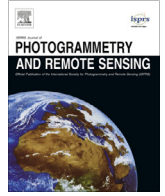




Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences

Yuanxin Ye ^{a,b}, Jie Shan ^{b,c,*}^a Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610031, China^b School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China^c School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:

Received 23 October 2012

Received in revised form 15 January 2014

Accepted 25 January 2014

Keywords:

Image registration

Multispectral remote sensing image

SR-SIFT

Local self-similarity

ABSTRACT

Image registration is a crucial step for remote sensing image processing. Automatic registration of multispectral remote sensing images could be challenging due to the significant non-linear intensity differences caused by radiometric variations among such images. To address this problem, this paper proposes a local descriptor based registration method for multispectral remote sensing images. The proposed method includes a two-stage process: pre-registration and fine registration. The pre-registration is achieved using the Scale Restriction Scale Invariant Feature Transform (SR-SIFT) to eliminate the obvious translation, rotation, and scale differences between the reference and the sensed image. In the fine registration stage, the evenly distributed interest points are first extracted in the pre-registered image using the Harris corner detector. Then, we integrate the local self-similarity (LSS) descriptor as a new similarity metric to detect the tie points between the reference and the pre-registered image, followed by a global consistency check to remove matching blunders. Finally, image registration is achieved using a piecewise linear transform. The proposed method has been evaluated with three pairs of multispectral remote sensing images from TM, ETM+, ASTER, Worldview, and Quickbird sensors. The experimental results demonstrate that the proposed method can achieve reliable registration outcome, and the LSS-based similarity metric is robust to non-linear intensity differences among multispectral remote sensing images.

© 2014 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.

1. Introduction

With the rapid development of remote sensing technology, remote sensing images from different sensors usually differ in spectral and spatial resolutions. Remote sensing images from different sensors or different spectral bands can provide complementary information about Earth's surface. Registration of these images in a common geographic coordinate is necessary for Earth observation. As a fundamental task in image processing, image registration aligns two or more images with overlapping scenes taken at different times, from different viewpoints, or by different sensors (Zitova and Flusser, 2003). For remote sensing image processing, image registration is also a prerequisite step for image mosaic, object identification, image fusion, and change detection. An automatic solution to this problem is highly desired as conventional image

registration techniques often require manual collection of tie points¹ between the images, which is often toilsome and time consuming. Different spectral bands often reflect different radiometric characteristics of the same scene resulting in significant intensity differences among the images acquired from different spectra, especially between visible and infrared imagery (Kern and Pattichis, 2007). The significant intensity differences can be visualized in Fig. 1, showing a pair of images acquired by TM band 1 (visible) and TM band 5 (infrared) sensors at the same time. Due to very different minute details now available between the multispectral images, the detection of tie points became much more difficult than ever before. Therefore, automatic registration of multispectral remote sensing images could be challenging.

Generally, automatic image registration consists of two main steps.

- (1) Image matching: detect the tie points between the reference and sensed image.

* Corresponding author at: School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

E-mail address: jshan@purdue.edu (J. Shan).

¹ The correspondence points between the reference and sensed image.

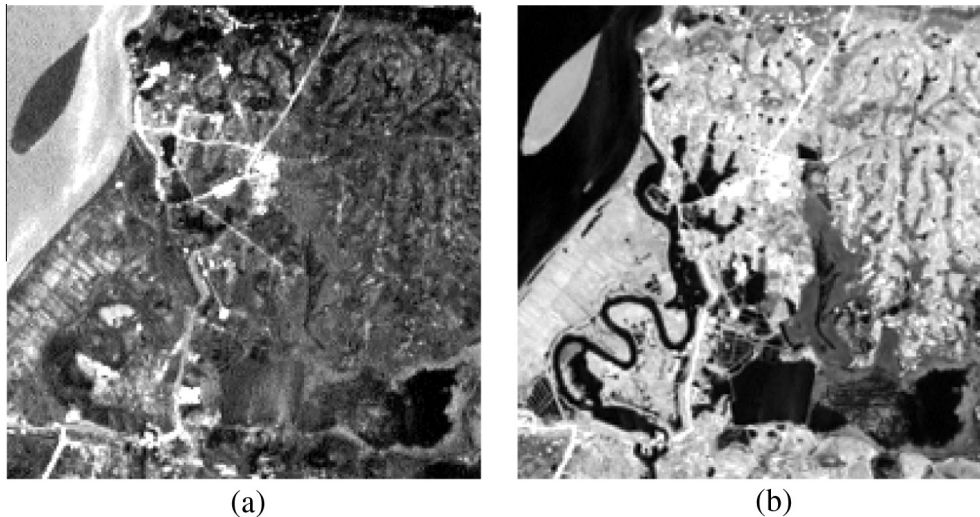


Fig. 1. The significant intensity differences between the multispectral images covering the same scene. (a) the TM band 1 image, (b) the TM band 5 image. The two images were taken at the same time.

- (2) Image rectification: Determine a transform model using the matched tie points and rectify the sensed image.

