying areas.

## 1.2. Contributions

Image tie points in real aerial photogrammetry must be matched from two or more images according to the forward-track and lateral-track distribution of images. An important aspect of digital aerial triangulation is the need to match several large images simultaneously. In earlier years, tie points had to be marked on every image by the stereovision ability of a human operator. Hand-crafted image matching technology has led photogrammetry towards automation, and learned image matching will promote photogrammetry towards artificial intelligence.

In this paper, a novel image matching approach for a large aerial image in real and difficult surveying areas with weak texture images based on SuperGlue is presented. The aim of the proposed LR-SuperGlu method is to solve large and weak texture image matching (especially, multi-images tie points in forward overlap and lateral overlap) in real surveying

areas and make learned image matching possible in photogrammetry applications. More precisely, the main contributions of this paper are as follows:

- (1) Overcome the shortcomings of traditional methods for poor texture and large images in real and difficult surveying area. The classic SIFT method has huge applications in teaching in schools and software in corporations, but poor texture image matching cannot be solved well by SIFT, especially, for multiple image tie points in a difficult survey area. Fortunately, deep learning approaches have shown great power in weak texture image matching, such as SuperGlue. So, the proposed LR-SuperGlue will outperforms SIFT.
- (2) Make the SuperGlue method from open and small images to practical applications for large image matching in real aerial photogrammetry. Given the proposed image matching method, other image matching methods via deep learning may go into photogrammetry. Finally, photogrammetry will develop from automation to intelligent, in the theory and practice.

## 2. Methodology

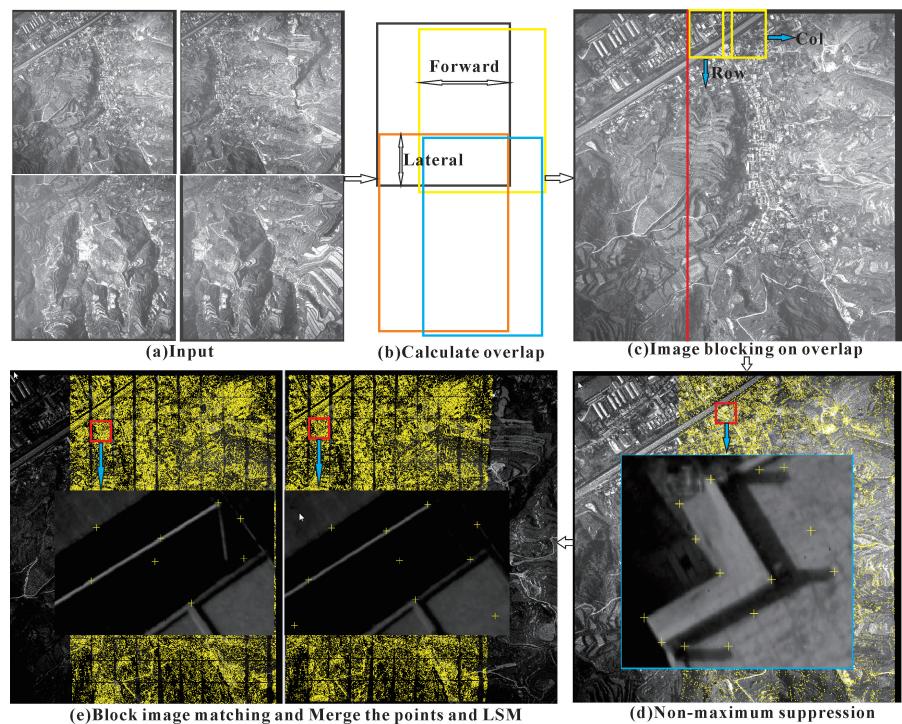
In this section, the workflow method is described first, and then the key steps are presented in detail. The implementation details of LR-SuperGlue are presented based on two pairs of large stereo images IL, IR and IU, ID from a real surveying area. IL and IR are the left and right images in the pairs in the forward overlap. IU and ID is the up and down image in the pairs in the lateral overlap.

### 2.1. Workflow

Given a pair of large stereo images IL, IR or IU, ID, one of the core steps of the LR-SuperGlue method is to process the “large” image because the original SuperGlue method only matches the small stereo images on GPU (the specific pixel size depends on the GPU size). Experimental results show that if the image size exceeds  $2000 \times 2000$  with the default parameters, the program just exits with GPU size 8 G because of too many match points. At the same time, if the texture of the input image is very rich (e.g., urban aerial image or forest image), the image size cannot exceed  $1000 \times 1000$ . Since GPU is very expensive, we must modify its parameters and do image blocking in the face of large images.

To make the two block-images (included in IL, IR or IU, ID) include the same ground area as far as possible (corresponding blocks), we should accurately calculate the overlap of input large stereo and consider the invalid pixels with 0 values. In general, the workflow of the proposed method follows an iterative manner within multi-block images. The main steps of the proposed method are as follows (Figure 3):

- (1) Accurately calculate the overlap of input large stereo image through a certain feature matching algorithm (e.g., SIFT or SuperGlue in this paper).
- (2) Image blocking for large stereo image with block size  $1000 \times 1000$  (if the GPU size is enough, the block size may be to  $2000 \times 2000$ ) based on the overlap size in row (y) or column direction (x). If the images have invalid pixels in the image edge, image blocking should consider this situation.
- (3) Modify the default parameters of original SuperGlue (model test parameters not the trained model parameters). The specific parameters are the input image size and the window size of non-maximum suppression.
- (4) Perform block image matching with the improved SuperGlue. If the image gray contrast is poor, image enhancement processing can be performed.
- (5) Merge the block matching results and optimize the initial matching point with LSM for subpixel accuracy. The initial matching point may be processed with RANSAC (Random Sample Consensus) if needed.



**Figure 3.** Workflow of proposed LR-SuperGlue. (a) is the input images; (b) calculates the overlap of two images in forward and lateral direction; (c) image block matching with SuperGlue(the yellow box represents the image block); (d) is the NMS; (e) merges block points and RANSAC, LSM (the yellow “+” represents the matching point).

## 2.2. Accurately Calculate the Overlap and Start Matching Position

Accurately calculating the overlap of stereo images is the key point of image blocking, which helps to make sure all overlap pixels are matched. In general, we can acquire the forward overlap and lateral overlap from flight information or may calculate them via POS data if it exists. However, the overlap obtained by the above approach often is inaccurate and POS data may be missing, especially in the early years. At this point, the accurate overlap is calculated by the pyramid image matching strategy.

