



# Article Matching Multi-Source Optical Satellite Imagery Exploiting a Multi-Stage Approach

Yuxuan Liu<sup>1</sup>, Fan Mo<sup>2</sup> and Pengjie Tao<sup>1,\*</sup>

- <sup>1</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; liuyx@whu.edu.cn
- <sup>2</sup> Satellite Surveying and Mapping Application Center, NASG, Beijing 100048, China; mof@sasmac.cn
- \* Correspondence: pjtao@whu.edu.cn

Received: 22 October 2017; Accepted: 30 November 2017; Published: 1 December 2017

Abstract: Geometric distortions and intensity differences always exist in multi-source optical satellite imagery, seriously reducing the similarity between images, making it difficult to obtain adequate, accurate, stable, and well-distributed matches for image registration. With the goal of solving these problems, an effective image matching method is presented in this study for multi-source optical satellite imagery. The proposed method includes three steps: feature extraction, initial matching, and matching propagation. Firstly, a uniform robust scale invariant feature transform (UR-SIFT) detector was used to extract adequate and well-distributed feature points. Secondly, initial matching was conducted based on the Euclidean distance to obtain a few correct matches and the initial projective transformation between the image pair. Finally, two matching strategies were used to propagate matches and produce more reliable matching results. By using the geometric relationship between the image pair, geometric correspondence matching found more matches than the initial UR-SIFT feature points. Further probability relaxation matching propagated some new matches around the initial UR-SIFT feature points. Comprehensive experiments on Chinese ZY3 and GaoFen (GF) satellite images revealed that the proposed algorithm performs well in terms of the number of correct matches, correct matching rate, spatial distribution, and matching accuracy, compared to the standard UR-SIFT and triangulation-based propagation method.

**Keywords:** multi-source optical satellite imagery; image matching; propagation matching; geometric correspondence matching; probability relaxation matching

# 1. Introduction

Image matching is the process of finding corresponding points on multi-view images of the same region; these images can be acquired by different sensors at different times [1]. Image matching is not only the precondition for automatic image registration [2,3], but also the foundation for three-dimensional (3D) modeling [4], change detection [5], image fusion [6], and environmental monitoring [7,8]. The need for multi-source-optical satellite imagery matching has quickly increased along with the increasing diversity of available optical satellite image data. For example, China launched the ZY3-02 satellite [9] on 30 May 2016, which is the second ZY3 series satellite. The first satellite was the ZY3-01 satellite [10], launched on 9 January 2012. With excellent performance, ZY3-01 plays an important role in stereo mapping. ZY3-02 uses a similar method as ZY3-01, but it includes some improvements, having a higher resolution for the optical cameras and adding a laser range finder to improve the vertical positional accuracy without ground control points. In addition, China initiated its high resolution earth observation program in 2010 [11]. This program plans to launch seven high resolution satellites, of which GF1, GF2, GF3, and GF4 satellites have been launched. GF1 and GF2 are optical remote sensing satellites. The panchromatic and multi-spectral resolution of GF1 are 2 m and 8 m, respectively, and those of GF2 are 1 m and 4 m, respectively. Determining how to fully use these optical satellite

images from a variety of remote sensors has become an imperative, and the matching of multi-source optical satellite images is one of the critical aspects in this field.

Matching corresponding points between multi-source optical satellite images is difficult due to the geometric deformation caused by scale change, rotation, and view angle transformations, as well as the nonlinear intensity differences caused by different radiation resolutions, and the effects of the atmosphere and radiation noise [12]. Image matching methods, based on local invariant features, are widely used for multi-source optical satellite images due to their robustness for image scale transformation. The most representative image matching method is the scale-invariant feature transform (SIFT) algorithm [13,14]. However, when the standard SIFT operator is applied to multi-source optical satellite image matching, the number of matched feature points is limited; a large percentage of mismatches occur and the correctly matched points are not uniformly distributed, resulting in match failure. To improve the performance of SIFT, researchers have reformed it mainly by: (1) improving the feature extraction operator [15–18]; (2) improving the feature descriptor [19–21]; and (3) improving the matching strategy [22–26]. For example, to improve the feature extraction ability in textureless areas, Sedaggat et al. [17] proposed the uniform robust SIFT (UR-SIFT) algorithm. According to the characteristics of the image scale pyramid, the expected number of feature points of each layer and octave was pre-defined, then the layer images were divided into grids, and the number of feature points retained in each grid was determined by the local image information entropy, feature point contrast, and the number of feature points in the grid. The strategy used to filter the points resulted in the retained feature points having high robustness and were evenly distributed.

To improve the feature descriptor, Ye et al. [21] introduced gradient direction information to SIFT, then the descriptor was constructed based on the gradient orientation restriction SIFT (GOR-SIFT) and shape context, which describes the edge information. The Euclidean distance and the chi-squared statistic  $(x^2)$  similarity measure were used to enhance the matching success rate. Ling et al. [23] presented a matching method integrating global Shuttle Radar Topography Mission (SRTM) and image segmentation. Before matching, the geometric epipolar lines were determined with the aid of SRTM, a few seed points were extracted by the Local Contrast (LC) operator, and the image was segmented, which was first conducted by the marker-based watershed algorithm. The image was further processed by Region Adjacency Graph (RAG), and good performance was achieved for the matching by combining the characteristics of distance, angle, and normalized correlation coefficient (DANCC). Triangulation-based methods, which are free of any other external data, have also been proven to be effective for image matching [24–26]. First, a few points are matched to build the initial triangles. Next, a certain number of feature points are detected within the corresponding triangles, and these feature points are matched under the triangle constraint. Then, the new matched points are inserted into the triangles, which are updated dynamically. Finally, the process is repeated until termination conditions are met, namely that the triangles are small enough or cannot match new points. For the matching of points in triangles, Geng [27] proposed the use of the nearest six matches around a candidate point to calculate the affine transformation in the local area for predicting the match. Then, the normalized correlation coefficients (NCCs) of the candidate point, with the points around the predicted location, were computed. If the NCC of the point with the highest value was larger than a pre-defined value, they were identified as a matching point pair.