Image matching is a prerequisite for image registration and the quality of tie points influences the registration accuracy. According to the different ways of image matching, image registration methods can be classified into two categories, namely, feature-based and area-based methods (Zitova and Flusser, 2003).

Feature-based methods first extract the features (e.g., points, lines, contours) from the images and then detect the tie points based on the similarity of these features. Representative feature-based methods include contour-based method (Li et al., 1995), invariant moment-based method (Dai and Khorram, 1999), linear or edge feature-based method (Dare and Dowman, 2001), and phase congruency-based feature extraction method (Wong and Clausi, 2007, 2010). Recently, local feature descriptors have been rapidly developed in the computer vision field. The most representative local feature descriptor is the Scale Invariant Feature Transform (SIFT, Lowe, 2004), which has been widely used in the registration of remote sensing images because of its invariant to image scale and rotation changes (Li et al., 2006; Yu et al., 2008). Moreover, in order to further improve the matching performance of SIFT, some researchers have proposed a series of enhanced SIFT algorithms such as Colored SIFT (Abdel-Hakim and Farag, 2006), Scale Restriction SIFT (SR-SIFT, Yi et al., 2008), Affine-SIFT (Morel and Yu, 2009), and Uniform Robust SIFT (Sedaghat et al., 2011). However, being fundamentally similar to SIFT, these algorithms are designed for images with linear intensity changes; matching images with non-linear intensity differences is still difficult (Kelman et al., 2007; Tsai et al., 2010). Achieving enough tie points for precise registration using SIFT-based methods alone is difficult due to intensity differences among multispectral or multisensor remote sensing images. Despite being sensitive to the non-linear intensity differences, SIFT-based methods are still useful for initial registration of multispectral or multisensor images, during which two images are coarsely aligned only using a small number of tie points (Yu et al., 2008).

Different from feature-based methods, area-based methods, sometimes called template matching, first define a template window in the sensed image, and then search for a correspondence window in the reference image using different kinds of similarity metrics. The centers of matched windows are regarded as tie

points. The similarity metrics play a crucial role in the area-based methods. The conventional similarity metrics mainly include the sum of square difference (SSD), the normalized cross correlation (NCC), the mutual information (MI) and et al. SSD performs similarity evaluation through comparing the differences of grey values among images directly. As such, it is usually vulnerable to complex intensity variations. As a classical similarity metric, NCC has been widely applied for image registration because of its invariant to linear intensity changes, however, it cannot properly handle images with non-linear intensity differences (Fan et al., 2010; Hel-Or et al., 2011). MI is robust to non-linear intensity differences and has been successfully applied in the registration of multispectral or multisensor images (Cole-Rhodes et al., 2003; Kern and Pattichis, 2007). However, the registration methods based on MI is computationally expensive (Hel-Or et al., 2011), which may pose a restriction in practice. In addition, area-based methods often suffer from image distortions due to the rectangular window used in the matching process. This type of window may not cover the same part of the scene between the images with complex geometric distortions. In order to solve this problem, a pre-registration procedure could be required to reduce these distortions.

Overall, feature-based methods are more robust to the geometric distortions of images compared with area-based methods, whereas area-based methods have better resistance to the non-linear intensity differences among images. Therefore, this paper proposes an automatic registration method for multispectral images by integrating feature-based and area-based methods. The proposed method involves a coarse-to-fine registration scheme. The pre-registration is first achieved using the SR-SIFT algorithm, which improves the correct match rate for multispectral images compared with the original SIFT algorithm (Yi et al., 2008). Then, we introduce a new similarity metric based on the local self-similarity (LSS) descriptor to determine the correspondences in the fine registration stage. LSS is a local feature descriptor that captures the internal geometric layouts of images (Shechtman and Irani, 2007). Recently, the LSS descriptor has been successfully applied for the registration of thermal and visible videos, and proved to be able to handle complex intensity variations (Torabi and Bilodeau, 2012).

Upon a successful image matching, the subsequent image rectification is straightforward. Since the main objective of this paper is to develop a robust technique of tie-point detection to address the non-linear intensity differences among multispectral remote

sensing images, we use a piecewise linear (PL) transform (Goshtasby, 1986) to handle the complex distortions between the images. It provides an alternative way for image registration to conventional orthorectification with an accurate digital elevation model and known sensor geometry, both of which are sometimes difficult and complex to acquire.

The main contribution of this paper includes two aspects. One is a coarse-to-fine automatic registration scheme designed for multispectral remote sensing images; the other is the LSS integrated as a new similarity metric to detect tie points between multispectral images. The methodology is described in Section 2, followed by experiments and evaluation in Section 3. Conclusions are presented in Section 4.