In this paper, two robust image matching methods are selected to calculate the overlap. It turns out that the overlap of most of the images can be accurately calculated by SIFT and RANSAC in a lot of experiments. In order to ensure the robustness of the calculation method of overlap, SuperGlue is used to replace SIFT automatically in the proposed algorithm if SIFT does not work. If the stereo images are IL, IR, the X-overlap ( $x_0$ ) will be got (Equation (1)), and if the stereo images are IU, ID, the Y-overlap ( $y_0$ ) will be got (Equation (2)).

$$xo = \text{float}(\text{Col}_{Ori}/\text{Col}_{Pyr} \times \frac{1}{n} \sum_{i=1}^n (\text{Lx}_{iOri} - \text{Rx}_{iOri})) \quad (1)$$

where,  $Col_{Ori}$  is the width of original Large image;  $Col_{Pyr}$  is the width of pyramid image;  $Lx_{iOri}$  is the value of left pyramid image matching point;  $Rx_{iOri}$  is the of right pyramid image matching point.

$$yo = \text{float}(Row_{Ori}/Row_{Pyr}) \times \frac{1}{n} \sum_{i=1}^n (Uy_{iOri} - Dy_{iOri}) \quad (2)$$

where,  $Row_{Ori}$  is the height of Original Large image;  $Row_{Pyr}$  is the height of pyramid image;  $Uy_{iOri}$  is the  $y$  value of up pyramid image matching point;  $Dy_{iOri}$  is the  $y$  value of down pyramid image matching point.

To accurately image block in the next step, the start matching row and col position needs to be calculated ( $sy$  for IL, IR,  $sx$  for IU, ID) Equation (3)).

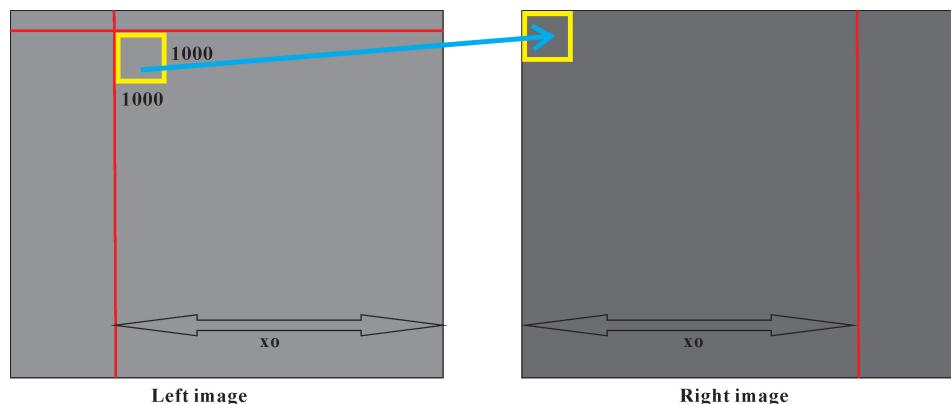
$$sy = L/Uy_{0Ori} - R/Dy_{0Ori}, sx = L/Ux_{0Ori} - R/Dx_{0Ori} \quad (3)$$

### 2.3. Image Blocking Based on Overlap

As mentioned above, the original SuperGlue method on GPU only matches small images (other deep learning methods as well). For this reason, image blocking based on overlap appears to be of the essence. Given or image blocking will be performed in the following steps: (1) calculate the valid matching image region width for IL, IR and height for IU, ID (Equation (4)), which helps to obtain good image matching results and reduce the computation time of LR-SuperGlue; (2) set the block size according to the original image texture, and  $1000 \times 1000$  pixels as the block size is used in this paper (Figure 3c).

$$valid_{LR} = xo \times Col_{Ori}, valid_{UD} = yo \times Row_{Ori} \quad (4)$$

The image blocking of IL, IR is easy, but sometimes it does not work for IU, ID in the lateral image if pyramid image matching does not work well. At the same time, due to the complexity of image texture, the block size should be calculated automatically if possible. To reduce the uncertainty of image blocking, a pre-overlap of the stereo image is used for the current stereo image overlap adopting an iterative approach until image matching works well which is different from the general image blocking strategy (Figure 4).



**Figure 4.** Image blocking of LR-SuperGlue in forward images, lateral images do same (the yellow box and blue arrow represent the corresponding image block).

### 2.4. SuperGlue Matching after Image Blocking

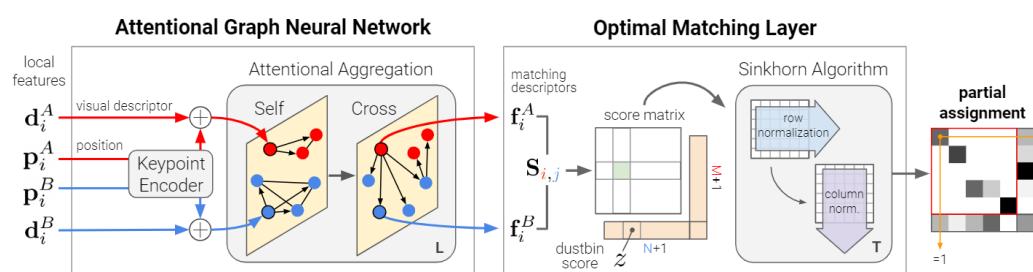
Block matching of SuperGlue can be performed after image blocking. SuperGlue is a feature matching Network based on deep learning, which is composed of two main modules: The Graph Neural Network (GNN) based on attention and the optimal matching layer (as shown in Figure 5). Attention GNN encodes feature points and descriptors as a feature matching vector, and then uses self-attention and cross-attention to enhance feature matching performance. After attention GNN is completed, the optimal matching layer is entered. First, the matching degree score matrix is obtained by calculating the inner product of feature matching vector, and then the optimal feature allocation matrix is solved by the Stinkhorn algorithm. Given two images on the left and right, each image has the position of feature points and corresponding descriptors, which can be used to represent image features. The matching network uses SuperPoint method to extract feature points.

After feature points are extracted, a complete graph of left and right images is constructed (see Figure 6), whose nodes are each feature point in the image. This graph contains two different undirected edges: one connects the feature points inside the left and right images (using self-attention), and the other connects the feature points between

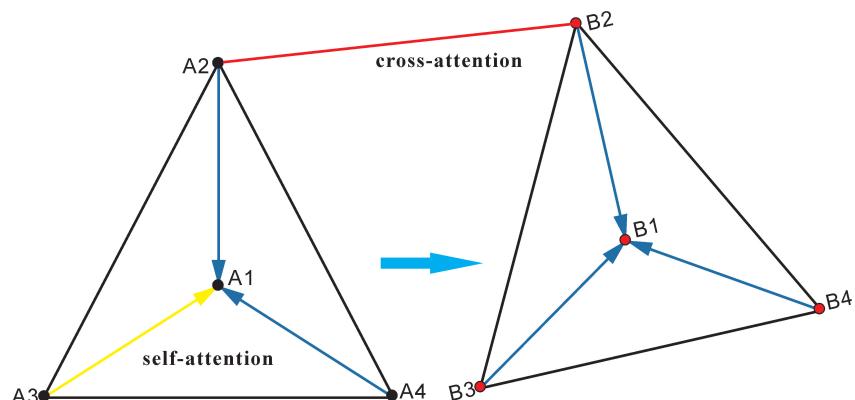
the left and right images (using cross-attention). The attention mechanism simulates the process of human eyes searching for matching points back and forth in the key areas of the left and right images. It is represented by the weight of attention. The larger the attention is, the more similar the two features are. If it is not the desired match, focus on the surrounding feature points until a match is found or no match is found. In the above process, the network model increases the specificity of neighborhood feature points by actively searching for them. Specifically, attention weight  $\alpha_{ij}$  is the Softmax of similarity of query and search feature points, namely:

$$\alpha_{ij} = \text{softmax}_i(q_i^T k_j) \quad (5)$$

where,  $q_i$  represents the self-attention mapping query value of feature points, and  $k_j$  represents the key value of corresponding attributes; Softmax represents an activation function and maps the output results to  $(0, 1)$ .