Based on previous studies, this paper presents a multi-stage matching approach, called "EMMOSI" (Effective Matching of Multi-source Optical Satellite Imagery) for multi-source optical satellite images, which does not need any data other than the images themselves. The framework of the matching method is shown in Figure 1. Firstly, well-distributed feature points at each scale were extracted by the UR-SIFT operator. Next, the initial matching was conducted, and the transformation between multi-source images was computed. Finally, a new matching propagation method was used that includes the following two steps: (1) geometric correspondence matching is used to obtain some new correct matches; and (2) probabilistic relaxation matching was used to match the feature points that were not successfully matched in previous steps. The matches obtained in all steps were the

final matches. Experimental results on a variety of optical satellite images showed that the proposed method can largely improve the use of feature points and increase the number of correct matches, which were more evenly distributed.



Figure 1. The framework of Effective Matching of Multi-source Optical Satellite Imagery (EMMOSI).

#### 2. Methodology

#### 2.1. UR-SIFT Feature Extraction

UR-SIFT [17] is based on the SIFT algorithm; it improves the robustness and accuracy of feature points by controlling the number, quality, and distribution. The main goal of the algorithm is to retain a certain number of high-quality features across the entire range of the image. The specific steps are as follows:

(1) Determine the approximate total number of feature points. The expected total number of feature points is set as 0.4% of the number of image pixels, with an upper limit of 5000, but, when the image is small, these parameters can be appropriately scaled to ensure that sufficient number of feature points are obtained.

(2) Determine the number of feature points for each scale space image as required for the SIFT algorithm. The initial feature points are local extreme points and 10% of which with the lowest contrasts are usually discarded. To ensure that the retained feature points can be evenly distributed, the number of feature points for each scale is pre-defined according to the scale coefficient, which can be calculated by SIFT [13]. Assuming that the total number of feature points is N, the number of l layers in the o octave is:

$$N_{ol} = N \times F_{ol}$$

$$\sum_{o=1}^{ON} \sum_{i=1}^{LN} F_{ol} = 1$$
(1)

where  $F_{ol}$  is the proportion value of *l* layer of *o* octave.  $F_{ol}$  is negatively correlated with the scale coefficient  $SC_{ol}$ . If we set  $f_0 = F_{11}$  as the proportion parameter of the first layer of the first octave whose scale factor is  $SC_{11}$ , then the proportion parameter of the *l* layer of *o* octave can be calculated by Equation (2). In addition, the sum of all the proportion values is 1, so  $F_0$  can be computed, and the other proportion values are obtained.

$$F_{ol} = \frac{SC_{11}}{SC_{ol}} f_0 \tag{2}$$

(3) Determine the number of feature points in each grid of each layer image. To ensure that the feature points of each layer image are uniformly distributed, each layer image is divided into virtual grids. For a grid cell  $Grid_i$  whose index is k of the l layer of o octave, the number of feature points

 $N_Grid_k$  retained in the certain grid is computed by the information entropy  $E_k$  of the grid image, the number  $n_k$ , and the average contrast  $C_k$  of feature points located in the grid image.

$$N_Grid_k = N_{ol} \left[ \frac{W_E E_k}{\sum_i E_k} + \frac{W_n n_k}{\sum_i n_k} + \frac{(1 - W_E - W_n)C_k}{\sum_i C_k} \right]$$
(3)

where  $W_E$  and  $W_n$  are the weight factors of entropy and feature number, respectively.

(4) Choose the feature points to be retained. Assuming that the total number of feature points is  $N_{cell}$ , the detected feature points in each grid are ordered according to their contrasts, the feature points arranged in the top  $3 \times N_{cell}$  are retained, and the remainder discarded. Then, the locations of the feature points are determined to sub-pixel accuracies, and the edge effects are removed according to standard SIFT algorithm. The local entropies of the retaining feature points are then computed, and  $N_{cell}$  feature points with the highest entropies are retained.

(5) Based on the standard SIFT, the main intensity slope directions of the feature points are calculated, and the feature descriptors are constructed.

Figure 2 displays an example of the feature extraction comparison results of Standard SIFT and the UR-SIFT algorithm. The results reveal that UR-SIFT performs well with respect to the number and spatial distribution of extracted feature points.



**Figure 2.** Feature extraction: (a) ZY3-02 image: multispectral,  $338 \times 429$  pixels; (b) scale invariant feature transform (SIFT) feature extraction result; and (c) uniform robust SFIT (UR-SIFT) feature extraction result.

## 2.2. Initial Matching

Initial matching was conducted using the minimum Euclidean distance between descriptors. The ration R, between the distance to the closest neighbor and that of the second-closest, was used to eliminate the wrong matches. We set R = 0.6. For optical satellite images, many mismatches after constraint still existed, thus a cross-matching strategy was introduced. In cross matching, if  $p_i \in P$  in the reference image is matched to  $q_j \in Q$  in the input image, and  $q_j \in Q$  is successfully matched to  $p_i \in P$  conversely, then the point pair  $(p_i, q_j)$  is temporarily retained. The Random Sample Consensus (RANSAC) [28] algorithm was used to refine the obtained matches. The final refined matches were used to compute the transformation matrix H between the image pair. Since the terrain elevation variations were much smaller than the distance from the ground to the sensor, homography well describes the transformation between two satellite images. The homography relationship between a pair of corresponding satellite image feature points can be expressed by Equations (4) and (5).

$$\begin{cases} x_2 = \frac{x_1h_{11} + y_1h_{12} + h_{13}}{x_1h_{31} + y_1h_{32} + 1} \\ y_2 = \frac{x_1h_{21} + y_1h_{22} + h_{23}}{x_1h_{31} + y_1h_{32} + 1} \end{cases}$$
(4)

$$\begin{cases} x_1 = \frac{x_2 h'_{11} + y_2 h'_{12} + h'_{13}}{x_2 h'_{31} + y_2 h'_{32} + 1} \\ y_1 = \frac{x_2 h'_{21} + y_2 h'_{22} + h'_{23}}{x_2 h'_{31} + y_2 h'_{32} + 1} \end{cases}$$
(5)

where  $(x_1, y_1)$  is the coordinates of a point on the reference image,  $(x_2, y_2)$  is the coordinates of its corresponding point on the input image,  $h_{ij}$  represents the parameters of homography matrix from the reference image to the input image, and  $h'_{ij}$  represents the parameters of the inverse of homography matrix.

