Article

Automatic Registration Method for Fusion of ZY-1-02C Satellite Images

Qi Chen, Shugen Wang, Bo Wang and Mingwei Sun *

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China;
E-Mails: daoqiqi@gmail.com (Q.C.); wangsg@whu.edu.cn (S.W.); bo.wang.rs@gmail.com (B.W.)

* Author to whom correspondence should be addressed; E-Mail: mingweis@whu.edu.cn;
Tel.: +86-13607144988; Fax: +86-27-6877-8010.

Abstract: Automatic image registration (AIR) has been widely studied in the fields of medical imaging, computer vision, and remote sensing. In various cases, such as image fusion, high registration accuracy should be achieved to meet application requirements. For satellite images, the large image size and unstable positioning accuracy resulting from the limited manufacturing technology of charge-coupled device, focal plane distortion, and unrecorded spacecraft jitter lead to difficulty in obtaining agreeable corresponding points for registration using only area-based matching or feature-based matching. In this situation, a coarse-to-fine matching strategy integrating two types of algorithms is proven feasible and effective. In this paper, an AIR method for application to the fusion of ZY-1-02C satellite imagery is proposed. First, the images are geometrically corrected. Coarse matching, based on scale invariant feature transform, is performed for the subsampled corrected images, and a rough global estimation is made with the matching results. Harris feature points are then extracted, and the coordinates of the corresponding points are calculated according to the global estimation results. Precise matching is conducted, based on normalized cross correlation and least squares matching. As complex image distortion cannot be precisely estimated, a local estimation using the structure of triangulated irregular network is applied to eliminate the false matches. Finally, image resampling is conducted, based on local affine transformation, to achieve high-precision registration. Experiments with ZY-1-02C datasets demonstrate that the accuracy of the proposed method meets the requirements of fusion application, and its efficiency is also suitable for the commercial operation of the automatic satellite data process system.

1. Introduction

Image registration is a fundamental image processing technique in the areas of medical imaging, computer vision, and remote sensing (RS). In the field of RS, automatic image registration (AIR) has been a widely studied problem for decades. AIR has been achieved in many different datasets, including multi-platform [1,2], multi-modal [3–6], multi-scale [7,8], and multi-angle [2,9] RS images. Considerable accuracy and efficiency have been obtained. A typical registration method can be divided into three steps: image matching, transformation model estimation, and image resampling.

Image matching is a technique that identifies corresponding structures such as point, line, and surface through certain criteria. The algorithms of this technique can be broadly classified into two categories: area-based matching (ABM) and feature-based matching (FBM). ABM calculates a certain measurement using gray data in the fixed-size window from two images and treats the center points of windows as corresponding points when the measured value exceeds the threshold. The most commonly used algorithms are methods based on mutual information [10–12], frequency-domain correlation [13,14], and cross correlation [4,8,9,15,16]. The main disadvantage of ABM lies in the excessive computing time spent on searching for the corresponding point; a matching strategy based on the image pyramid is usually helpful in accelerating the processing speed [4,15]. In FBM algorithms, salient features are extracted from different images, and then corresponding features are determined by comparing the similarities of their descriptions. Matching methods based on point feature, such as scale invariant feature transform (SIFT) [17] and speeded up robust features (SURF) [18], are widely used in RS. Other methods based on features, including lines [7], edges [2,19], contours [20], and shapes [21], are also used in numerous applications. FBM based on SIFT is proven to have excellent performance [22] and has been successfully applied to various AIR cases of RS images [3,23–26]. However, for extremely large RS images, the direct use of SIFT-based matching faces difficulties in obtaining evenly distributed corresponding points. Another problem to consider is excessive memory consumption. To mitigate this problem, several researchers conducted SIFT matching based on image segmentation [27–29], but this approach always results in lower computational efficiency. To overcome these limitations, a coarse-to-fine matching strategy that integrates the two types of algorithms has been proposed [8,9,16]. In this strategy, FBM is first performed to accomplish coarse matching, and then ABM, as a process of precise matching, used the results of coarse matching to obtain the appropriate corresponding points. High accuracy and efficiency can be achieved through this strategy, particularly when processing extremely large RS images.

The second step of image registration is to estimate the image transformation model with the results of image matching. The affine and polynomial transformation models are commonly adopted for RS image registration. However, under certain situations, such as when the image size is extremely large or when complex image distortion exists, the aforementioned two models can make only a rough estimation of the transformation, but the fitting precision is insufficient and usually uneven on the

image. To implement high-precision registration of images with significant distortions, some local fitting methods, such as B-spline functions [11], thin-plate splines [9,21,30], and local affine transformation based on triangulated irregular network (TIN) [4,8,15,29] have been applied. The experiment results of Arévalo *et al.* [31] demonstrate that local methods outperform global methods. The presence of noise and limitations of the algorithm inevitably cause false matches in the image matching results. To address this issue, certain criteria should be adopted to reject the false matches. Comparison of the residuals solved by least squares is a commonly used method. In addition, numerous algorithms that focused on removing false matches have been proposed and widely used in image registration, such as random sample consensus [32], maximum likelihood estimation sample consensus [33], and maximum distance sample consensus [34].

The final process of registration, that is, image resampling, can be conducted using the estimated parameters of the transformation model to warp the input image to the reference image. To achieve sub-pixel accuracy, image interpolation methods, such as bilinear interpolation and cubic convolution interpolation, are usually employed for resampling [35].
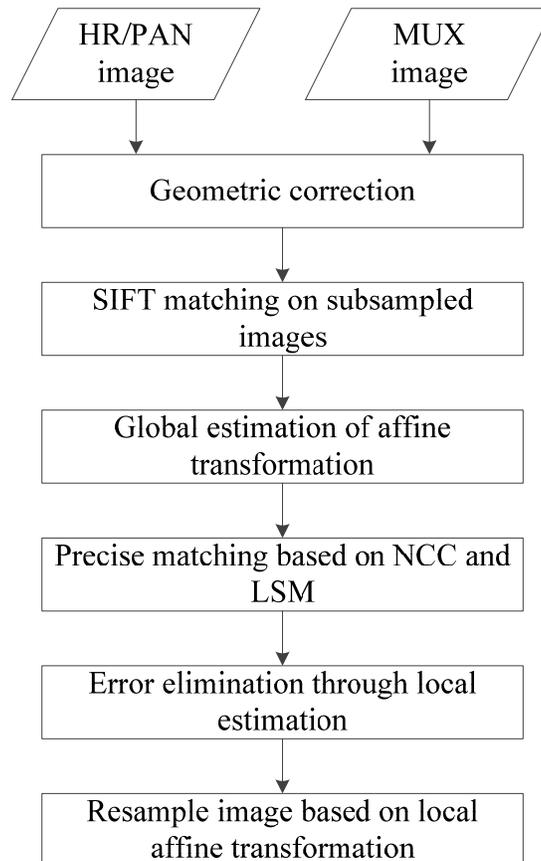
ZY-1-02C satellite, launched on 22 December 2011, is equipped with a multispectral (MUX) camera (10 m resolution, including infrared, red, and green bands), a panchromatic (PAN) camera (5 m resolution), and a panchromatic high-resolution (HR) camera (2.36 m resolution). As China's first satellite that is customized specifically for the land resource department, an automatic satellite data processing system (ASDPS) was established on land to ensure its commercial operation. In accordance with the ASDPS design, the PAN and MUX images captured from the same area are geometrically corrected first, and then image fusion processing is performed to obtain color infrared images with higher resolution. Theoretically, the geographical coordinates of the corrected images from different sensors should be extremely close, and the images can be fused directly. However, the positioning accuracy of different images is influenced by the limited manufacturing technology of the charge-coupled device, focal plane distortion, and unrecorded spacecraft jitter [36]. In our actual data processing, the geometrically corrected images of ZY-1-02C still exhibit a significant accuracy difference, which indicates that the image registration process should be conducted before image fusion.

An AIR method applied to the fusion of ZY-1-02C satellite imagery is proposed in this paper. First, the images are geometrically corrected with equivalent ground sampling distance (GSD), a coarse matching based on SIFT for the subsampled corrected images is performed, and a rough global estimation is made with the matching results. Feature points are then extracted using a Harris detector, and the coordinates of their corresponding points are calculated according to the results of the global estimation. Precise matching is conducted based on normalized cross correlation (NCC) and least squares matching (LSM). To fit the complex image distortion, which cannot be precisely estimated, a local estimation using the structure of TIN is applied to eliminate the false matches. Finally, using the optimized TIN generated in the process of error elimination, image resampling is conducted based on local affine transformation to achieve high-precision registration. Experiments with ZY-1-02C datasets demonstrate that the accuracy of the proposed method meets the requirements of fusion application, and its efficiency is also suitable for the commercial operation of ASDPS.