**Figure 5.** SuperGlue Network [29].



**Figure 6.** Complete graph of feature points ( $A_1, A_2, A_3, A_4$  and  $B_1, B_2, B_3, B_4$ ).

## 2.5. Non-Maximum Suppression of SuperGlue Matching Points

The aim of non-maximum suppression (NMS) is to reduce memory footprint and reduce the time consumption of relative orientation (or more complicated calculations for bundle block adjustment in further research). Theoretically speaking, original SuperGlue can obtain good image matching results via image blocking (e.g.,  $1000 \times 1000$ ). However, in practice, complex image texture (a surveying area includes rich textures and weak textures images) often makes part of the image matching unsuccessful, resulting in airline networks failing to be tied. NMS selects the matching point with the highest accuracy within the set window size. In this paper, the window size of 4 pixels is selected as the initial NMS window size ( $ws$ ). In the proposed approach, there are several cases automatically selecting window size, as follows (Equation (6)):

$$ws = \begin{cases} 4, & \text{initial.} \\ 8, & \text{lateral.} \\ 16, & \text{forward.} \end{cases} \quad (6)$$

## 2.6. Optimize Image Tie Points with RANSAC and LSM

Existing literature shows that feature matching can only achieve pixel-level accuracy, as well as SuperGlue. Different from other literature evaluating image matching quality with recall (Equation (7)), precision, rmse (root-mean-square error) [32], etc., on simulated data with pixel-level accuracy, aerial photogrammetry needs subpixel accuracy.

$$\text{recall} = \text{cm}/\text{tm.}, \text{precision} = \text{cm}/(\text{cm} + \text{fm}). \quad (7)$$

where,  $\text{cm}$  (correct match) and  $\text{fm}$  (false match) are the number of correctly and falsely matched point pairs in the matching results, respectively, and (total match) is the total number of existing correctly matched point pairs in the initial matched point sets.

So, after the initial matching points are obtained, LSM [2] method is used to optimize them to obtain subpixel accuracy image tie points for relative orientation (Algorithm 1). Meanwhile, matching points of large images merged from a lot of block images should be optimized with the RANSAC method [49], which can remove the mismatch points (outlier) [50–53].

---

### Algorithm 1 Large image matching in real surveying area

---

**Input:** large images from forward and lateral direction

**Output:** large image matching point sets

```

1: Pseudo-code:
2: //step1:calculate the overlap of two images
3: PointSets = PyramidSIFT(image1,image2);
4: for each  $i \in [0, \text{Num}]$  do
5:   if ( $IL, IR$ ) then
6:      $xo = xo + abs(Lx_i - Rx_i);$ 
7:   end if
8:   if ( $IU, ID$ ) then
9:      $yo = yo + abs(Uy_i - Dy_i);$ 
10:  end if
11: end for
12:  $xo = xo/\text{Num}, yo = yo/\text{Num};$ 
13: //step2:image blocking based on overlap
14: if ( $IL, IR$ ) then
15:   imageblockL = largeimage(Rect(sx, 0, 1000, 1000));
16:   imageblockR = largeimage(Rect(0, sy, 1000, 1000));
17: end if
18: if ( $IU, ID$ ) then
19:   imageblockU = largeimage(Rect(sy, 0, 1000, 1000));
20:   imageblockD = largeimage(Rect(sx, 0, 1000, 1000));
21: end if
22: //step3: block image matching with SuperGlue and NMS automatically
23: while  $i < BlockNum$  do
24:   do SuperGlue via Python
25: end while
26: initialNMS = 4;
27: lateralNMS = 8;
28: forwardNMS = 16;
29: //step4:merge the block matching results
30: largeresult = add(blockimageresult);
31: //step5:optimize the initial matching point with LSM and RANSAC
32: LSM(RANSAC(largeresult));

```

---

### 3. Experiments and Results

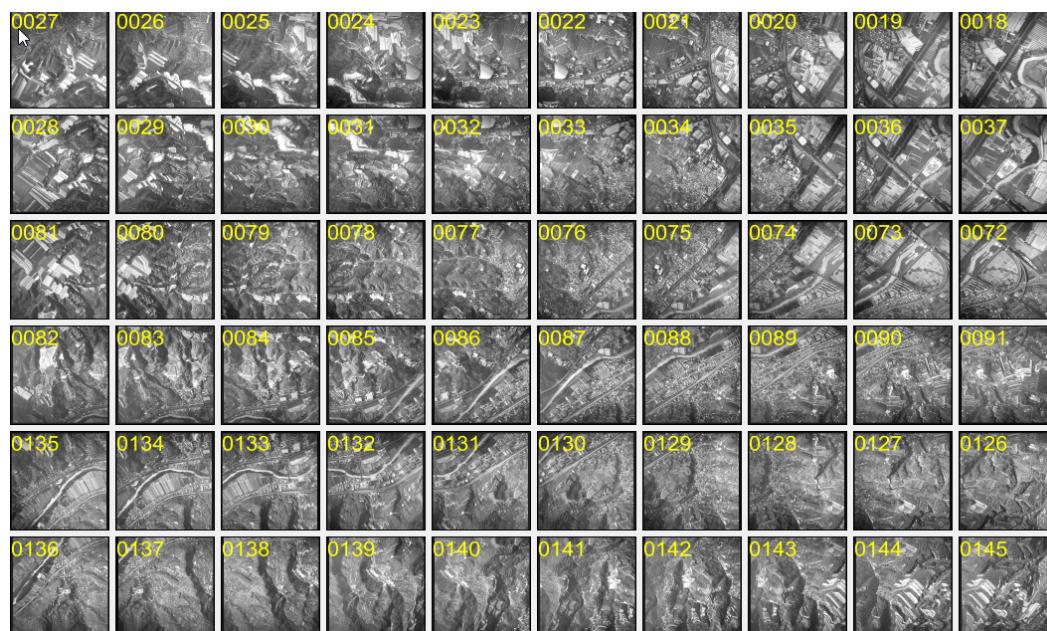
In this section, description of test data is presented firstly. Next, in order to better illustrate the effectiveness of this method Compared with Block-SIFT method, the matching quality measures are elaborated in detail. Finally, detailed experimental results and analysis of two methods are described. The experimental platform is Windows 10 (CPU is E5-1603-2.8 GHz, the memory is 16 G—1066 MHz, and the GPU is 8 G-NVIDIA Quadro K5200). The algorithm implementation tool is Visual Studio 2015 (C++ language for Block-SIFT and others) and Pycharm 2017 (C++ and Python language for LR-SuperGlue).

#### 3.1. Test Data

The LR-SuperGlue method is proposed to mainly solve large image matching for airline network tie in real surveying, 60 images with 6 airlines are used in experiment in this paper. The details of the data are shown in Table 1 and Figure 7.