To evaluate the calculated homography matrix, we used the cost function proposed in Navy et al. [29], expressed by Equation (6):

$$C(h) = \frac{1}{N} \left( \sum_{i}^{N} d(h(p_i), p_j) + \sum_{j}^{N} d(h^{-1}(p_j), p_i) \right)$$
(6)

where *C* is the cost function value, *d* is the distance function between two spatial points, *N* is number of matches, and  $(p_i, p_j)$  is a pair of matching points.

#### 2.3. Propagation Matching

As the initial matches only occupy a small percentage of the whole feature points in multi-source optical satellite image matching, most of the feature points are wasted. Therefore, to fully use the remaining feature points, geometric correspondence matching and probability relaxation matching were used to complete the propagation process.

#### 2.3.1. Geometric Correspondence Matching

Initial matching only used the local information of the feature points, but did not take advantage of the geometric relationship between the image pair. Thus, we first used the geometric constraint to select a few candidate points, and further used the Modified NCC (M-NCC) method to conduct feature matching. The traditional NCC method, based on rectangle windows, is not invariant to rotation and scale changes. However, rotation and scale changes often exist in multi-source optical satellite images. Thus, we modified the traditional NCC method by warping the correlation windows to enable robustness in the presence of such geometric distortions. Specifically, we opened a rectangular window in the reference image and projected it to the input image, using the homography obtained in initial matching, which is likely to be an irregular quadrangle. Then, we resampled the irregular quadrangle to the same size as the rectangular window using bilinear interpolation. Finally, we calculated the correlation coefficient for the rectangle windows according to the standard NCC. The specific steps of geometric correspondence matching are as follows:

(1) The feature points on the reference image, which were not successfully matched in the initial matching, were first selected out. Since the transformation matrix H was estimated in initial matching, the corresponding locations on the input image were predicted using Equation (4). As shown in Figure 3, P is a seed point and P' is its corresponding point on the input image. Because H may contain some offset, the feature point on the input image, which is nearest to P', may not be the correct matching point, thus we chose to retain all the points whose distances to P' were smaller than n pixels. If points  $\{Q_1, Q_2, \ldots, Q_n\}$  existed on the input image, then we calculated their M-NCCs with P. After sampling the correlation window, the M-NCC value was calculated with Equation (7). If  $Q_j$  had the highest M-NCC value with P, and the value was larger than the threshold r, then, in reverse, we calculated its corresponding point  $Q'_j$  on the reference image using Equation (5). If the distance between P and  $Q'_j$  was smaller than n, and their M-NCC value was largest among all the feature points

on the reference image, whose distances to  $Q'_j$  were smaller than n, and the value was also larger than r, then P and  $Q_j$  were initially identified as a pair of matching points.

$$\rho(\mathbf{c},\mathbf{r}) = \frac{\sum_{i=1}^{w} \sum_{j=1}^{s} (g_{(i,j)} \times g_{(i+r,j+c)}) - \frac{1}{w \times s} (\sum_{i=1}^{w} \sum_{j=1}^{s} g_{(i,j)}) (\sum_{i=1}^{w} \sum_{j=1}^{s} g'_{(i+r,j+c)})}{\sqrt{\left[\sum_{i=1}^{w} \sum_{j=1}^{s} g_{(i,j)}^{2} - \frac{1}{w \times s} (\sum_{i=1}^{w} \sum_{j=1}^{s} g_{(i,j)})^{2}\right] \left[\sum_{i=1}^{w} \sum_{j=1}^{s} g'_{(i+r,j+c)}\right]^{2} - \frac{1}{w \times s} (\sum_{i=1}^{w} \sum_{j=1}^{s} g'_{(i+r,j+c)})^{2}}}$$
(7)

where *w* and *s* are the size of the matching window,  $g_{(i,j)}$  is the gray value at point (i, j) on the reference image, and  $\rho(c, r)$  is the correlation coefficient.



**Figure 3.** Geometric correspondence matching. *P* is a seed point, *P'* is its predicted point calculated by Equation (1),  $Q_j$  is the matched point, and the other points marked by a yellow circle are those points whose distance to *P* are within *n* pixels.

(2) Mismatches elimination: We re-calculated the homography matrix  $H_1$  by using the new matches and the initial matches. Then, the mismatches process was performed by using the global root mean square error (RMSE). If the RMSE was larger than  $r_1$  pixels, we eliminated the points with largest error one by one until the RMSE was smaller than  $r_1$  pixels. After that, the mean square error  $\sigma_x$  and  $\sigma_y$  in the horizontal and vertical directions were calculated. If a point's horizontal error was larger than  $E_x$ , or the vertical error was larger than  $E_y$ , then the point was deleted.

(3) Taking the retained matches and  $H_1$  as input, we repeated Steps 1 and 2 until the number of correct matches no longer changed. In general, the matches increased a little after several repetitions, but, given the time requirement involved, this paper used three as the maximum number of repetitions.