2. Methodology

The proposed approach can be divided into six steps and each step is specified in Figure 1.

Figure 1. Flowchart of the entire process.

2.1. Geometric Correction

Before the process of geometric correction, the long strip satellite images were segmented into scenes by ASDPS, and the rational polynomial coefficient parameters for each scene of the images were provided. Therefore, the rational function model (RFM) is adopted for geometric correction in our approach. As a result of the limited positioning accuracy of the satellite, the average elevation of the positioning range is used for image correction. SRTM-DEM [37], which is publicly available in our industry, is employed as the data sources of elevation.

In the RFM, image pixel coordinates are defined as the ratios of polynomials of ground coordinates [38]:

$$\begin{cases} r_n = \dfrac{P_1(X_n, Y_n, Z_n)}{P_2(X_n, Y_n, Z_n)} \\ c_n = \dfrac{P_3(X_n, Y_n, Z_n)}{P_4(X_n, Y_n, Z_n)} \end{cases} \tag{1}$$

where (r_n, c_n) and (X_n, Y_n, Z_n) are the normalized coordinates of the image space and ground space points, respectively. $P_i(i = 1, 2, 3, 4)$ are the cubic polynomials that represent the rigorous geometric sensor model of satellite imagery.

To eliminate the scale differences between images, the equivalent GSD is adopted for geometric correction so that precise matching can be implemented smoothly. Even after geometric correction, a significant accuracy difference remains between images of ZY-1-02C (see analysis in Section 3.1). In this situation, direct NCC matching results in a dramatic increase in time consumption and matching error. Therefore, a coarse matching for the subsampled corrected images is performed in advance. With the results of rough global estimation through coarse matching, the searching range of precise matching is sharply reduced, and precision and efficiency are evidently improved.

2.2. SIFT Matching on Subsampled Images

2.2.1. Image Subsampling

In our related work, the coarse-to-fine matching strategy is generally applied using a multi-level image pyramid created by methods that can preserve more image information, such as the wavelet-based pyramid approach [8,16]. However, in our approach, as the image sizes become excessively large, multi-level image pyramid and complex subsampling method evidently reduce efficiency. Thus, we subsample the images only once to perform coarse matching, and the interval sampling method is used to generate the subsampled images. An appropriate size of the subsampled images can be calculated using the following equation:

$$\begin{cases} interval = \max\{\text{ceil}(w / N), \text{ceil}(h / N)\} \\ w_{sub} = \text{floor}(w / interval) \\ h_{sub} = \text{floor}(h / interval) \end{cases} \tag{2}$$

where *interval* is the sampling interval; w and h are the width and height of the full-size image, respectively; w_{sub} and h_{sub} are the width and height of the subsampled image, respectively; and N is the maximum acceptable size of the subsampled image. ceil() takes the number and rounds it to the nearest integer above its current value, whereas floor() rounds it to the nearest integer below its current value. The width and height calculated by Equation (2) is extremely close to but less than N.

To avoid repeatable loading of the large image, interval sampling is conducted while saving the corrected images in the process of geometric correction. Aside from its use in coarse matching, the subsampled image can also be filed to ASDPS as a browse map of the corrected image.

2.2.2. SIFT Matching

SIFT matching is a highly successful FBM algorithm proposed by Lowe [17]. The corresponding points are determined by comparing their descriptors. Thus, region searching is not required in SIFT matching. After eliminating false matches, the matching results possess high reliability. Meanwhile, the consumption of memory and time also becomes acceptable because of the small size of the subsampled images.

The subsampled MUX image has to be transformed to grayscale before SIFT matching. To minimize the information losses, principal component analysis (PCA), which was first proposed by

Hotelling [39], is performed, and the first principal component is used for matching. Moreover, because the low contrast of the image may result in the filtering of most feature points extracted by SIFT, histogram equalization is used to process the grayscale image in advance. Wang *et al.* [28] proved that histogram equalization improves the contrast sensitivity of the images and significantly enhances the possibility of extracting and matching features. Thereafter, SIFT is performed on the processed subsampled images.

2.3. Global Estimation of Affine Transformation

A number of corresponding points can be obtained via SIFT matching. Multiplying the coordinates of these points by the value of *interval* determines the positions of these points on the full-size images. The image size of the subsampled images is significantly less than that of full-size ones; thus, the precision of coarse matching is limited. However, affine transformation remains capable of estimating the accuracy difference between images to a certain extent. In this study, we consider an HR/PAN image as the reference image with the MUX image as the input image. The parameters of affine transformation can be solved through a least squares method.

The matching results inevitably contain false matches, which can be eliminated by comparing the residual error of each point pair and the root mean square error (RMSE). The overall accuracy of the coarse matching points is not extremely high. Thus, point pairs with its residual error more than two times of RMSE are removed. The remaining matching points are used to calculate a new RMSE, and the iteration continues until the residual errors of all the remaining point pairs are less than twice the RMSE.

2.4. Precise Matching Based on NCC and LSM

Approximate parameters of image transformation are obtained through global estimation. With the parameters, the position of the corresponding points can be predicted, and the point should be searched out in a relatively small window around the position. Similar to the process of coarse matching, the full-size MUX image should be converted into grayscale through PCA before precise matching. In our approach, the process of precise matching is divided into two steps: feature extraction and matching based on NCC and LSM.

2.4.1. Feature Extraction

In the featureless and textureless area, determining a reliable matching result is difficult. This situation can be avoided by matching after feature extraction.

The Harris detector adopted in our approach is widely known for its high speed and stable result [40]. To extract an appropriate number of feature points that are evenly distributed, the image is segmented into grids with a certain grid size. The features with strongest Harris value are then extracted from each grid. Precision of the Harris detector can be achieved only at the pixel level, whereas that of the Förstner operator can be achieved at the sub-pixel level [41]. As a result, using the method proposed by Zhang *et al.* [42], we employ the Harris detector to extract more apparent features and then refine the feature point to the sub-pixel level using the Förstner operator. As the resolution of

an MUX image is lower than that of an HR/PAN image, some detailed information existing in an HR/PAN image cannot be found in an MUX image. Thus, we decide to extract feature points from an MUX image and then search for the corresponding points on an HR/PAN image.

2.4.2. Matching Based on NCC and LSM

After extracting feature points from MUX images, the coordinates of the corresponding points on HR/PAN images can be estimated through affine transformation. Around the estimated position, a large searching window need not be set to find the corresponding points, which ensures the high efficiency and reliability of matching results when an NCC-based approach is used.
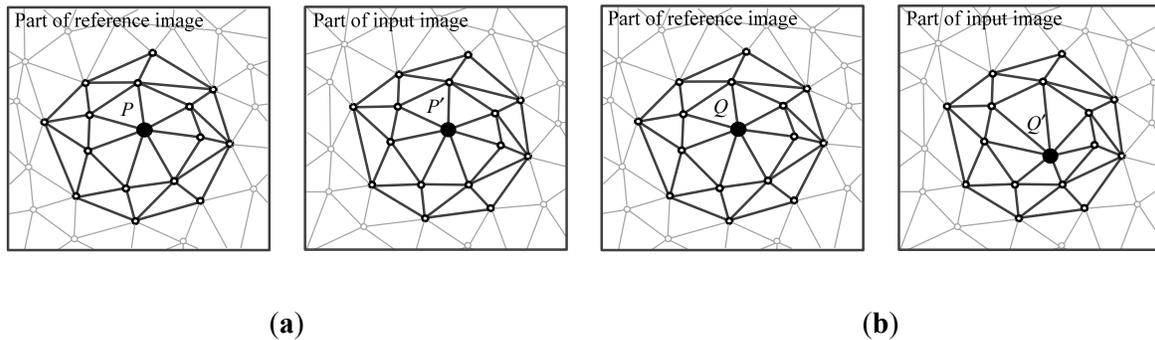
NCC matching can only achieve pixel-level precision. To achieve sub-pixel matching precision, the coordinates of corresponding points obtained by NCC are used as initial values to be refined using LSM. Proposed by Ackermann [43], LSM uses adequate information in the image window for adjustment calculation, which can improve precision at the sub-pixel level.

2.5. Error Elimination through Local Estimation

Numerous corresponding points can be obtained through precise image matching. False matches must be eliminated from the corresponding points before use in image registration. The complicated image distortion in PAN/HR and MUX images causes difficulty in employing a mathematical model for the direct and precise description of image transformation. The model cannot, therefore, serve as a criterion to eliminate false matches. However, influences of distortions can be minimized in most cases when the judging area is reduced to a relatively small one [44]. In this case, local image transformation can be approximated accurately by a simple model, such as affine transformation, and the false matches can be eliminated effectively.