**Table 1.** Information of surveying area.

| Surveying Area Name | Width × Height of Image | Pixel Size ( $\mu\text{m}$ ) | f of Camera (mm) | Total Number of Image | Forward Overlap | Lateral Overlap |
|---------------------|-------------------------|------------------------------|------------------|-----------------------|-----------------|-----------------|
| SX500               | 28,820 × 30,480         | 4.6                          | 142.4169         | 6 × 10 = 60           | 63%             | 30%             |

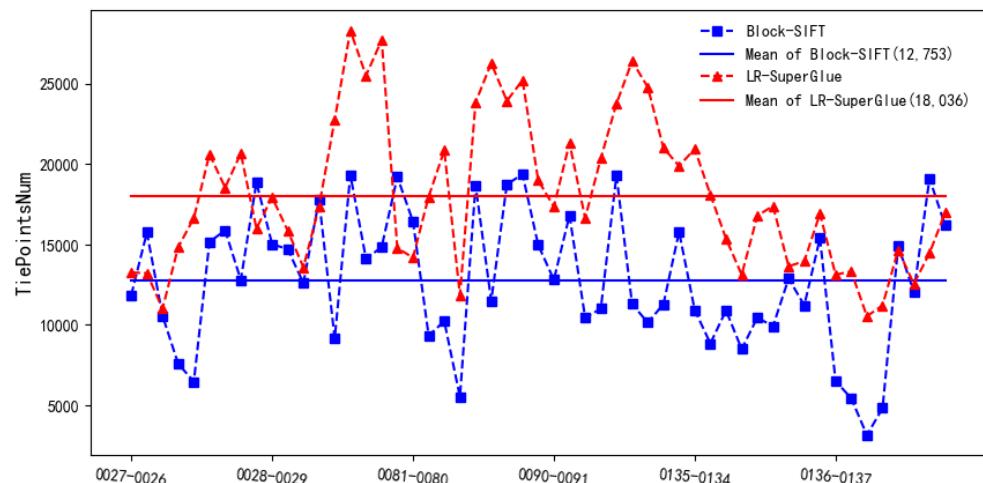


**Figure 7.** Images from a real surveying area. The western part of the survey area is a typical loess plateau landform with ravines and ravines and the terrain is very complex and contains many special texture images. The yellow number indicates the image number.

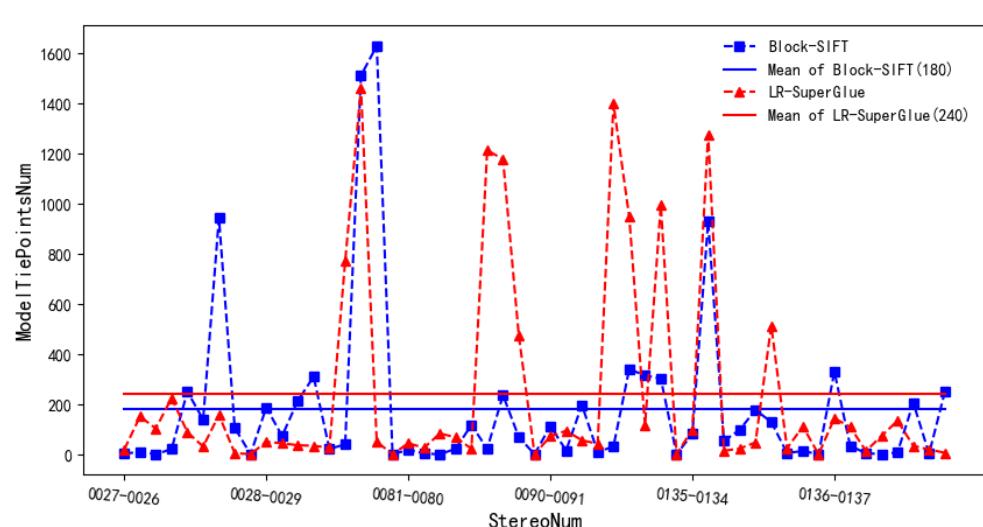
#### 3.2. Matching Results in Forward Images

The performance evaluation of LR-SuperGlue for the forward images is presented in this section. The comparative results for the Block-SIFT method, and for number of matching points, number of model tie and relative orientation accuracy, are shown in Figures 8 and 9. As can be seen, the LR-SuperGlue significantly outperforms the Block-SIFT in most stereo model cases, especially for poor texture images. Specifically, in terms of model tie points and relative orientation accuracy, image tie point matching by the Block-SIFT method makes some stereo models lack tie points, but the LR-SuperGlue method can obtain tie points in these stereo models. Meanwhile, the relative orientation accuracy (unit weight median error) of the LR-SuperGlue method is higher than the block-SIFT method on the whole (the unit weight median error of block-SIFT is 3.64  $\mu\text{m}$ , and the LR-SuperGlue's is 2.85), and almost all of them are about 0.6 pixel. The main reason is that

the SIFT method relies on the image gray, and it cannot extract effective feature points in the weak texture area of the image, resulting in no matching points or only a small number of matching points between these images. However, during model training, the SuperGlue method adopts a large number of image samples with different textures and geometric changes of images, so that the method can match a large number of corresponding points in weak texture and large geometric deformation areas, and the accuracy meets basically the requirements of aerial photogrammetry due to the use of graph neural network.