## 2.3.2. Probabilistic Relaxation Matching

Due to the complexity of the local texture structures of optical satellite images, homography cannot adequately express the transformation relationship between the images in some local areas, resulting in an offset of correct matches larger than n pixels. The M-NCCs for correct matches may be smaller than r, but we used strict values for n and r in geometric matching, resulting in the elimination of some correct matches. Additionally, the problem existed where the correctly matched point might have been close to the corresponding point but was not extracted in the feature extraction stage. Given these problems, we used the probability relaxation method [30,31] to match those feature points. Probability relaxation matching takes advantage of the constraints from surrounding matches to enhance robustness against geometric distortion, and uses the points around the predicted point as candidate points to increase the probability of finding the correct matching point, which greatly improves the matching success rate. The specific steps are as follows:

(1) Select the feature points which may be correctly matched, and finding their candidate matching points. If the feature points did not match successfully using geometric matching, but their transformation errors were less than  $n_1$  pixels, then we considered that these feature points may be correctly matched. The larger the  $n_1$ , the more correct matches obtained, but the time required to complete the matches was greater. Taking the 16 feature points around the corresponding point of a certain point *P*, their M-NCCs with *P* were calculated point by point. The candidate points were

selected by setting a threshold  $r_2$ . If the M-NCC value of one point was larger than  $r_1$ , then the point was defined as one candidate matching point. For example, if the initial point was  $I_i$ ,  $I_j$  represents its candidate matching point,  $t_i$  is the number of candidate matching points, and  $\rho_{ij}$  is their corresponding M-NCC. The initial probability that  $I_i$  is the correct matching point of  $I_i$  is:

$$P_{ij} = \rho_{ij} / \sum_{k=0}^{t_i} \rho_{ik} ; k = 0, \dots, t_i$$
(8)

(2) Calculate the compatibility between the point pair to be matched and a neighboring match. P(i, j) is the probability of match  $I_i \leftrightarrow I_j$ ,  $I_k$  is one of the matched points located in the neighborhood of  $I_i$ , and  $I_j$  is the corresponding point of  $I_i$ . To quantify the compatibility between the match  $I_i \leftrightarrow I_j$  and its neighbouring match  $I_k \leftrightarrow I_l$ , the following compatible coefficient function C(i, j; k, l) was introduced:

$$C(i,j;k,l) = \frac{T}{\exp\left[(\Delta p_x^2 + \Delta p_y^2)/\beta\right]} \Delta p_x = (x_j - x_i) - (x_l - x_k), \ \Delta p_y = (y_j - y_i) - (y_l - y_k)$$
(9)

where  $\Delta p_x$  is the displacement of  $I_i$  with its neighbouring point  $I_k$  in x direction, and  $\Delta p_y$  is the displacement in y direction. The larger is the  $\Delta p$ , the smaller is the compatibility. T and  $\beta$  are constant values.

(3) Calculate the match probability for each candidate point and find the correct matching point. The initial probability for each candidate was calculated with Equation (8). The probabilities P(i, j) are updated by the following rule:

$$P^{(o+1)}(i,j) = \frac{P^{(o)}(i,j)Q^{(o)}(i,j)}{\sum_{s=1}^{t_i} P^{(o)}(i,s)Q^{(o)}(i,s)}$$
(10)

$$Q^{(o)}(i,j) = \prod_{I_k \in \Omega(I_i)} P^{(o)}(i,j)C(i,j;k,l)$$
(11)

where  $\Omega(I_i)$  represents the matches around point  $I_i$ , which are determined by Euclidean distances, and o is the iteration number. The function  $Q^{(o)}(i, j)$  is the support that  $I_i \leftrightarrow I_j$  receives at the oth iteration from its neighboring match  $I_k \leftrightarrow I_l$ . The iteration process stops when one of the following two conditions is achieved: (a) the pre-defined number of iterations is reached, or (b) for all the candidate points, one of the match probabilities  $P(i, j)(j = 1, ..., t_i)$  exceeds  $1 - \partial$ , where  $\partial \ll 1$ . The match with the highest probability, which is larger than  $1 - \partial$ , is retained. If the pre-defined number has been reached and no probability meets the conditions, meaning that point  $I_i$  is a false match.

(4) Mismatches elimination: Although probabilistic relaxation matching makes full use of the information of neighboring matches, it cannot ensure that all the new matches are correct, thus we needed to perform error elimination. Since we needed to consider the exceptions in local areas, RMSE could not be used, so we also used the cross matching method here. Taking one matched point on the input image as an initial point, and corresponding points around it on the reference image as candidate points, we repeated Steps 1–3. If the located point was the same as on the reference image, then we retained it; otherwise, we deleted it.

#### 3. Experiments and Analysis

We compared EMMOSI with the standard UR-SIFT and the triangulation-based propagation method (TPM), which shares the same purpose as this paper. The concrete implementation process of TPM was as follows: we took the matching points from standard UR-SIFT as the input to create the initial triangulations, and then the centroids of each triangle in the reference were taken as the candidate points for the following matching. The reason for this was to improve the distribution uniformity of the propagation points. Finally, the method proposed in Geng [27] was used in the matching stage. All three methods were programmed by C++.

#### 3.1. Description of Experimental Datasets

To evaluate the performance of EMMOSI, two series of satellite images from ZY3 and GF were used. ZY3 experimental data included the forward-view and nadir-view images of the ZY3-01 satellite, whose ground sampling distances (GSDs) are 3.5 m and 2.1 m, respectively, and the nadir-view, backward-view, and multi spectral images of the ZY3-02 satellite, whose GSDs are 2.1 m, 2.1 m, and 5.8 m, respectively. GF experimental images included one GF1 panchromatic image whose GSD is 2.0 m, and GF2 panchromatic and multi spectral images, whose GSDs are 0.8 m and 3.2 m, respectively.

All input image pairs were selected from different satellite sensors or different imaging years. ZY3 satellite images covered the mountain topography and complex terrain with monotonous and intense textures, and the GF satellite images covered building areas. Image pairs 1–5 are ZY3 images, and image pairs 6 and 7 are GF images. To comprehensively evaluate the performance of the proposed algorithm under different conditions, we used the following seven imaging condition combinations shown in Table 1:

Table 1. Seven	different imaging	condition	combinations.

No.	Different Remote Sensors	Different Viewpoint Cameras	Different Resolutions	Different Spectral Modes	Different Acquisition Times
1					$\checkmark$
2	$\checkmark$				$\checkmark$
3	$\checkmark$	$\checkmark$			$\checkmark$
4	$\checkmark$	$\checkmark$	$\checkmark$		$\checkmark$
5	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
6	$\checkmark$		$\checkmark$		$\checkmark$
7	$\checkmark$		$\checkmark$	$\checkmark$	<u></u>

Large illumination differences and scene changes existed. In addition, local geometric distortions and relief displacement existed in all image pairs. Their specific parameters are shown in Table 2, and the experimental images are displayed in Figure 4.