Using a TIN structure, the adjacency between points can be determined simply and efficiently. Wang *et al.* [45,46] proposed a method of outlier detection using the TIN structure, which was successfully applied to automatic registration of terrestrial laser scanning point clouds. Similarly, in our approach, the TIN structure is used to calculate the adjacency between neighboring points, but our approach is based on affine invariance of local points while the Wang *et al.* [45,46] method is based on distance invariance. As shown in Figure 2, image distortion becomes subtle and regular when the judging area is limited within a small local part. In this situation, if estimation is conducted by using an affine transformation, the coordinate differences between correct matches can be fitted well with the transformation (Figure 2a), whereas residual errors of the false matches become evidently greater than those of the surrounding points (Figure 2b) and are then detected as gross errors to be eliminated.

The divide-and-conquer approach proposed by Lee *et al.* [47] is a practical and commonly used algorithm to construct a Delaunay triangulation. However, the TIN structure may contain very long and narrow triangles at the border of the structure. These triangles will undermine the local estimation and registration accuracy. They are detected by comparing the perimeter of each triangle with the mean perimeter of all triangles in the TIN structure. Triangles at the border with perimeters that are larger than the mean perimeter by 3 times the variance of perimeters are removed iteratively.

Figure 2. Analysis of false match identification. (**a**) Correct match. (**b**) False match.

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

Based on the preceding analysis, the proposed gross error elimination method is demonstrated as follows:

(1) A TIN is constructed using the coordinates of matching points, and the points in TIN are judged one by one in the following steps;

(2) Several nearest neighboring points around the current judging point are collected based on the TIN structure. The neighboring points are determined by an iterative method: first, all the points adjacently connected to the judging point are collected; then, more points that are adjacently connected to the collected points are found and gathered continually. In our approach, the recommended number of iteration times is 2, as shown in Figure 2;

(3) Based on the coordinates of the collected matching points, affine transformation parameters of the local distortion can be estimated;

(4) The residual error of the judging point is calculated using the affine parameters obtained previously. If the error is greater than a certain threshold (which is twice the RMSE in our approach), the judging point and its corresponding point are eliminated as a false match. Otherwise, we go to step (2) to judge the next point;

(5) After traversing all the points in the TIN, we return to step (1) to reconstruct a new TIN using the remaining points. The process continues until the residual errors of all points meet the requirements.

2.6. Resample Image Based on Local Affine Transformation

The accuracy of the geometrically corrected image of ZY-1-02C is influenced by numerous factors. Thus, transformation models, such as affine and geometric polynomials, cannot describe the distortions precisely. However, in the local areas of images, affine transformation is sufficient to estimate the distortion. Therefore, image resampling is implemented using local affine transformation based on TIN to achieve high precision.

An optimized TIN has been obtained in the process of error elimination. For every triangle in the optimized TIN, the affine transformation is calculated. Based on the parameters of the transformation, the gray data of the triangle in the input image can be rectified to the reference image. When all the triangles have been rectified, image registration between the input image and reference image is achieved. To achieve registration precision at the sub-pixel level, the bilinear interpolation method is used to resample the input image.

3. Experiments and Results

3.1. Description and Accuracy Analysis of Test Data

To evaluate the performance of the proposed approach, three datasets from ZY-1-02C were used for the experiments. Among these datasets, two consist of a scene of a PAN image and a scene of an MUX image, whereas the third set consists of a scene of an HR image and a scene of an MUX image. The first dataset is in Mohe, a county in the northern border of China, with a mainly mountainous terrain. The second dataset is in Hangzhou, an eastern city of China, where the main terrain is urban and partly mountainous. The third dataset is in Tokyo, Japan, which consists of the main part of the city and a large sea area. Details of the datasets are described in Table 1.

Table 1. Overview of test datasets.

Dataset ID	Location	Track Number	Lat/Long	Date	Sensor	Resolution (m)
1	Mohe	000589	121.8/53.3	1 February 2012	PAN/MUX	5.0/10.0
2	Hangzhou	000847	120.1/30.3	12 February 2012	PAN/MUX	5.0/10.0
3	Tokyo	000918	139.9/35.7	24 February 2012	HR/MUX	2.36/10.0

Before image registration, geometric correction with equivalent GSD was applied to the images. Meanwhile, subsampled corrected images were generated. According to Equation (2), the sizes of the subsampled images were close to and less than 1,400 × 1,400 pixels. The corrected PAN/HR image was then used as the reference image and the corrected MUX image was used as the input image for registration. The information contained in the corrected images is described in Table 2.

Table 2. Images after geometric correction.

Dataset ID	GSD (m)	Reference Image		Input Image	
		Sensor	Size	Sensor	Size
1	5.0	PAN	15,119 × 14,318	MUX	15,136 × 14,505
2	5.0	PAN	14,534 × 13,882	MUX	14,528 × 13,926
3	2.5	HR	27,466 × 29,645	MUX	28,811 × 27,318

To clarify the necessity of image coarse matching before precise matching, a number of checkpoints evenly distributed on the images were manually measured, and the coordinate differences between the input and reference images were analyzed. Generally, corresponding points can be determined by using the geographic information of the images. Therefore, we firstly obtained the geographic coordinates of the checkpoints, which are (X_{ref}, Y_{ref}) and (X_{input}, Y_{input}), and then calculated the differences $\Delta x_{geo} = |X_{ref} - X_{input}|/GSD$ and $\Delta y_{geo} = |Y_{ref} - Y_{input}|/GSD$. The comparative results of the coordinates of checkpoints are shown in Figure 3 and Table 3. Δx_{geo} and Δy_{geo} are extremely high in all three datasets. Meanwhile, discrepancies exist between the values of different checkpoints. Camera distortion and platform instability are mainly responsible for the differences. Although indoor calibration for the cameras of ZY-1-02C was conducted, the in-orbit calibration was not done before the experiment. Under such conditions, a significantly large searching window has to be set if ABM is directly performed on the corrected ZY-1-02C images. However, this process results in considerably lower reliability of match results and more computing time.

Figure 3. Coordinate difference between observed values of checkpoints. For (**a**,**b**) Datasets 1 and 2, 16 pairs of checkpoints are measured; and for (**c**) Dataset 3, the scale of the image is larger and 36 pairs of checkpoints are measured.
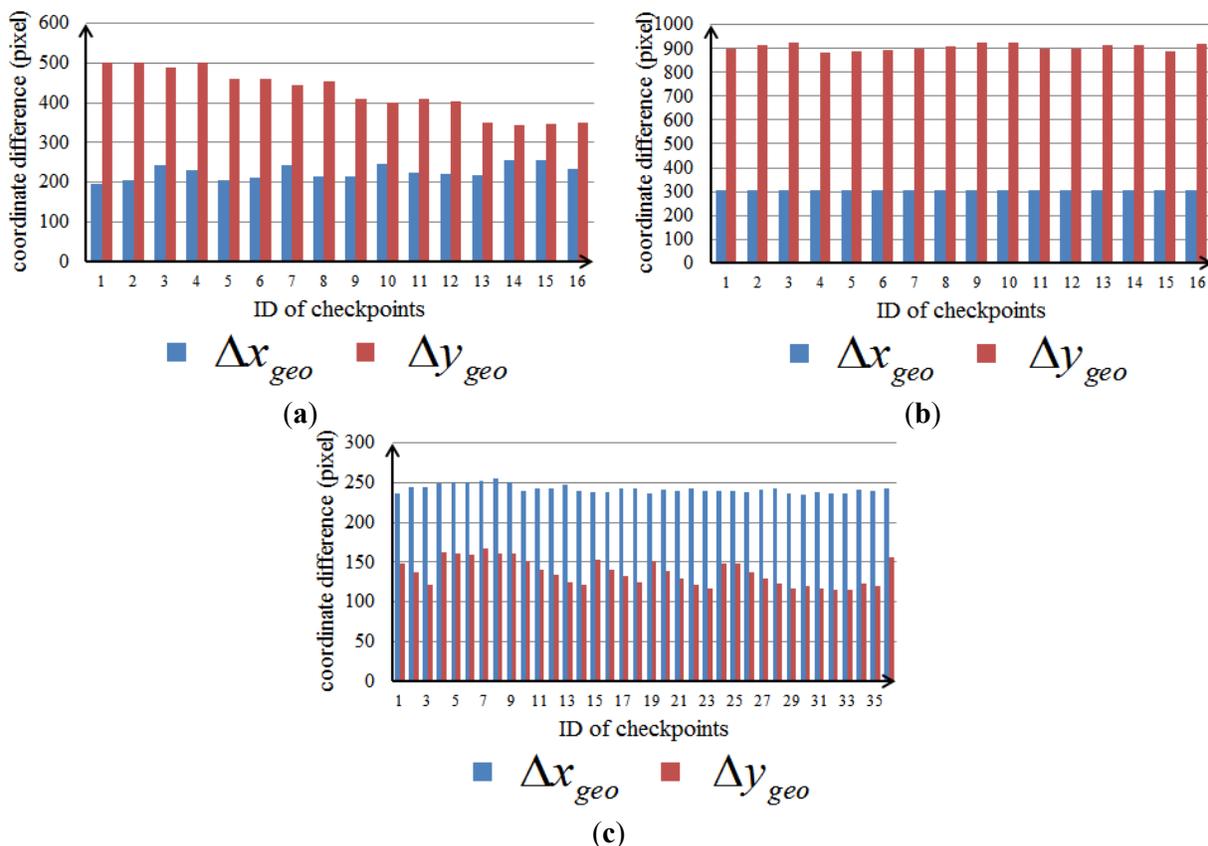

(a)

(b)

(c)

Table 3. Statistics of coordinate differences between corresponding checkpoints.