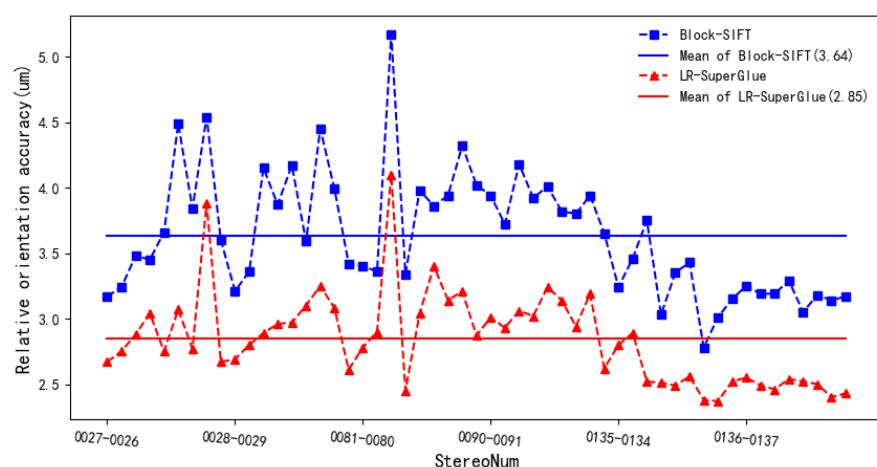


(a) Forward image tie points number



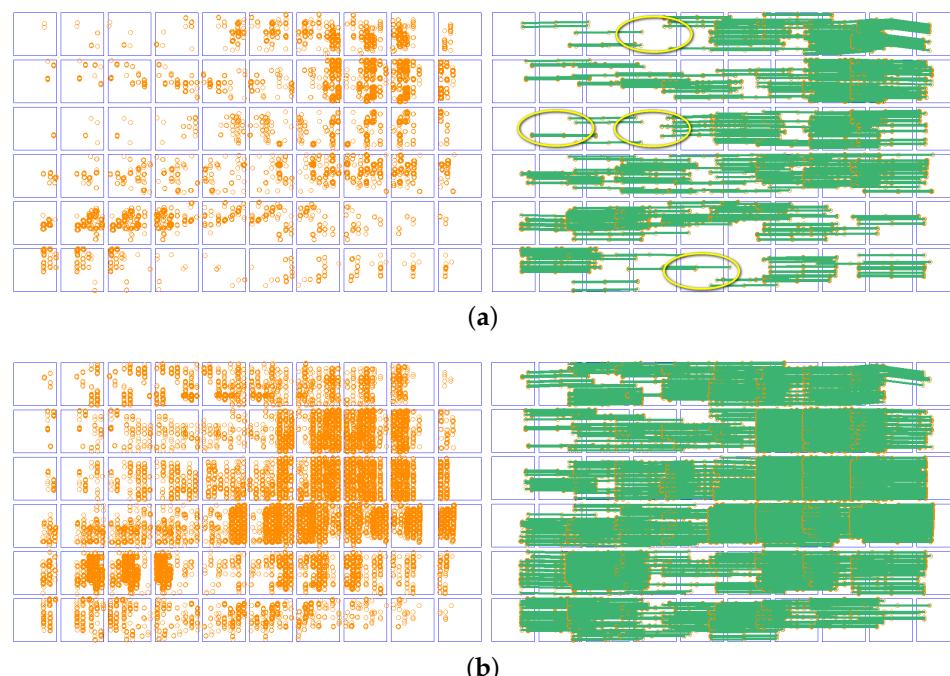
(b) Forward model tie points number

**Figure 8.** Forward image tie points number: (a) Forward image tie points are chosen from the stereo matching; (b) is the model tie points number of Block-SIFT and LR-SuperGlue method.



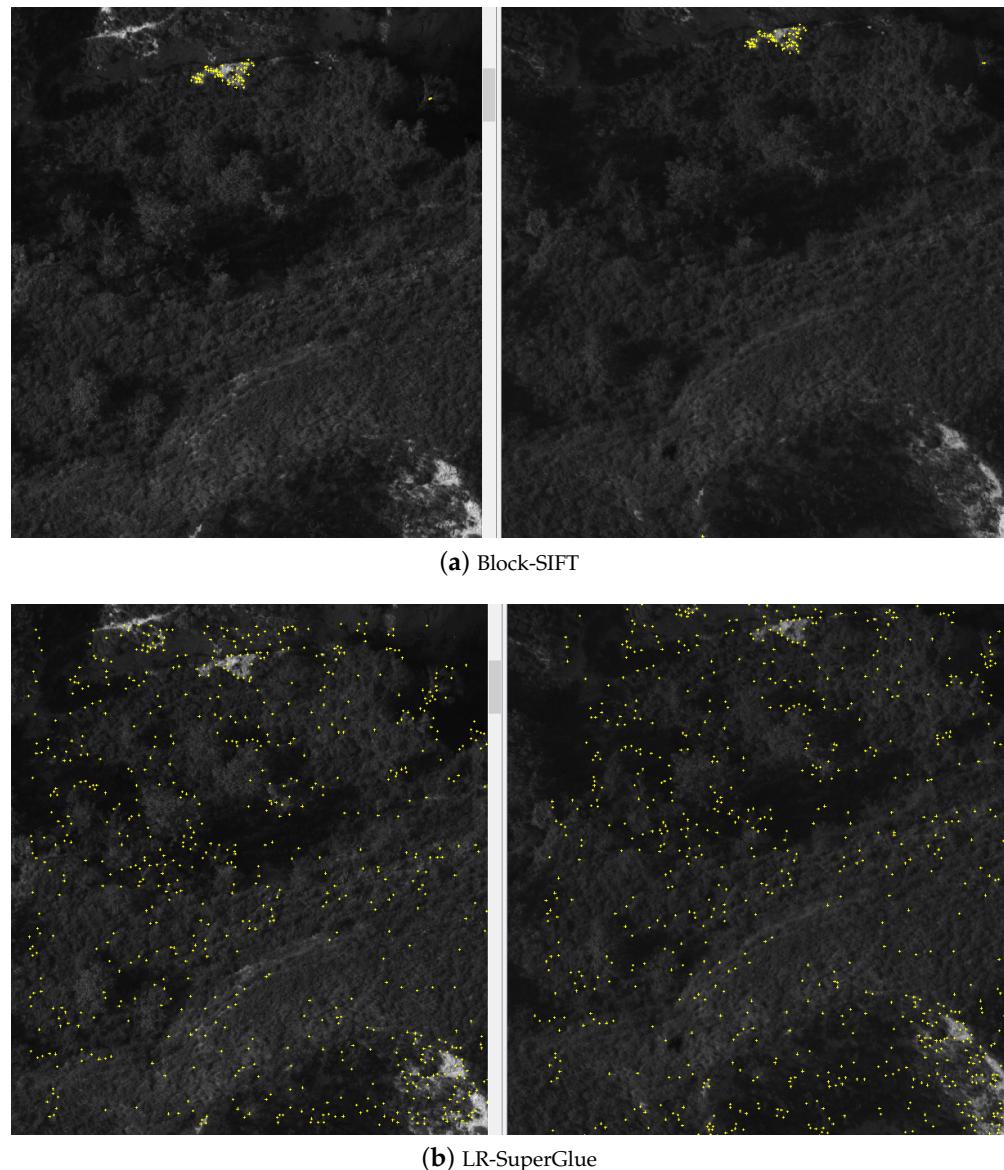
**Figure 9.** Relative orientation accuracy (unit weighted errors). The camera pixel size is  $4.6 \mu\text{m}$ . Relative orientation refers to the process of restoring the relative positional relationship of two photos in a stereo pair. Through the relative orientation, it is possible to achieve the intersection of pairs of rays of the corresponding points, achieving the purpose of constructing a stereo model.

In fact, a large number of matching points and then a certain number of reorientation points can be obtained by using both methods only from the independent stereo pair matching (Figure 8). In addition to orientation points (tie points of overlap in two adjacent forward images, namely image tie points), a certain number of model tie points of overlap in three adjacent forward images are needed in practical photogrammetry. Only in this way can a whole airline be connected. As can be seen from Figure 10, the matching points obtained through LR-Superglue can connect each airline and obtain more model tie points on the whole, and the relative orientation meets the requirement of accuracy. However, the matching points obtained through the block-SIFT method cannot be connected in the first (stereo model 0025-0024 and 0024-0023) and the third route (stereo model 0079-0078 and 0078-0077).



**Figure 10.** Show of forward image model tie points (orange circle). There are 0 or 1 model tie point in the yellow circle. (a) Block-SIFT (Left is model tie points and right is tie lines); (b) LR-SuperGlue (Left is model tie points and right is tie lines (green line)).

Therefore, there is a certain disharmony between image matching and model tie, especially in the weak texture region and the transition region between strong texture and weak texture. In fact, the block-SIFT method is already very robust and is used by many photogrammetric software. However, when conducting a survey area similar to the image in this paper, local models may not be tied. When there is no tie points between models, the manual addition of model connection points is often required in the actual photogrammetry production, but this requires a higher cost. Compared with the block-SIFT method, the proposed method does not have the case of no model tie points (although some models only have a few model tie points). The main reason is that the LR-Superglue method can match more high-precision image corresponding points, even in weak texture areas (Figure 11).