No.	Figure	Satellite	View View/Spectral Mode	Image Size	GSD (m)	Acquisition Date	Location
1	4a 4b	ZY3-01 ZY3-01	Nadir-View Nadir-View	$\begin{array}{c} 802\times720\\ 803\times718 \end{array}$	2.1 2.1	5 October 2013 21 January 2014	China-Guangzhou
2	4c 4e	ZY3-01 ZY3-02	Nadir-View Nadir-View	932 ×1009 932 ×1027	2.1 2.1	17 May 2016 5 June 2016	China-Fuxin
3	4d 4f	ZY3-01 ZY3-02	Forward-View Backward-View	$\begin{array}{c} 600 \times 603 \\ 856 \times 828 \end{array}$	3.5 3.5	17 May 2016 5 June 2016	China-Fuxin
4	4d 4e	ZY3-01 ZY3-02	Forward-View Nadir-View	$\begin{array}{c} 600\times 603\\ 932\times 1027\end{array}$	3.5 2.1	17 May 2016 5 June 2016	China-Fuxin
5	4e 4g	ZY3-01 ZY3-02	Nadir-View Multi Spectral	$\begin{array}{c} 932 \times 1009 \\ 344 \times 374 \end{array}$	2.1 5.8	17 May 2016 5 June 2016	China-Fuxin
6	4h 4i	GF1 GF2	Panchromatic Panchromatic	$\begin{array}{c} 344 \times 314 \\ 841 \times 766 \end{array}$	2.0 0.8	4 November 2016 30 April 2017	China-Guangzhou
7	4h 4j	GF1 GF2	Panchromatic Multi Spectral	$\begin{array}{c} 344 \times 314 \\ 212 \times 193 \end{array}$	2.0 3.2	4 November 2016 30 April 2017	China-Guangzhou

Table 2. Experimental image pairs.



Figure 4. Cont.



**Figure 4.** Experimental satellite images. (**a**–**c**): nadir-view image of ZY3-01; (**d**) forward-view image of ZY3-01; (**e**) nadir-view image of ZY3-02; (**f**) backward-view image of ZY3-02; (**g**) multi-spectral image of ZY3-02; (**h**) panchromatic image of GF1; (**i**) panchromatic image of GF2; and (**j**) multi-spectral image of GF2.

## 3.2. Setting of Parameters

## 3.2.1. Parameters in UR-SIFT Feature Extraction

The octave number of UR-SIFT is self-adaptive to the image size, and the number of layers was set as 3. The scale layer gridding was set to approximately  $100 \times 100$  pixels, and the values of  $W_e$  and  $W_n$  were set to 0.2 and 0.5, respectively, according to Sedaghat et al. [17].

#### 3.2.2. Parameters in Geometric Correspondence Matching

To find the proper value of *n* and *r* in the first step, different values were assigned to *n*, changing from 0.5 to 3 in increments of 0.5, while the different values assigned to *r* changed from 0.6 to 0.9 in increments of 0.1. Many experimental results show that when n = 1 and r = 0.8, the value of the cost function of the final homography is always within one pixel and the matching points obtained were considered as correct matches. In the second step,  $r_1$  was set as one, which is reliable for eliminating the wrong matches [17].  $E_x$  and  $E_y$  were set as  $3\sigma_x$  and  $3\sigma_y$ , respectively. The error was Gaussian distribution, and the accuracy of a correct match was around 0.3–0.4 pixels. If the horizontal or vertical error of a match was larger than  $3\sigma_x$  and  $3\sigma_y$ , it was probably a mismatch. This may have eliminated a few correct matches, but it ensured the high accuracy of the retained matches.

#### 3.2.3. Parameters in Probabilistic Relaxation Matching

 $n_1$  and  $r_2$  were used to select candidate match points. The larger the  $n_1$  and the smaller the  $r_2$ , the more candidate match points were found and the more matches were obtained, which may have included some wrong matches. Many experiments revealed that, when  $n_1 = 2$  and  $r_2 = 0.7$ , a good balance between the number and accuracy of the matches is achieved. *T* and  $\beta$  were set as 1000 and 10 according to Zheng et al. [31].

### 3.2.4. Evaluation Criteria and Implementation Details

The matching quality was evaluated from the following four factors: the number of correct matches, correct matching rate, distribution quality, and matching accuracy.

(1) The number of correct matches: For all the matches, if the residual error of one point was less than 1.2 pixels, it was taken as a correct match, otherwise it was defined as a false match [23].

(2) Correct matching ratio: Correct matching ratio was the proportion of correct matches compared to the total matches.

(3) Distribution quality: To evaluate the distribution quality of matched points, the statistical distribution quality factor  $S_{cat}$  proposed in Goncalves et al. [32] was used. The smaller is the  $S_{cat}$ , the more uniform is the distribution.  $S_{cat}$  evaluates the distribution of feature points by counting the distance between each match and other matches, and storing the intermediate value, which was used to evaluate distribution quality. If this value was larger than 95% of the mean of the image dimensions, good distribution was achieved.

(4) Matching accuracy: In general, matching accuracy assessment requires ground truth. For image matching, manually measured correspondences are always used as references. However, measuring large numbers of correspondences manually, and the measured accuracy, depends on the accuracy of human observation, which is subjective and may vary from person to person. Thus, we required more practical methods. In this paper, the matching accuracy assessment method, based on RMSE of adjustment, was used. Using the matches as tie points between images, we conducted free network adjustment for the image pair, and the RMSE of the adjustment was taken as the image matching accuracy. For satellite images, the imaging model used was the Rational Function Model (RFM). We performed free network adjustment based on RFM by using the matches, then each point had a residual error after adjustment, upon which we calculated the RMSE. Theoretically, RMSE consists of the following three parts: image matching error, imaging model error, and image internal geometric distortion. The image matching error is the chief component, whereas the other two are secondary components, which can be ignored for satellite images with good geometric quality. Thus, RMSE can objectively reflect the image matching accuracy, although the actual positional accuracy of matched points was better than RMSE.