Coordinate Difference		Minimum (pixel)	Maximum (pixel)	Mean (pixel)
Dataset 1	Δx_{geo}	195.34	255.67	226.36
	Δy_{geo}	343.17	502.30	427.01
Dataset 2	Δx_{geo}	303.42	307.53	304.89
	Δy_{geo}	882.04	927.03	906.54
Dataset 3	Δx_{geo}	235.05	255.15	242.09
	Δy_{geo}	115.35	167.09	137.10

3.2. Determining Global Affine Parameters

According to the proposed strategy, SIFT matching was conducted on the subsampled images, and a rough estimation of global affine transformation was made based on the matching results. Figure 4 shows the results of SIFT matching. The SIFT parameters used are the recommended values in the study conducted by Lowe [17]. The number of matching points in each dataset is sufficient to determine the global affine parameters. The coordinate differences after affine transformation can be calculated using the determined parameters: assuming that the coordinate of a point on the input image is (x_{input}, y_{input}), according to affine transformation, the estimated value of the corresponding point

on the reference image can be computed as $(\overline{x_{ref}}, \overline{y_{ref}})$. The difference between the estimated and observed values can then be calculated as $\Delta x_{affine} = |x_{ref} - \overline{x_{ref}}|$ and $\Delta y_{affine} = |y_{ref} - \overline{y_{ref}}|$. Figure 5 shows the coordinate differences of the three datasets. It can be seen that, although more corresponding points are obtained in Dataset 1, the coordinate differences between the estimated and observed values of checkpoints are greater than those in Datasets 2 and 3. We believe that this condition is caused by the large topographic variations in the area of Dataset 1. The statistics in Table 4 reflect that the coordinate differences along the x-axis and y-axis directions are both controlled within a certain range. Even the maximum value is less than 22 pixels, which indicates that the searching range for the corresponding points can be significantly reduced with the results of coarse matching.

Figure 4. Results of SIFT matching on subsampled images. Blue crosses indicate estimated points and red crosses indicate excluded points. (**a**) Dataset 1. 1,941 pairs of corresponding points were matched, of which 41 pairs were false matches. (**b**) Dataset 2. 163 pairs of corresponding points were matched, of which 22 pairs were false matches. (**c**) Dataset 3. 308 pairs of corresponding points were matched, of which 20 pairs were false matches.
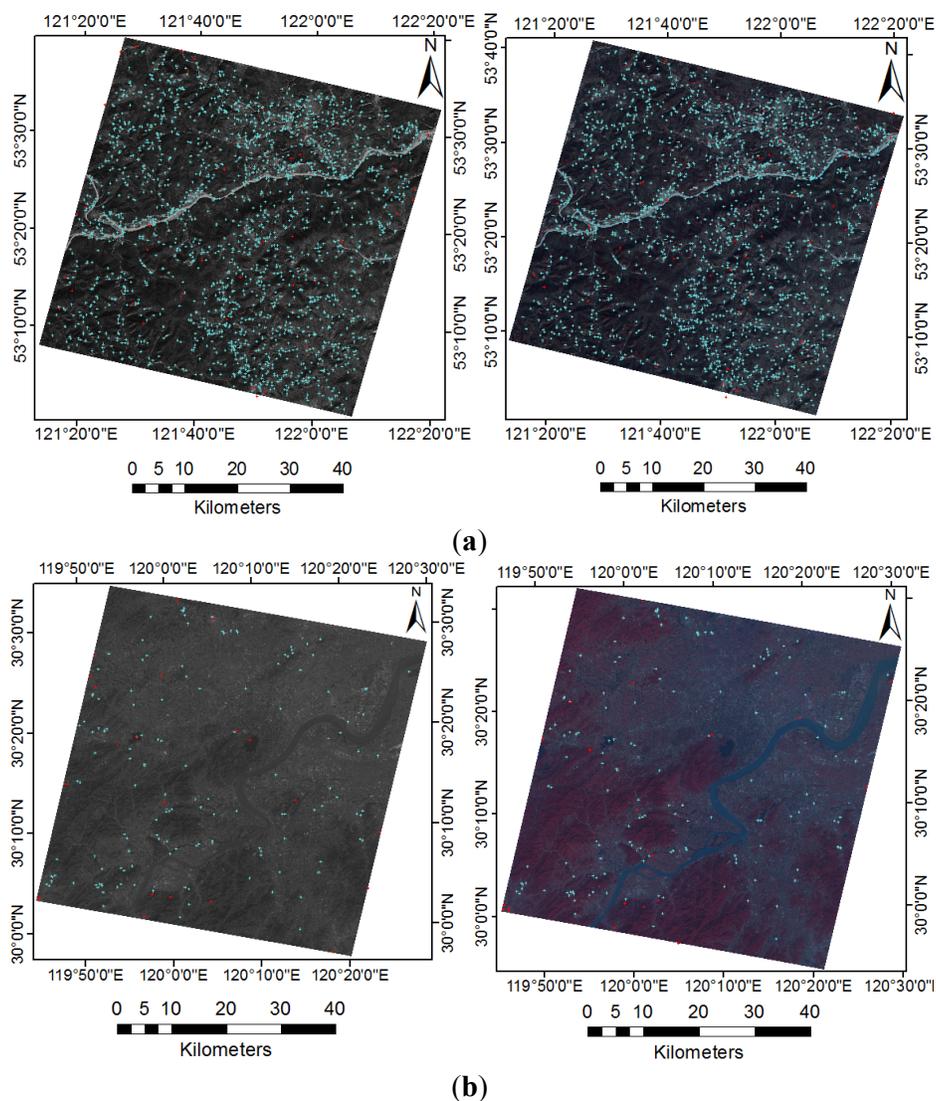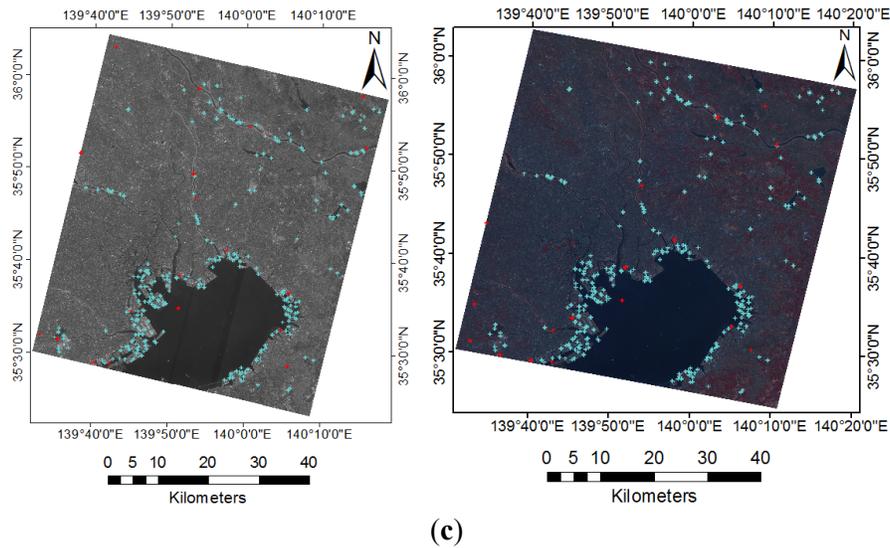

(**a**)

(**b**)

Figure 4. *Cont.*

(**c**)