**Figure 11.** Matching result of weak texture image: (a) There are only a few matching points (yellow "+") by Block-SIFT method; (b) There are much more matching points by LR-SuperGlue method.

In terms of time consumption, LR-SuperGlue has no great advantage. If both methods use CPU devices, a stereo matching will take about 15 min for both methods in test data. If both methods use GPU devices, LR-SuperGlue takes 3 min and Block-SIFT takes 12 min.

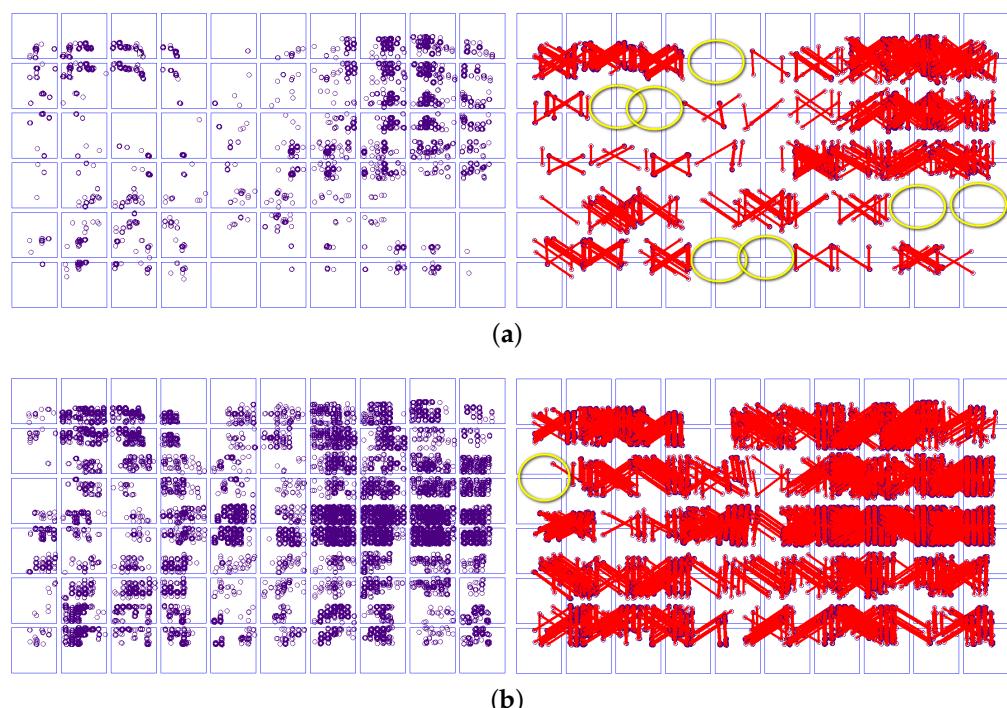
However, it is possible to use multithreading with a CPU device, whereas a GPU device is difficult because GPU has limited capacity and very expensive currently.

### 3.3. Matching Results in Lateral Images

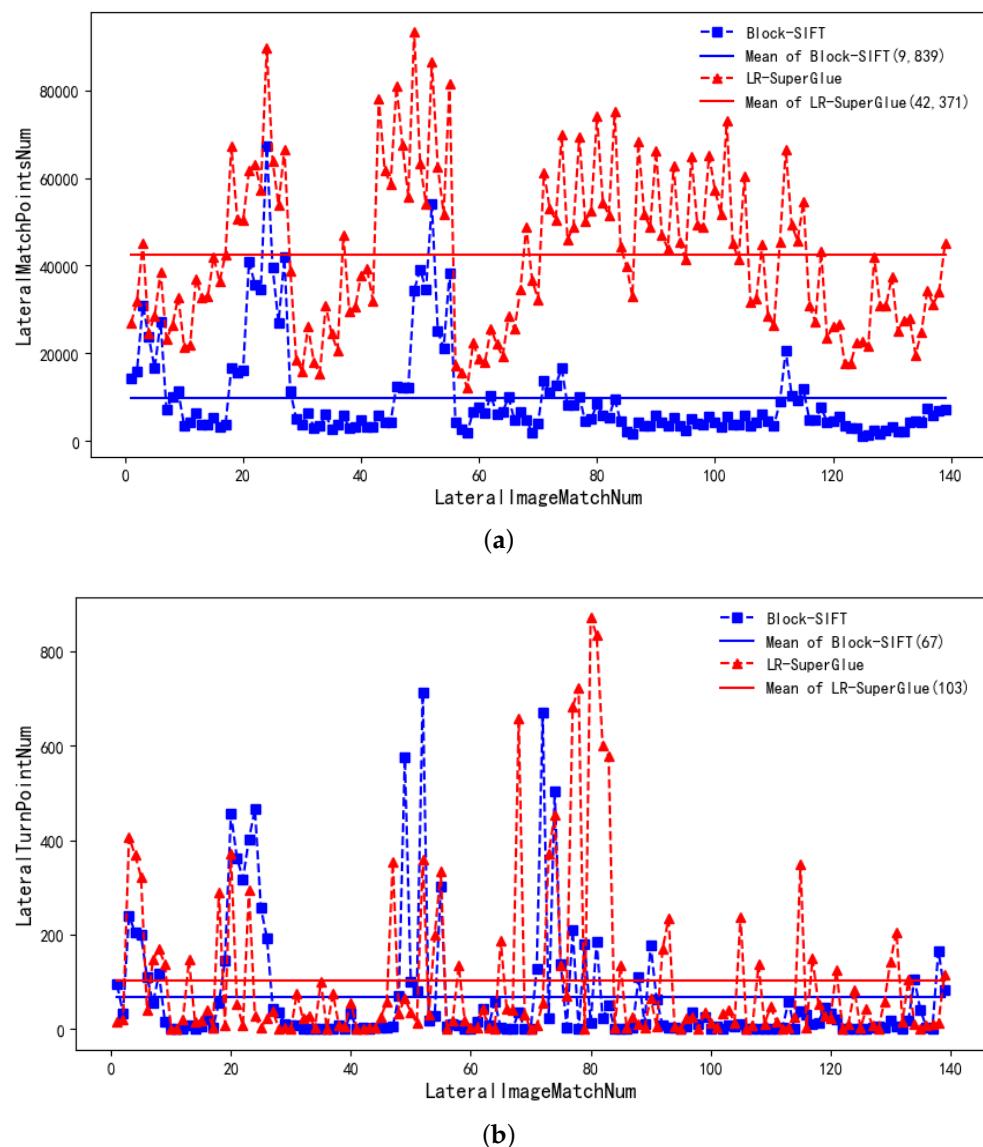
The performance evaluation of LR-SuperGlue for the lateral images is presented in this section. The comparative results for the Block-SIFT method and for the number of matching point and turning points between lateral images are shown in Figure 12 and Figure 13. As can be seen, the LR-SuperGlue significantly outperforms the Block-SIFT in most of overlap image cases. The main reason is same as the previous section about forward images matching for weak texture images.