# 3.2.5. Comparative Results and Analysis

Table 3 displays the number of extracted features, number of correct matches, correct matching rate, distribution quality measure  $S_{cat}$ , and matching accuracy measured by RMSE for UR-SIFT, TPM, and EMMOSI.

No.	Algorithm -	Extracted Features				ŝ	
		Reference Image	Input Image	Correct Matches	Correct Matching Kate	Scat	KMSE/Pixel
	UR-SIFT	4617	4630	147	100.0%	0.1117	0.314
1	TPM	4617	4630	268	98.5%	0.0354	0.389
	EMMOSI	4617	4630	887	100.0%	0.0257	0.328
	UR-SIFT	6548	6373	156	100.0%	0.0467	0.266
2	TPM	6548	6373	272	97.1%	0.0541	0.448
	EMMOSI	6548	6373	696	99.1%	0.0236	0.282
3	UR-SIFT	5167	6131	49	100.0%	0.6680	0.258
	TPM	5167	6131	74	92.5%	0.7430	0.399
	EMMOSI	5167	6131	289	97.0%	0.5451	0.318
4	UR-SIFT	5167	6373	60	100.0%	0.1129	0.225
	TPM	5167	6373	105	96.3%	0.6389	0.395
	EMMOSI	5167	6373	302	97.4%	0.0806	0.320
	UR-SIFT	6548	2067	36	94.7%	0.6780	0.343
5	TPM	6548	2067	42	79.2%	0.8062	0.363
	EMMOSI	6548	2067	80	83.5%	0.0282	0.365
6	UR-SIFT	1553	6419	28	100.0%	0.4094	0.170
	TPM	1553	6419	41	100.0%	0.4908	0.332
	EMMOSI	1553	6149	77	87.5%	0.3456	0.305
7	UR-SIFT	1553	643	26	96.3%	0.0985	0.311
	TPM	1553	643	35	88.6%	0.6443	0.358
	EMMOSI	1553	643	59	89.4%	0.0206	0.304

**Table 3.** Experimental results of uniform robust scale invariant feature transform (UR-SIFT), triangulation-based propagation method, TPM and the proposed method, Effective Matching of Multi-Source Optical Satellite Imagery (EMMOSI).

From the results shown in Table 3, despite the large number of extracted features, the standard UR-SIFT achieved few matches, because it only uses the feature descriptors, which are computed from a small local area. Although TPM can increase the number of correct matches, the increased rate is

limited. In contrast, the proposed method significantly increased the number of matches with the aid of geometric correspondence between images and surrounding matched points, several orders of magnitude above the standard UR-SIFT. The matches of UR-SIFT were basically correct, and EMMOSI also had a high correct matching rate, which was higher than TPM in most cases.

Typically, the initial matches from UR-SIFT are unevenly distributed; they are mainly concentrated in regions with rich texture, whereas few matches exist in low texture areas. From the value of  $S_{cat}$ of TPM, TPM reduces rather than increases the uniformity of feature. The matching results of TPM are strongly dependent on the distribution of the initial feature points; the new inserted points were all in the regions of the initial triangulations. If the initial points are not well-distributed, TPM will not improve the uniformity of matches. Conversely, the value of  $S_{cat}$  of EMMOSI was much lower compared with standard UR-SIFT, indicating that a uniform distribution of the matches of EMMOSI was achieved.

The RMSE values reveal that the matching results from UR-SIFT are accurate. UR-SIFT matches can be reliably taken as initial points for TPM. The matching accuracies of TPM were lower than UR-SIFT because TPM mainly relies on accurate key points. If a few new inserted points are false, these wrong matches will seriously influence the final matching results. The result from the proposed method proves that EMMOSI can always achieve high accuracy, higher than that achieved by TPM. In addition to RMSE, Figure 5 shows the residual error histogram of the correct matches of EMMOSI for all the test image pairs. The residual errors were all within 1.2 pixels, and most of the values were within 0.8 pixel, which also proves the acceptable matching accuracy of the proposed algorithm.

Figure 6 displays the matching results of the first pair that represents the mountainous topography, on two ZY3-01 satellite images acquired in different seasons, so the land cover changed considerably. In Figure 6, the UR-SIFT matches have a good distribution, because the images have good texture in mountainous areas, but the matches are relatively sparse. TPM added a few matches in the first frame selection area, but did not add any match in the second frame selection compared to UR-SIFT. Initial triangulations are present in the second selection area, but the reason for no new matches is that the selected candidate feature points in these triangulations were not matched. However, regardless of the standard used to select the candidate matching points, the failure of matching some of the selected points will occur, and a few points that do not meet the specific criteria may match successfully. By using the geometry relationship between the image pair, EMMOSI found more matches than UR-SIFT in initial UR-SIFT feature points, and the probability relaxation matching propagated some new matches. As a result, EMMOSI achieved more evenly distributed matches than UR-SIFT and TPM, indicating that the proposed method can be applied to mountainous areas.

Figure 7 displays the matching result for the fourth image pair, which contains one forward-view image from the ZY3-01 satellite and one nadir-view image from the ZY3-02 satellite, covering complex terrain, and the two frame selection areas have less texture. In Figure 7, UR-SIFT failed to match in both selection areas, because the feature points are not distinct and cannot be matched when relying on Euclidean distance in textureless areas. TPM failed to match in the first selection area and only achieved a few matches in the second selection area. The reason for the matching failure for the first selection area is that there was no initial triangulation, thus no candidate feature points were selected for matching in that area. In contrast, EMMOSI found many matches in both selection areas because our method effectively uses UR-SIFT feature points by geometric correspondence matching, and increases some matched points located in the surrounding areas of the matched UR-SIFT points using probability relaxation matching. Overall, the matches of UR-SIFT and TPM were distributed unevenly. TPM only added a small number of new matches to UR-SIFT, and EMMOSIEMMOSI achieved many more matches than TPM and these matches were evenly distributed across the whole image, indicating that the proposed method can be applied to complex terrain, especially in textureless regions.