Figure 5. Coordinate differences between estimated and observed values. (**a**) Dataset 1. (**b**) Dataset 2. (**c**) Dataset 3.
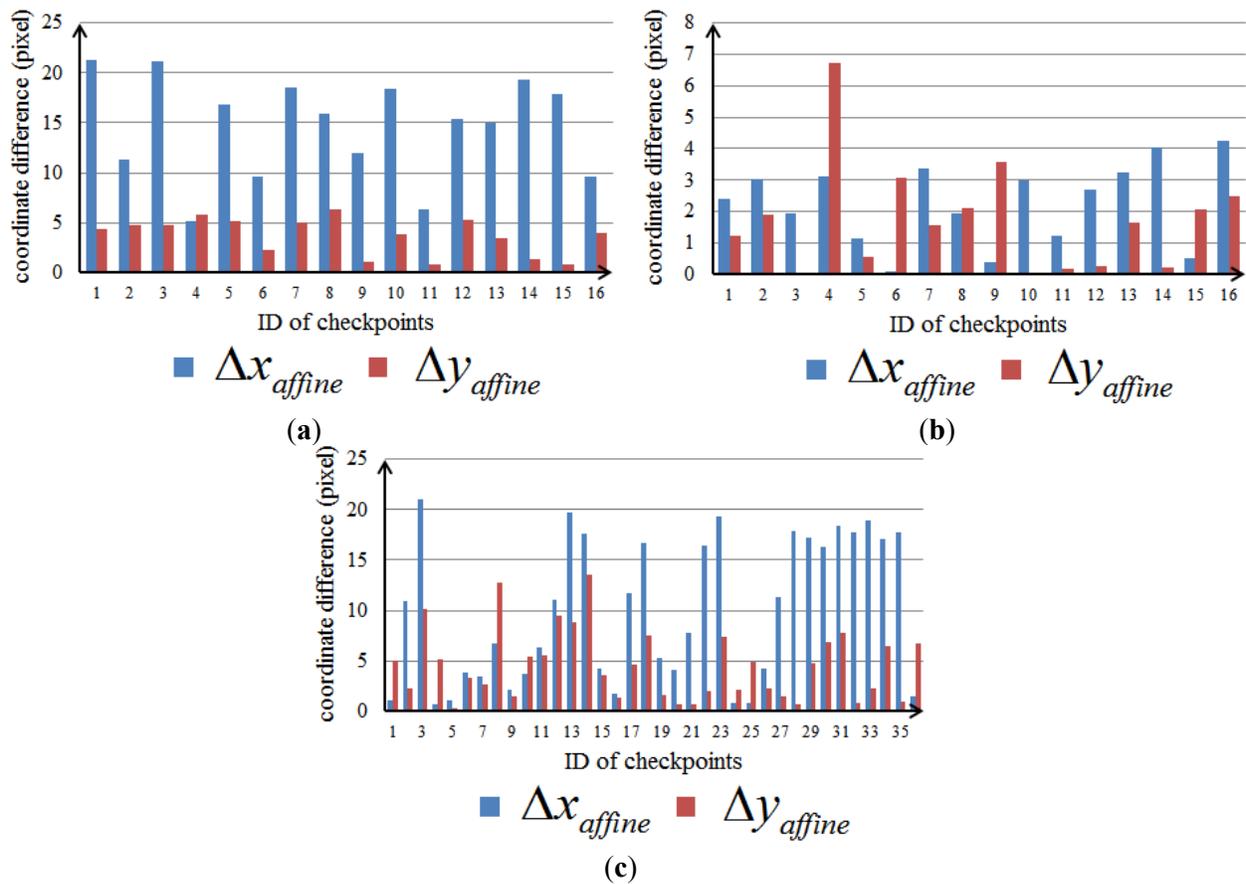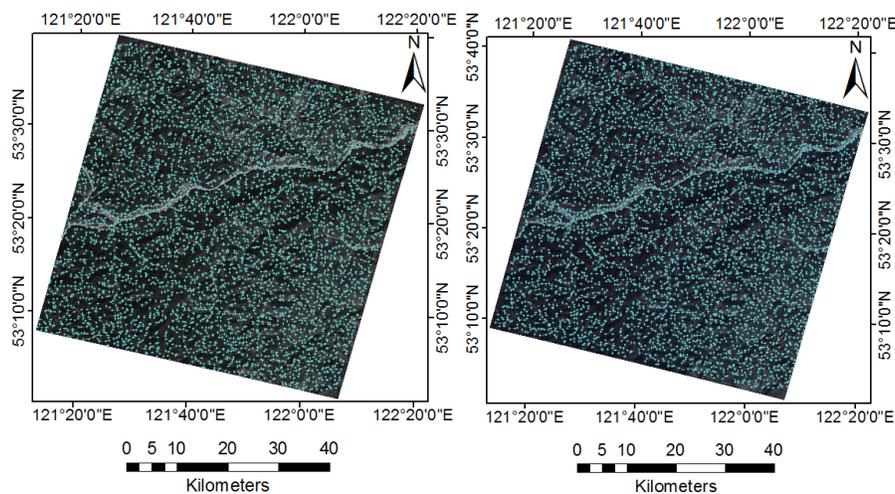

(**a**)

(**b**)

(**c**)

Table 4. Statistics of coordinate differences between estimated and observed values.

Coordinate Difference		Minimum (pixel)	Maximum (pixel)	Mean (pixel)
Dataset 1	Δx_{affine}	5.10	21.24	14.62
	Δy_{affine}	0.76	6.34	3.68
Dataset 2	Δx_{affine}	0.08	4.23	2.27
	Δy_{affine}	0.03	6.74	1.73
Dataset 3	Δx_{affine}	0.64	21.03	9.91
	Δy_{affine}	0.35	13.56	4.54

3.3. Acquisition of a Large Number of Evenly Distributed Corresponding Points

With the results of coarse matching, a large number of evenly distributed corresponding points can be acquired using the NCC-based matching method, and sub-pixel level matching accuracy can be achieved based on LSM. Figure 6 presents the results of precise matching. The grid size used in our approach to extract Harris feature points was 150×150 pixels, the template window size is 13×13 pixels, the searching window size is 51×51 pixels, and the threshold of correlation coefficient is 0.85. As shown in the figure, sufficient corresponding points can be obtained for all three datasets, even though in Dataset 3, the HR image resolution is over four times that of the MUX image. The density of corresponding points of Dataset 3 is evidently lower than that of the other two. Verification shows that a significant difference in resolution exists between the HR and MUX images, which hinders a number of feature points generated in the MUX image by the Harris detector from matching a point as image details are more abundant in the HR image.

Figure 6. Results of precise matching. Only correct matches are indicated on the images. (**a**) Dataset 1. A total of 6,456 Harris feature points were extracted and 4,945 pairs of points were successfully matched, of which 52 pairs were excluded as false matches. (**b**) Dataset 2. A total of 5,977 Harris feature points were extracted and 3,467 pairs of points were successfully matched, of which 117 pairs were excluded as false matches. (**c**) Dataset 3. A total of 20,431 Harris feature points were extracted and 7,224 pairs of points were successfully matched, of which 1,348 pairs were excluded as false matches.

(**a**)

Figure 6. *Cont.*

(b)

(c)

3.4. Results of Image Registration

Image registration was conducted by resampling the input image using local affine transformation based on TIN. In ASDPS, the images after registration are supposed to be fused by IHS algorithm [48], and the fused images can reflect the effect of registration adequately. The results of image registration and fusion are shown in Figures 7–9.

3.5. Efficiency and Accuracy Assessment

All the algorithms mentioned in this paper were implemented using C++ language. To verify the efficiency of the proposed method, a desktop computer with Microsoft Windows 7 operating system was adopted for the experiments, and the main hardware environment consisted of a quad-core CPU with a speed of 3.20 GHz and 4 GB memory. The statistics show that three datasets respectively