After the completion of forward image matching and model tie, the main purpose of lateral image matching is to carry out turning points of adjacent route, that is, to carry out route tie. In fact, the turning points between the adjacent routes are based on relative orientation points (Figure 8a). The number of turning points is not only related to the image matching between adjacent routes, but also related to the distribution and number of reorientation points. Compared with the model tie within the airline, the strip connection does not require connection points for per adjacent images. As long as there are a certain number of connection points between adjacent strips, the free route network construction of the whole survey area can be completed.

As can be seen from Figure 13a, when matching adjacent images between each pair of airstrips, the proposed method can obtain more matching points than the block-SIFT method. However, in terms of the number of turning points between airstrips, although the method in this paper can obtain more turning points on the whole, no turning points occur in both methods. At the same time, the block-SIFT tie point between some images is 0 especially the first and second routes, as well as the third and fourth routes, but the proposed method has a few turn points (Figure 13b).



**Figure 12.** Show of lateral image tie points (purple circle). There are 0 or 1 tie point in the yellow circle. (a) Block-SIFT (Left is tie points and right is tie lines (red line)); (b) LR-SuperGlue (Left is tie points and right is tie lines).



**Figure 13.** Lateral image matching and tie points number. (a) Lateral image matching points number; (b) Lateral image turning points number.

#### 4. Conclusions and Future Work

Image tie point matching is an essential task in real aerial photogrammetry, especially for three overlap regions in the forward direction (model tie points). In recent decades, a large number of image matching methods have emerged, which can obtain a better matching effect between two images under certain conditions. However, in current photogrammetry production, the SIFT method is still the main method, and the RANSAC and LSM methods are used for point location optimization. The SIFT method is very robust, but in actual production, only a few or no corresponding points can be matched between a few images (mainly repeated texture or poor texture image), which leads to the failure of model tie, and then the model tie points can only be manually added. Manual addition can solve the problem, but it will cost more and add a limited number of model tie points.

With the development of deep learning, many scholars have proposed many image matching methods based on deep learning. These methods are often better in weak texture and small image matching than traditional methods, but they cannot be directly applied to large image tie point matching in real photogrammetry. Considering the actual photogrammetry needs and the shortcomings of SIFT, this paper proposes a deep learn-

based LR-Superglue matching method for large aerial image tie points matching. The main steps are as follows: Firstly, the pyramid matching strategy is used to accurately calculate the image overlap degree, then the block matching and matching point screening (non-local maximum suppression) are carried out, then the matching point combination is carried out, and finally the mismatching elimination and point location optimization are carried out. Based on the proposed method, the experimental results are compared with the block-SIFT images. In the comparative experiment, the number of model tie points, relative orientation accuracy and the number of turning points between airstrips are mainly used as evaluation indexes. The results show that compared with the block-SIFT method, the proposed method can realize the construction of the free route network in the actual survey area, while the block-SIFT method has a few models with no tie points and the relative directional accuracy is lower than the proposed method. Therefore, the LR-Superglue method solves the problem of missing or insufficient orientation points of weak texture survey area model to a certain extent, and also proves that the deep learning method can be initially used in practical photogrammetry, which is conducive to promoting the further development of intelligent photogrammetry in terms of geometric processing, not just intelligent interpretation of remote sensing images.

Although the LR-Superglue method achieves better experimental results in image tie point matching and free route network construction than the block-SIFT method, this paper only uses one surveying area for comparison experiment and does not complete the final bundle block adjustment (considering the temporary absence of control points in this surveying area). Therefore, in order to apply the proposed method to practical photogrammetry, it is necessary to carry out experiments in more surveying areas and use control points to do the bundle block adjustment and evaluate the accuracy in the future study, as well as generate a DOM (Digital Orthophoto Map). In addition, the deep learning matching model more suitable for the characteristics of aerial images should be strengthened.

**Author Contributions:** Conceptualization, X.Y. (Xiuli Yuan), X.Y. (Xiuxiao Yuan) and J.C.; Data curation, X.Y. (Xiuli Yuan); Investigation, X.Y. (Xiuli Yuan) and X.Y. (Xiuxiao Yuan); Methodology, X.Y. (Xiuxiao Yuan) and J.C.; Resources, X.Y. (Xiuxiao Yuan); Writing—original draft, X.Y. (Xiuli Yuan), X.Y. (Xiuxiao Yuan); funding acquisition, J.C. and X.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the National Natural Science Foundation of China [Grants Number 41771479] and the National High-Resolution Earth Observation System (the Civil Part) [Grant No. 50-H31D01-0508-13/15].

**Acknowledgments:** Thanks for the test data provided by the Fifth Academy of Aerospace Science and technology 508.

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

## References

1. Yao, W. Autocorrelation Techniques for Soft Photogrammetry. Ph.D. Thesis, Iowa State University, Ames, IA, USA, 1997.
2. Ackermann, F. Digital image correlation: performance and potential application in photogrammetry. *Photogramm. Rec.* **1984**, *11*, 429–439. [[CrossRef](#)]
3. Moravec, H.P. Rover Visual Obstacle Avoidance. In Proceedings of the 7th International Joint Conference on Artificial Intelligence, IJCAI '81, Vancouver, BC, Canada, 24–28 August 1981; Volume 81, pp. 785–790.
4. Förstner, W.; Gülich, E. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, 2–4 June 1987; pp. 281–305.
5. Harris, C.; Stephens, M. A combined corner and edge detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; Volume 15, pp. 147–151.
6. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
7. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [[CrossRef](#)]

8. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In Proceedings of the ECCV 2006, 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 430–443.
9. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
10. Calonder, M.; Lepetit, V.; Vincent, S.; Strecha, C.; Fua, P. Brief: Binary robust independent elementary features. In Proceedings of the European Conference on Computer Vision, Berlin/Heidelberg, Germany, 5–11 September 2010; pp. 778–792.
11. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.
12. Sedaghat, A.; Mohammadi, N. Illumination-Robust remote sensing image matching based on oriented self-similarity. *ISPRS J. Photogramm. Remote Sens.* **2019**, *153*, 21–35. [[CrossRef](#)]
13. Ke, Y.; Sukthankar, R. PCA-SIFT: A more distinctive representation for local image descriptors. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, Washington, DC, USA, 27 June–2 July 2004; Volume 2, p. II.
14. Morel, J.M.; Yu, G. ASIFT: A new framework for fully affine invariant image comparison. *SIAM J. Imaging Sci.* **2009**, *2*, 438–469. [[CrossRef](#)]
15. Sedaghat, A.; Ebadi, H. Remote sensing image matching based on adaptive binning SIFT descriptor. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5283–5293. [[CrossRef](#)]
16. Sun, Y.; Zhao, L.; Huang, S.; Yan, L.; Dissanayake, G. L2-SIFT: SIFT feature extraction and matching for large images in large-scale aerial photogrammetry. *ISPRS J. Photogramm. Remote Sens.* **2014**, *91*, 1–16. [[CrossRef](#)]
17. Sedaghat, A.; Mokhtarzade, M.; Ebadi, H. Uniform robust scale-invariant feature matching for optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4516–4527. [[CrossRef](#)]
18. Li, J.; Hu, Q.; Ai, M. RIFT: Multi-modal image matching based on radiation-invariant feature transform. *arXiv* **2018**, arXiv:1804.09493.
19. Ye, Y.; Bruzzone, L.; Shan, J.; Bovolo, F.; Zhu, Q. Fast and robust matching for multimodal remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9059–9070. [[CrossRef](#)]
20. Xiao, G.; Luo, H.; Zeng, K.; Wei, L.; Ma, J. Robust Feature Matching for Remote Sensing Image Registration via Guided Hyperplane Fitting. *IEEE Trans. Geosci. Remote Sens.* **2020**, *60*, 1–14. [[CrossRef](#)]
21. Huang, X.; Wan, X.; Peng, D. Robust feature matching with spatial smoothness constraints. *Remote Sens.* **2020**, *12*, 3158. [[CrossRef](#)]
22. Ma, J.; Jiang, X.; Fan, A.; Jiang, J.; Yan, J. Image matching from handcrafted to deep features: A survey. *Int. J. Comput. Vis.* **2021**, *129*, 23–79. [[CrossRef](#)]
23. Jiang, W.; Trulls, E.; Hosang, J.; Tagliasacchi, A.; Yi, K.M. COTR: Correspondence Transformer for Matching Across Images. *arXiv* **2021**, arXiv:2103.14167.
24. Jin, Y.; Mishkin, D.; Mishchuk, A.; Matas, J.; Fua, P.; Yi, K.M.; Trulls, E. Image matching across wide baselines: From paper to practice. *Int. J. Comput. Vis.* **2021**, *129*, 517–547. [[CrossRef](#)]
25. Zagoruyko, S.; Komodakis, N. Learning to compare image patches via convolutional neural networks. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4353–4361.
26. He, H.; Chen, M.; Chen, T.; Li, D. Matching of remote sensing images with complex background variations via Siamese convolutional neural network. *Remote Sens.* **2018**, *10*, 355. [[CrossRef](#)]
27. Han, X.; Leung, T.; Jia, Y.; Sukthankar, R.; Berg, A.C. Matchnet: Unifying feature and metric learning for patch-based matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3279–3286.
28. Yi, K.M.; Trulls, E.; Lepetit, V.; Fua, P. Lift: Learned invariant feature transform. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 467–483.
29. Sarlin, P.E.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learning feature matching with graph neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4938–4947.
30. Sun, J.; Shen, Z.; Wang, Y.; Bao, H.; Zhou, X. LoFTR: Detector-free local feature matching with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 8922–8931.
31. Lindeberg, T. Edge detection and ridge detection with automatic scale selection. *Int. J. Comput. Vis.* **1998**, *30*, 117–156. [[CrossRef](#)]
32. Sedaghat, A.; Mohammadi, N. Uniform competency-based local feature extraction for remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 142–157. [[CrossRef](#)]
33. Wei, D.; Zhang, Y.; Liu, X.; Li, C.; Li, Z. Robust line segment matching across views via ranking the line-point graph. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 49–62. [[CrossRef](#)]
34. Li, J.; Hu, Q.; Ai, M.; Zhong, R. Robust feature matching via support-line voting and affine-invariant ratios. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 61–76. [[CrossRef](#)]
35. Dong, Y.; Jiao, W.; Long, T.; Liu, L.; He, G.; Gong, C.; Guo, Y. Local deep descriptor for remote sensing image feature matching. *Remote Sens.* **2019**, *11*, 430. [[CrossRef](#)]

36. Yang, T.Y.; Hsu, J.H.; Lin, Y.Y.; Chuang, Y.Y. DeepCD: Learning deep complementary descriptors for patch representations. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3314–3322.
37. Luo, Z.; Shen, T.; Zhou, L.; Zhu, S.; Zhang, R.; Yao, Y.; Fang, T.; Quan, L. Geodesc: Learning local descriptors by integrating geometry constraints. In Proceedings of the Computer Vision—ECCV 2018—15th European Conference, Munich, Germany, 8–14 September 2018; pp. 168–183.
38. Zhang, Z.; Lee, W.S. Deep graphical feature learning for the feature matching problem. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 5087–5096.
39. Balntas, V.; Johns, E.; Tang, L.; Mikolajczyk, K. PN-Net: Conjoined triple deep network for learning local image descriptors. *arXiv* **2016**, arXiv:1601.05030.
40. Tian, Y.; Fan, B.; Wu, F. L2-Net: Deep learning of discriminative patch descriptor in euclidean space. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 661–669.
41. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superpoint: Self-supervised interest point detection and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 224–236.
42. Chen, S.; Chen, J.; Rao, Y.; Chen, X.; Fan, X. A Hierarchical Consensus Attention Network for Feature Matching of Remote Sensing Images. *ISPRS J. IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [[CrossRef](#)]
43. Zhu, H.; Jiao, L.; Ma, W.; Liu, F.; Zhao, W. A novel neural network for remote sensing image matching. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2853–2865. [[CrossRef](#)] [[PubMed](#)]
44. Wang, S.; Quan, D.; Liang, X.; Ning, M.; Guo, Y.; Jiao, L. A deep learning framework for remote sensing image registration. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 148–164. [[CrossRef](#)]
45. Ye, Y.; Tang, T.; Zhu, B.; Yang, C.; Li, B.; Hao, S. A multiscale framework with unsupervised learning for remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [[CrossRef](#)]
46. Zhang, Z.; Xu, Y.; Cui, Q.; Zhou, Q.; Ma, L. Unsupervised SAR and Optical Image Matching Using Siamese Domain Adaptation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [[CrossRef](#)]
47. Cui, S.; Ma, A.; Zhang, L.; Xu, M.; Zhong, Y. MAP-net: SAR and optical image matching via image-based convolutional network with attention mechanism and spatial pyramid aggregated pooling. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [[CrossRef](#)]
48. Xu, C.; Liu, C.; Li, H.; Ye, Z.; Sui, H.; Yang, W. Multiview Image Matching of Optical Satellite and UAV Based on a Joint Description Neural Network. *Remote Sens.* **2022**, *14*, 838. [[CrossRef](#)]
49. 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](#)]
50. Ma, J.; Zhao, J.; Tian, J.; Yuille, A.L.; Tu, Z. Robust point matching via vector field consensus. *IEEE Trans. Image Process.* **2014**, *23*, 1706–1721. [[CrossRef](#)] [[PubMed](#)]
51. Ramos, J.S.; Watanabe, C.Y.; Traina, C., Jr.; Traina, A.J. How to speed up outliers removal in image matching. *Pattern Recognit. Lett.* **2018**, *114*, 31–40. [[CrossRef](#)]
52. Brachmann, E.; Rother, C. Neural-guided RANSAC: Learning where to sample model hypotheses. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 4322–4331.
53. Cavalli, L.; Larsson, V.; Oswald, M.R.; Sattler, T.; Pollefeys, M. Adalam: Revisiting handcrafted outlier detection. *arXiv* **2020**, arXiv:2006.04250.