**Figure 5.** Residual error histogram of correct matches of EMMOSI in all test image pairs: (a) Nadir-view image of ZY3-01 and Nadir-view image of ZY3-01; (b) Nadir-view image of ZY3-01 and Nadir-view image of ZY3-02; (c) forward-view image of ZY3-01 and backward-view image of ZY3-02; (d) forward-view image of ZY3-01 and Nadir-view image of ZY3-02; (e) Nadir-view image of ZY3-01 and multi-spectral image of ZY3-02; (f) panchromatic image of GF1 and panchromatic image of GF2; and (g) panchromatic image of GF1 and multi-spectral image of GF2.



**Figure 6.** Matching results for the first experimental image pair: (**Left**) Nadir-view image of ZY3-01; and (**Right**) Nadir-view image of ZY3-01.

Figure 8 displays the matching result of the seventh image pair, including one GF1 panchromatic image and one GF2 multi spectral image, which cover an urban area. UR-SIFT and TPM matched few points in the selected areas, while EMMOSI found many matches. Overall, the EMMOSI had more matches than UR-SIFT and TPM. Additionally, the EMMOSI matches were evenly distributed throughout the image, indicating that the proposed method can be applied to urban areas.



**Figure 7.** Matching results for the fourth experimental image pair: (Left) forward-view image of ZY3-01; and (**Right**) Nadir-view image of ZY3-02.



**Figure 8.** Matching results for the seventh experimental image pair in an urban area: (**Left**) panchromatic image of GF1; and (**Right**) multi spectral image of ZY3-02.

#### 4. Discussion

In this study, we present an effective method, EMMOSI, for multi-source optical satellite imagery matching, using seven image pairs with different imaging combinations to evaluate its performance. Two widely used state-of-the-art methods were compared and analyzed based on number of matches, correct matching rate, spatial distribution, and accuracy of matched points. As the experimental results indicated, EMMOSI is superior to UR-SIFT and TPM for number and distribution of correct matches.

Although UR-SIFT can independently extract adequate and well-distributed feature points regardless of the image texture being dense or sparse, it only achieves few matches by using the Euclidean distance between descriptors. These matches are too few to meet the requirements for image registration. TPM uses the UR-SIFT matches as seed points to build initial triangulations, finds new feature points in the initial triangulations, and matches these new points under the triangulation constraints. UR-SIFT matches are not always well-distributed and the newly found feature points

are all within the initial triangulations, losing the well-distributed property of the initial UR-SIFT feature points, and adds extra feature extraction work. Then, the new matches are inserted into the triangulations, which are dynamically updated. However, we cannot ensure that all the new matches are correct, and if some false matches are inserted, the following building process of triangulations is considerably influenced, and may cause more false matches.

To acquire sufficient matches with good distribution, we first used the UR-SIFT operator to extract adequate and well-distributed feature points, laying a good foundation for the following feature matching stage. In initial matching, we obtained a few matches and the transformation matrix between the image pair. The value of the cost function of the transformation matrix revealed that the matrix was reliable. To make full use of the UR-SIFT feature points, we first conducted geometric correspondence matching basing on the transformation matrix, which uses the geometric relationship between the image pair, rather than local image information. This process constrains the corresponding points to a small searching area, so it can find more matches and these matches are well-distributed with high reliability. Since many matches were obtained, we could find some matches around the unmatched feature points from previous steps, which is the precondition of probabilistic relaxation matching. Probabilistic relaxation matching uses spatial distance constraints to conduct matching, which is robust for local image deformation.

To ensure a high accuracy, several constraints were implemented during the whole matching flow. In the initial matching process, a strict value for NNDR, cross matching, and RANSAC were used to ensure that the initial matches were correct. Then, in the geometric correspondence matching process, we used RMSE in global, x and y directions to eliminate the wrong matches, and, in the probabilistic relaxation matching stage, the searching area was limited to a small neighboring area, and cross matching was used to refine the matches.

Our experimental image pairs covered diverse terrain, including mountainous topography, complex terrain with intense and textureless regions, and urban areas. EMMOSI performed well for all topographies. In summary, EMMOSI is effective and stable for multi-source optical satellite imagery matching.

## 5. Conclusions

In this paper, EMMOSI, a multi-stage matching approach, is proposed for multi-source optical satellite imagery. The proposed algorithm includes the following steps: UR-SIFT feature extraction, initial matching, geometric correspondence matching, and probabilistic relaxation matching. The main contribution is the proposed matching propagation strategy. Geometric correspondence matching uses the geometric transformation information between the image pair, which finds the matches that satisfy the geometric transformation. Probabilistic relaxation matching uses spatial distance constraints from surrounding correct matches to find matches in local, distorted, and weak texture areas. Comprehensive experiments on images derived from different sensors with different viewing directions, geometric resolutions, spectral mode, and times proved that our method increased the number of correct matches and the accuracy and improved their spatial distribution.

The proposed method can be applied to a variety of remote sensing applications that require feature point matching, such as image registration and image mosaicking. In addition, the proposed propagation method can be added to other matching methods, which will broaden its application. For future work, a wider range of optical satellite images will be applied to evaluate the robustness of the proposed method, and we will try to improve the algorithm efficiency.

**Acknowledgments:** This paper was substantially supported by the National Key R&D Program of China (Grant No. 2017YFB0503004), the National Natural Science Foundation of China (Project Nos. 41301525 and 41571440), the Surveying and Mapping Public Welfare Project of China (No. 201512012), and the High Resolution Remote Sensing, surveying and mapping Application Demonstration System Research Program (Issue No. 1). We owe great appreciation to the anonymous reviewers for their critical, helpful and constructive comments and suggestions.

**Author Contributions:** Yuxuan Liu, Fan Mo and Pengjie Tao conceived and designed the experiments; Yuxuan Liu performed the experiments and analyzed the data; Fan Mo contributed experimental images; Pengjie Tao contributed analysis tools; and Yuxuan Liu and Pengjie Tao wrote the paper.