consumed 68.2 s, 83.2 s, and 561.8 s in the registration, which completely meet the time requirements of the commercial operation of ASDPS.
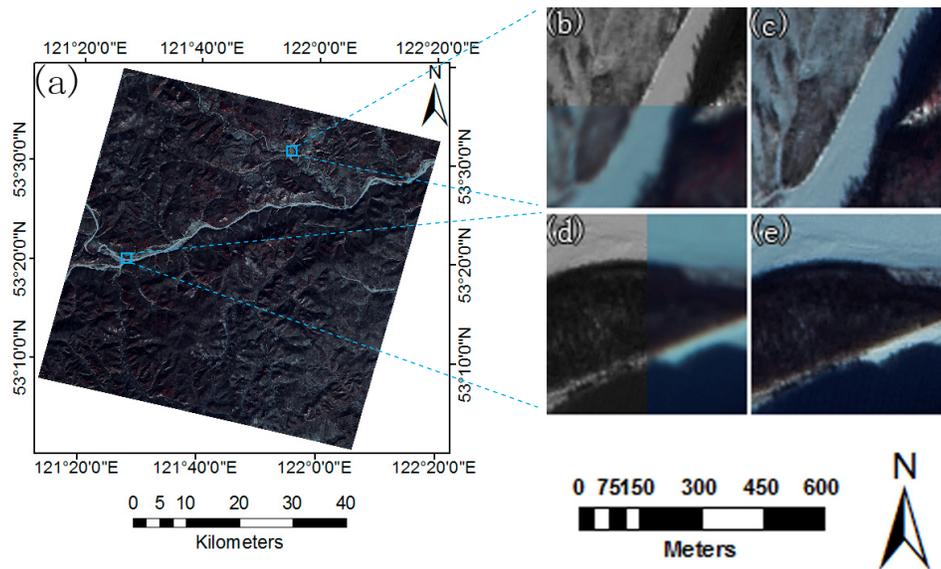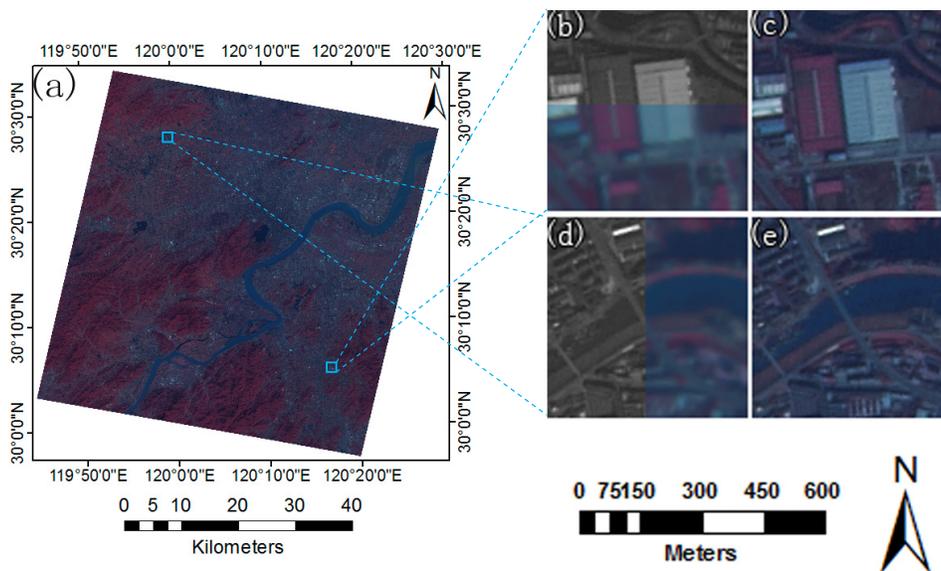
Figure 7. Registration and fusion results of Dataset 1. (**a**) provides an overall view of fused images, (**b**) shows the swipe effect of registered images in the first cyan rectangular area, (**c**) shows the fusion result, (**d**,**e**) show the result in the second rectangular area.
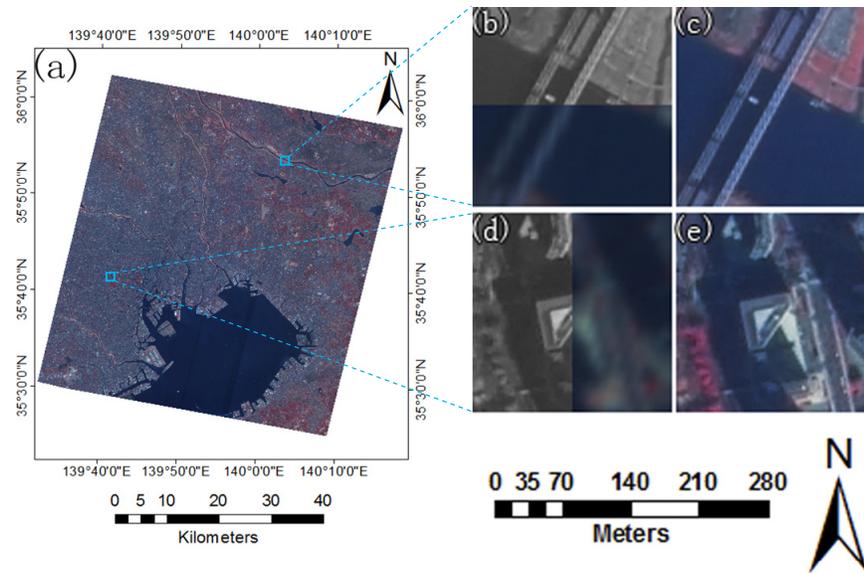

Figure 8. Registration and fusion results of Dataset 2. (**a**) provides an overall view of fused images, (**b**) shows the swipe effect of registered images in the first cyan rectangular area, (**c**) shows the fusion result, (**d**,**e**) show the result in the second rectangular area.

Figure 9. Registration and fusion results of Dataset 3. (**a**) provides an overall view of fused images, (**b**) shows the swipe effect of registered images in the first cyan rectangular area, (**c**) shows the fusion result, (**d**,**e**) show the result in the second rectangular area.

The comparative experiments were conducted by using ENVI 5.0. Similar to our approach, Förstner operator is chosen to extract feature points, and matching method of cross correlation is selected to obtain corresponding points. In ENVI, at least three pairs of seed tie points must be measured manually before the automatic generation of matching points, and the false matches should also be removed from matching points by human intervention.

For each set of data, registration tests based on three methods provided in ENVI were conducted, including affine transformation, quadratic polynomial, and triangulation. The registration and fusion results are shown in Figures 10–12. Misalignment evidently exists between the registered images, particularly those obtained through the affine transformation and quadratic polynomial methods, which directly resulted in the "edge phenomenon" [15] in the fused images.
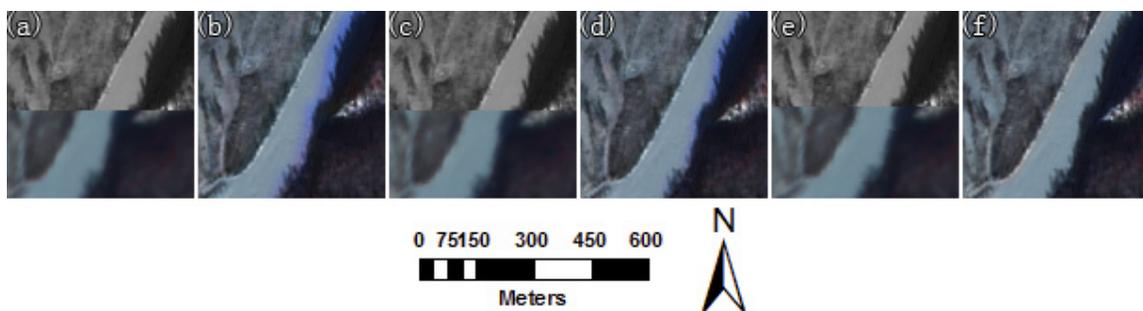
Figure 10. Registration and fusion results of Dataset 1 using ENVI methods. (**a**) shows the swipe effect of registered images through affine transformation method, (**b**) shows the fusion results. (**c**,**d**) show the result through quadratic polynomial method, (**e**,**f**) show the result through triangulation method.

Figure 11. Registration and fusion results of Dataset 2 using ENVI methods. (**a**) shows the swipe effect of registered images through affine transformation method, (**b**) shows the fusion results. (**c**,**d**) show the result through quadratic polynomial method, (**e**,**f**) show the result through triangulation method.
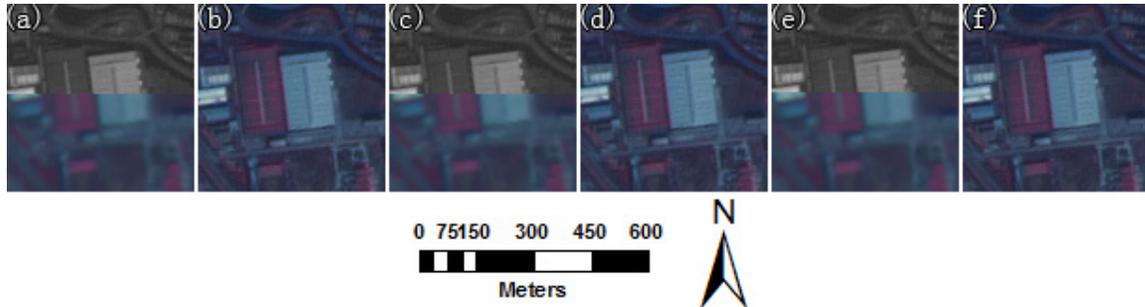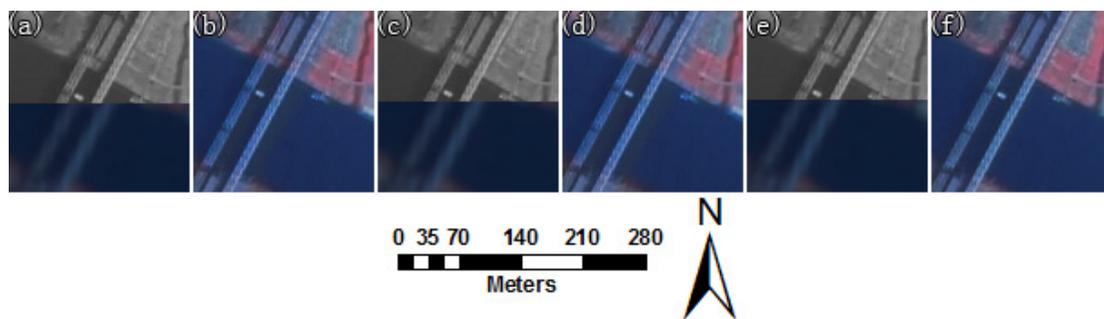

Figure 12. Registration and fusion results of Dataset 3 using methods of ENVI. (**a**) shows the swipe effect of registered images through affine transformation method, (**b**) shows the fusion results. (**c**,**d**) show the result through quadratic polynomial method, (**e**,**f**) show the result through triangulation method.
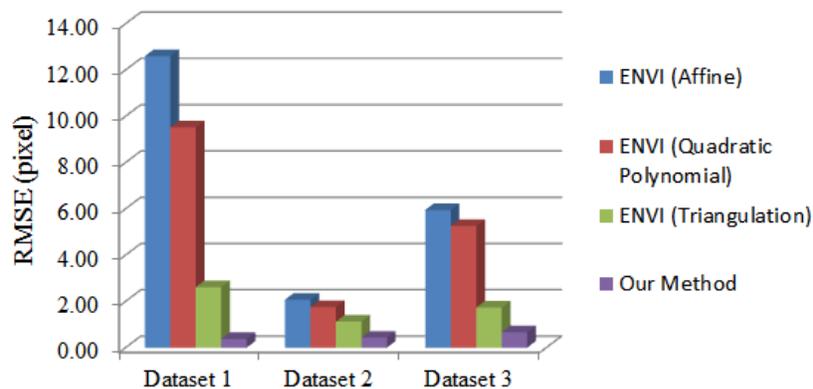

As our method is fully automatic, the efficiency is significantly higher than the methods of ENVI. To examine the registration accuracy, RMSE of the registered images by different methods were compared, which is calculated by 50 pairs of checkpoints. More methods of accuracy assessment can be found in Reference [49].

As shown in Table 5 and Figure 13, we compared the registration accuracy between methods in ENVI and our method. It can be seen that: first, the accuracy of our method reaches sub-pixel level, the overall precision of ENVI is lower than that of our method mainly because numerous evenly distributed corresponding points for registration cannot be obtained by the ENVI methods; then, it could be noticed that the accuracy of Dataset 1 is lower than others, which may be caused by the dramatic changes in local topography. We also found that the local methods, including ours and the triangulation method in ENVI, have significantly higher precision than global methods based on affine and quadratic polynomial transformation. This phenomenon indicates that complex distortion exists in geometrically corrected images of ZY-1-02C, and that the influence of image distortion can be properly addressed only with local transformation.

Table 5. Results of registration accuracy.

Method		RMSE (pixel)		
		Dataset 1	Dataset 2	Dataset 3
ENVI	Affine	12.59	2.07	5.95
	Quadratic Polynomial	9.51	1.75	5.26
	Triangulation	2.61	1.12	1.73
Our method		0.37	0.43	0.66

Figure 13. Comparison of registration accuracy.

4. Discussion

An AIR method with high precision and efficiency applied to fusion of ZY-1-02C imagery is proposed in this paper. A significant characteristic of the proposed method is the successful combination of the benefits from SIFT matching and NCC-based matching. As ASDPS is a commercial operation system with minimal human intervention, the proposed approach is a fully automatic procedure.

4.1. About the Coarse-to-Fine Registration

Our related studies have two representative approaches to realize a coarse-to-fine registration. One is to perform coarse matching on full-size images by means of FBM, such as SIFT or SURF, and perform precise matching on coarsely registered images by means of ABM [8,16]. This approach is feasible for regular-sized (less than 2,000 × 2,000 pixels) images. However, using FBM directly on the extremely large images processed in this article may consume excessive memory. Moreover, the matching results cannot be obtained through the use of an ordinary computer because the needed memory exceeds that of the computer hardware. The other approach only uses ABM and performs a progressive image matching based on an image pyramid, thereby reducing the time consumed for searching corresponding points [4,15]. However, constructing a multi-level image pyramid based on complex subsampling method (e.g., wavelet pyramid) for very large images reduces efficiency. When the difference in positioning accuracy between images is very significant, matching results obtained by ABM are not as reliable as those obtained by FBM.

To resolve the abovementioned issues, the corrected images were subsampled only once with a simple interval sampling method, and coarse matching based on SIFT was performed on the subsampled images. Relatively constant and suitable subsampled image sizes can be obtained with Equation (2). Thus, coarse matching can be conducted very efficiently, and the amount of memory consumed becomes acceptable. The overall workflow of the proposed matching strategy was dramatically simplified in comparison with methods based on image pyramids.

The error elimination applied in this article differs from the approaches reported in most studies. Traditional error elimination is mostly based on global estimation. The false matches or outliers are excluded by estimation based on least squares [8,9,28] or RANSAC [3,16,19,24,29]. However, complex distortion exists between images in our experiments, and, thus, a simple mathematical model cannot be directly used to describe the global transformation. In our approach, considering that the influences of distortions can be minimized when the judging area is reduced, we constructed a TIN structure; consequently, the false matches can be eliminated effectively by local affine estimation.

The main purpose of this article is to obtain fused ZY-1-02C images in spite of significant difference in positioning accuracy among images captured from different sensors. Our method can also be applied to similar problems in satellite data sources. However, the proposed method also has limitations. First, geometric correction with equivalent GSD was adopted before registration. The correction requires RPC parameters for each scene of images. Thus, our method cannot be directly applied to images with rotation and scale differences without RPC parameters. Moreover, using a coarse-to-fine strategy to achieve registration is unnecessary for satellite images with very small difference in positioning accuracy; NCC-based matching can be directly performed, and the corresponding points can be easily searched on the basis of the geographic information of the images.

4.2. Accuracies, Errors, and Uncertainties

In the experiments, 50 evenly distributed point pairs were manually selected for each dataset to assess the registration accuracy. The sub-pixel level accuracies were achieved for all the datasets. Comparative analysis revealed that our method is better than methods in ENVI.

The registration error in this article is mainly produced by three factors. First, the error may be generated in image matching. Theoretically, LSM can achieve sub-pixel level precision; however, the accuracy of individual point pairs may remain low because of different influences such as noises. Second, the error may be caused by local affine estimation. In most cases, affine transformation based on TIN structure can sufficiently estimate the local-image distortion. However, in image areas where the distribution of matching points is relatively sparse, some triangle may cross a larger range of image than others would. The estimation based on affine may be insufficient under this condition. Third, the error may be caused by image resampling. Bilinear interpolation adopted in our approach may also produce errors.

In the proposed method, the registration accuracy is significantly determined by the results of precise matching. Parameters such as the size of the template and searching window and the threshold of correlation coefficient are determinants of the matching results. In our experiments, the searching window size (51×51 pixels) is determined according to the coarse matching results, and the template window size (13×13 pixels) and the correlation coefficient threshold (0.85) are set from experience.

Whether there exists any parameter that could yield better results remains uncertain. Moreover, the maximum resolution discrepancy in our test datasets is about four times. We are also uncertain if the proposed approach can process images with higher differences in resolution.

5. Conclusions

Significant difference in positioning accuracy exists among images captured by different sensors of ZY-1-02C. A coarse-to-fine matching was applied to register images under this situation. Unlike other related works, instead of performing coarse matching directly on the full-size images or using a multi-level image pyramid to implement the coarse-to-fine matching, coarse matching based on scale invariant feature transform (SIFT) was performed for the subsampled images and a rough global estimation was made with the matching results. On the basis of the global estimation results, precise matching was conducted by means of normalized cross correlation and least squares matching. After eliminating the false matches through a local estimation by triangulated irregular network structure, image resampling was implemented based on local affine transformation to achieve registration.