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

# References

- 1. Gruen, A. Development and status of image matching in photogrammetry. *Photogramm. Record* **2012**, 27, 36–57. [CrossRef]
- Yan, L.; Roy, D.; Zhang, H.; Li, J.; Huang, H. An automated approach for sub-pixel registration of Landsat-8 operational land imager (OLI) and sentinel-2 multi spectral instrument (MSI) imagery. *Remote Sens.* 2016, *8*, 520. [CrossRef]
- 3. Gianinetto, M. Automatic co-registration of satellite time series. *Photogramm. Record* 2012, 27, 462–470. [CrossRef]
- 4. Bulatov, D.; Wernerus, P.; Heipke, C. Multi-view dense matching supported by triangular meshes. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 907–918. [CrossRef]
- 5. Du, P.; Liu, S.; Liu, P.; Tan, K.; Cheng, L. Sub-pixel change detection for urban land-cover analysis via multi-temporal remote sensing images. *Geo-Spat. Inf. Sci.* **2014**, 17, 26–38. [CrossRef]
- 6. Ghassemian, H. A review of remote sensing image fusion methods. *Inf. Fusion* **2016**, *32*, 75–89. [CrossRef]
- 7. Stumpf, A.; Malet, J.P.; Delacourt, C. Correlation of satellite image time-series for the detection and monitoring of slow-moving landslides. *Remote Sens. Environ.* **2017**, *189*, 40–55. [CrossRef]
- 8. Richter, R. Modelling and monitoring urban built environment via multi-source integrated and fused remote sensing data. *Int. J. Image Data Fusion* **2013**, *4*, 2–32.
- 9. Deng, J.; Yang, Y. ZY-3-02 Camera Achieved Higher Stereo Imaging Accuracy. Aerosp. China 2016, 17, 62–63.
- 10. Li, D.R. China's first civilian three-line-array stereo mapping satellite: ZY-3. *Acta Geod. Cartogr. Sin.* **2012**, *41*, 317–322.
- 11. Li, D.R.; Tong, Q.X.; Li, R.X.; Zhang, L.P. Current issues in high-resolution earth observation technology. *Sci. China Earth Sci.* **2012**, *55*, 1043–1051. [CrossRef]
- 12. Scheffler, D.; Hollstein, A.; Diedrich, H.; Segl, K.; Hostert, P. AROSICS: An automated and robust open-source image co-registration software for multi-sensor satellite data. *Remote Sens.* **2017**, *9*, 676. [CrossRef]
- 13. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
- Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* 2005, 27, 1615–1630. [CrossRef] [PubMed]
- 15. Castillo-Carrión, S.; Guerrero-Ginel, J.E. Sift optimization and automation for matching images from multiple temporal sources. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *57*, 113–122. [CrossRef]
- 16. Lingua, A.; Marenchino, D.; Nex, F. Performance analysis of the sift operator for automatic feature extraction and matching in photogrammetric applications. *Sensors* **2009**, *9*, 3745–3766. [CrossRef] [PubMed]
- 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. Paul, S.; Pati, U.C. Remote sensing optical image registration using modified uniform robust sift. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1–5. [CrossRef]
- 19. Sedaghat, A.; Ebadi, H. Distinctive order based self-similarity descriptor for multi-sensor remote sensing image matching. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 62–71. [CrossRef]
- 20. Ye, Y.; Shan, J. A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences. *ISPRS J. Photogramm. Remote Sens.* **2014**, *90*, 83–95. [CrossRef]
- 21. Ye, Y.; Shan, J.; Xiong, J. A matching method combining SIFT and edge information for multi-source Remote Sensing Images. *Geomat. Inf. Sci. Wuhan Univ.* **2013**, *10*, 1148–1151.
- 22. Gu, Y.; Ren, K.; Wang, P.; Gu, G. Polynomial fitting-based shape matching algorithm for multi-sensors remote sensing images. *Infrared Phys. Technol.* **2016**, *76*, 386–392. [CrossRef]
- 23. Ling, X.; Zhang, Y.; Xiong, J.; Huang, X.; Chen, Z. An image matching algorithm integrating global SRTM and image segmentation for multi-source satellite imagery. *Remote Sens.* **2016**, *8*, 672. [CrossRef]

- 24. Wu, B.; Zhang, Y.; Zhu, Q. Integrated point and edge matching on poor textural images constrained by self-adaptive triangulations. *ISPRS J. Photogramm. Remote Sens.* **2012**, *68*, 40–55. [CrossRef]
- 25. Zhu, Q.; Wu, B.; Tian, Y. Propagation strategies for stereo image matching based on the dynamic triangle constraint. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 295–308. [CrossRef]
- 26. Wu, B.; Zhang, Y.; Zhu, Q. A triangulation-based hierarchical image matching method for wide-baseline images. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 695–708. [CrossRef]
- 27. Geng, X. Research on photogrammetric processing for mars topographic mapping. *Acta Geod. Cartogr. Sin.* **2015**, *44*, 944.
- 28. 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]
- 29. Navy, P.; Page, V.; Grandchamp, E.; Desachy, J. Matching two clusters of points extracted from satellite images. *Pattern Recognit. Lett.* **2006**, *27*, 268–274. [CrossRef]
- 30. Deng, W.; Zou, H.; Guo, F.; Lei, L.; Zhou, S. Point-pattern matching based on point pair local topology and probabilistic relaxation labeling. *Vis. Comput.* **2016**, 1–11. [CrossRef]
- 31. Zheng, S.; Zhang, Z.; Zhang, J. Image relaxation matching based on feature points for DSM generation. *Geo-Spat. Inf. Sci.* 2004, 7, 243–248.
- 32. Goncalves, H.; Goncalves, J.A.; Corte-Real, L. Measures for an objective evaluation of the geometric correction process quality. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 292–296. [CrossRef]



© 2017 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 (http://creativecommons.org/licenses/by/4.0/).