The proposed method can achieve highly precise registration even when the sizes of the processed images are extremely large (maximum of 27,466 × 29,645 pixels). The experiments show that three test datasets consumed 68.2 s, 83.2 s, and 561.8 s in the registration with the accuracies of 0.37 pixels, 0.43 pixels, and 0.66 pixels, respectively. These findings suggest that our method is better than methods in ENVI. The accuracy and efficiency can fully meet application requirements of ZY-1-02C. Our method has been successfully applied at the China Center for Resources Satellite Data and Application, and the satellite data can be processed efficiently in a parallel computing environment.

However, it is important to mention that some further improvements are required. The adopted SIFT-matching approach can be replaced by a more efficient feature-based matching algorithm, such as speeded up robust features or Gradient location-orientation histogram [22]. Equivalent results can be achieved by these algorithms, which can reduce computation intensity. Meanwhile, during the process of geometric correction and image resampling, the gray value of the multispectral image was calculated twice with bilinear interpolation, which might result in a deviation from the gray value of the original image. Future studies aim to focus on solving the aforementioned problems.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Brook, A.; Ben-Dor, E. Automatic registration of airborne and spaceborne images by topology map matching with SURF processor algorithm. *Remote Sens*. **2011**, *3*, 65–82.

2. Wang, P.; Qu, Z.; Wang, P.; Ying, H.; Zhen, K. A coarse-to-fine matching algorithm for FLIR and optical satellite image registration. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 599–603.

3. Wang, B.; Lu, Q.; Li, Y.; Li, F.; Bai, L.; Lu, G.; Lai, R. Image registration method for multimodal images. *Appl. Opt.* **2011**, *21*, 1861–1867.

4. Bunting, P.; Labrosse, F.; Lucas, R. A multi-resolution area-based technique for automatic multi-modal image registration. *Image Vis. Comput.* **2010**, *28*, 1203–1219.

5. Hasan, M.; Pickering, M.R.; Jia, X. Modified SIFT for Multi-Modal Remote Sensing Image Registration. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012.

6. Song, Z.L.; Zhang, J. Remote sensing image registration based on retrofitted SURF algorithm and trajectories generated from lissajous figures. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 491–495.

7. Al-Ruzouq, R.I. Data fusion of multi-source imagery based on linear features registration. *Int. J. Remote Sens*. **2010**, *31*, 5011–5021.

8. Yu, L.; Zhang, D.; Holden, E.J. A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. *Comput. Geosci.* **2008**, *34*, 838–848.

9. Ma, J.; Chan, J.C.W.; Canters, F. Fully automatic subpixel image registration of multiangle CHRIS/proba data. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2829–2839.

10. Hu, Z.; Gong, W.; Cai, Z. Multi-resolution remote sensing image registration using differential evolution with adaptive strategy selection. *Opt. Eng.* **2012**, doi: 10.1117/1.OE.51.10.101707.

11. Loeckx, D.; Slagmolen, P.; Maes, F.; Vandermeulen, D.; Suetens, P. Nonrigid image registration using conditional mutual information. *IEEE Trans. Med. Imaging* **2010**, *29*, 19–29.

12. Suri, S.; Reinartz, P. Mutual-information-based registration of terraSAR-X and ikonos imagery in urban areas. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 939–949.

13. Wong, A.; Clausi, D.A. ARRSI: Automatic registration of remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1483–1493.

14. Lemoine, G.; Giovalli, M. Geo-correction of high-resolution imagery using fast template matching on a GPU in emergency mapping contexts. *Remote Sens.* **2013**, *5*, 4488–4502.

15. Hong, G.; Zhang, Y. Wavelet-based image registration technique for high-resolution remote sensing images. *Comput. Geosci.* **2008**, *34*, 1708–1720.

16. Lee, S.R. A coarse-to-fine approach for remote-sensing image registration based on a local method. *Int. J. Smart Sens. Intell. Syst.* **2010**, *3*, 690–702.

17. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

18. Bay, H.; Ess, A.; Tuytelaars, T.; van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359.

19. Cao, S.; Jiang, J.; Zhang, G.; Yuan, Y. An edge-based scale-and affine-invariant algorithm for remote sensing image registration. *Int. J. Remote Sens.* **2013**, *34*, 2301–2326.

20. Yang, Y.; Gao, X. Remote sensing image registration *via* active contour model. *AEU-Int. J. Electron. Commun.* **2009**, *63*, 227–234.

21. Huang, L.; Li, Z. Feature-based image registration using the shape context. *Int. J. Remote Sens.* **2010**, *31*, 2169–2177.

22. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1615–1630.

23. Zheng, Y.; Cao, Z.; Xiao, Y. Multi-spectral remote image registration based on SIFT. *Electron. Lett.* **2008**, *44*, 107–108.

24. Fan, B.; Huo, C.; Pan, C.; Kong, Q. Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 657–661.

25. Li, Q.; Wang, G.; Liu, J.; Chen, S. Robust scale-invariant feature matching for remote sensing image registration. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 287–291.

26. Sima, A.A.; Buckley, S.J. Optimizing SIFT for matching of short wave infrared and visible wavelength images. *Remote Sens.* **2013**, *5*, 2037–2056.

27. Goncalves, H.; Corte-Real, L.; Goncalves, J.A. Automatic image registration through image segmentation and SIFT. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2589–2600.

28. Wang, L.; Niu, Z.; Wu, C.; Xie, R.; Huang, H. A robust multisource image automatic registration system based on the SIFT descriptor. *Int. J. Remote Sens.* **2012**, *33*, 3850–3869.

29. Wu, Z.; Goshtasby, A. Adaptive image registration *via* hierarchical voronoi subdivision. *IEEE Trans. Image Process.* **2012**, *21*, 2464–2473.

30. Li, Y.; Davis, C.H. A Combined Global and Local Approach for Automated Registration of High-Resolution Satellite Images Using Optimum Extrema Points. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 7–11 July 2008.

31. Arévalo, V.; González, J. An experimental evaluation of non-rigid registration techniques on quickbird satellite imagery. *Int. J. Remote Sens.* **2008**, *29*, 513–527.

32. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography. In Proceedings of the Communications of the ACM, New York, NY, USA, June 1981.

33. Torr, P.H.S.; Zisserman, A. MLESAC: A new robust estimator with application to estimating image geometry. *Comput. Vis. Image Understand.* **2000**, *78*, 138–156.

34. Wong, A.; Clausi, D.A. AISIR: Automated inter-sensor/inter-band satellite image registration using robust complex wavelet feature representations. *Pattern Recognit. Lett.* **2010**, *31*, 1160–1167.

35. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 2nd ed.; Prentice Hall Press: Upper Saddle River, NJ, USA, 2002.

36. Ayoub, F.; Leprince, S.; Binet R.; Lewis, K.W.; Aharonson, O. Influence of Camera Distortions on Satellite Image Registration and Change Detection Applications. In Proceedings of the IEEE Geoscience and Remote Sensing Symposium, Boston, MA, USA, 7–11 July 1981.

37. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; *et al.* The shuttle radar topography mission. *Rev. Geophys.* **2007**, *45*, doi: 10.1029/2005RG000183.

38. Tao, C.V.; Hu, Y. A comprehensive study of the rational function model for photogrammetric processing. *Photogramm. Eng. Remote Sens.* **2001**, *67*, 1347–1358.

39. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441.

40. Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In Proceedings of the 4th Alvey Vision Conference, Alvey, UK, 31 August–2 September 1988.

41. Förstner, W.; Gülch, 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.

42. Zhang, Y.; Xiong, J.; Hao, L. Photogrammetric processing of low-altitude images acquired by unpiloted aerial vehicles. *Photogramm. Rec.* **2011**, *26*, 190–211.

43. Ackermann, F. Digital image correlation: Performance and potential application in photogrammetry. *Photogramm. Rec.* **1984**, *11*, 429–439.

44. Zhang, Y.; Wang, B.; Duan, Y. An algorithm of gross error elimination in image matching for large rotation angle images. *Geomat. Inf. Sci. Wuhan Univ.* **2013**, *38*, 1135–1138.

45. Kang, Z.; Zlatanova, S. New Point Matching Algorithm for Panoramic Reflectance Images. In Proceedings of the International Symposium on Multispectral Image Processing and Pattern Recognition, Wuhan, China, 15 November 2007.

46. Kang, Z.; Li, J.; Zhang, L.; Zhao, Q.; Zlatanova, S. Automatic registration of terrestrial laser scanning point clouds using panoramic reflectance images. *Sensors* **2009**, *9*, 2621–2646.

47. Lee, D.T.; Schachter, B.J. Two algorithms for constructing a delaunay triangulation. *Int. J. Comput. Inform. Sci.* **1980**, *9*, 219–242.

48. Chavezp, P.; Sides, S.C.; Anderson, J.A. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogramm. Eng. Remote Sens.* **1991**, *57*, 295–303.

49. Gonçalves, H.; Gonçalves, 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.