2. Methodology

Given a reference image and a sensed image acquired from different spectral bands or different sensors, image registration aims to determine an optimal alignment between the two images. The proposed registration method involves a two-stage process, namely, pre-registration and fine registration. In the pre-registration stage, candidate tie points are first selected using the SR-SIFT algorithm, and the outliers are removed by RANdom SAmple Consensus (RANSAC, Fischler and Bolles, 1981). Then, the sensed image is rectified through a projective transform. This rectified image is referred to as the “pre-registered image”. The purpose of pre-registration is to eliminate the obvious translation, rotation, and scale differences between the images. In the fine registration stage, the Harris corner detector is first applied to extract the evenly distributed interest points in the pre-registered image. Then, the LSS is employed as a new similarity metric for template matching in the reference image to detect the correspondences. The tie points with large errors are removed by a global consistency check method. Finally, the pre-registered image is rectified using a PL transform, which constructs the Delaunay triangulated irregular networks (TINs) with the tie points and applies an affine transform to rectify each triangle region. Fig. 2 shows the main process of our method.

2.1. Pre-registration by SR-SIFT

The SIFT algorithm first detects keypoints as the extrema in the Difference-of-Gaussian scale space, and assigns a main orientation

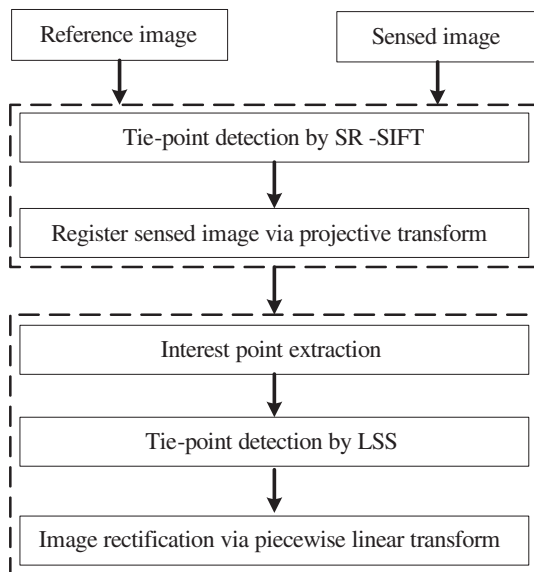


Fig. 2. Flowchart of the proposed image registration method.

for each keypoint. Then, the descriptor is computed using gradients within the local neighborhood with respect to the scale and orientation of each keypoint. The resulting descriptor is invariant to image scale and rotation changes. However, SIFT does not work well for images with non-linear intensity differences (Kelman et al., 2007; Li et al., 2011).

The SR-SIFT algorithm (Yi et al., 2008) was proposed to improve the performance of SIFT in matching multispectral image. The main improvements of SR-SIFT include the following two aspects.

- (1) Gradient orientation modification: As the pixel intensity of the same region between multispectral images is often different or even inverse (the river in Fig. 3), the SR-SIFT algorithm modifies the gradient orientation and limits it between 0 and π through the following equation:

$$\beta(\alpha) = \begin{cases} \alpha & \alpha \in [0^\circ, 180^\circ) \\ 360^\circ - \alpha & \alpha \in (180^\circ, 360^\circ) \end{cases} \quad (1)$$

where α is the gradient orientation, and β is the modified orientation.

- (2) Scale restriction criteria: Given a keypoint pair $k_1(x_1, y_1, \sigma_1, \theta_1)$ and $k_2(x_2, y_2, \sigma_2, \theta_2)$, the scale difference (SD) of the keypoint pair is defined as

$$SD(k_1, k_2) = |\sigma_1 - \sigma_2| \quad (2)$$

where σ is the scale of keypoint.

The SDs of the correct matches should be much smaller than those of the mismatches. According to this property, the scale restriction criterion is defined as

$$a < SD(k_1, k_2) < b \quad (3)$$

A match will be discarded if it does not satisfy the scale restriction criteria. To decide the values of a and b in Eq. (3), a histogram of SDs of all matches is formed, and the peak in the histogram is noted as \overline{SD} . Then, the values of a and b are respectively assigned with $\overline{SD} - W$ and $\overline{SD} + W$, where W is a constant and is set between 0.20 and 0.35 (Yi et al., 2008).

Given the above improvements, the SR-SIFT algorithm can raise the correct match rate for multispectral images. However, the number of tie points obtained using only SR-SIFT may be too low and unevenly distributed in the image. Fig. 3 shows an example of tie points detected by applying SR-SIFT between a SPOT 4 band 2 (visible) and band 4 (infrared) image. SR-SIFT can only obtain a limited number of tie points. A possible reason is that different spectral information results in the significant non-linear intensity differences between the images. Nevertheless, SR-SIFT could provide at least four tie points for the images, which is sufficient to determine a projective transform² for pre-registration to eliminate the obvious translation, rotation, and scale differences between the images.

In the pre-registration stage, SR-SIFT keypoints are first extracted from the reference and sensed image, and a one-to-one matching between the keypoints is performed using the Euclidean distance ratio between the first and the second nearest neighbor (hereafter assigned as d_{ratio}). Then, the matches that satisfy the scale restriction criteria are selected as tie points, followed by the application of RANSAC to remove the outliers. Finally, the sensed image is rectified by a projective transform.

2.2. Fine registration

Once the pre-registration is completed, the reference and the pre-registered images have been coarsely aligned. We can

² The projective transform can handle affine transform (translation, rotation, and scale) and perspective transform.

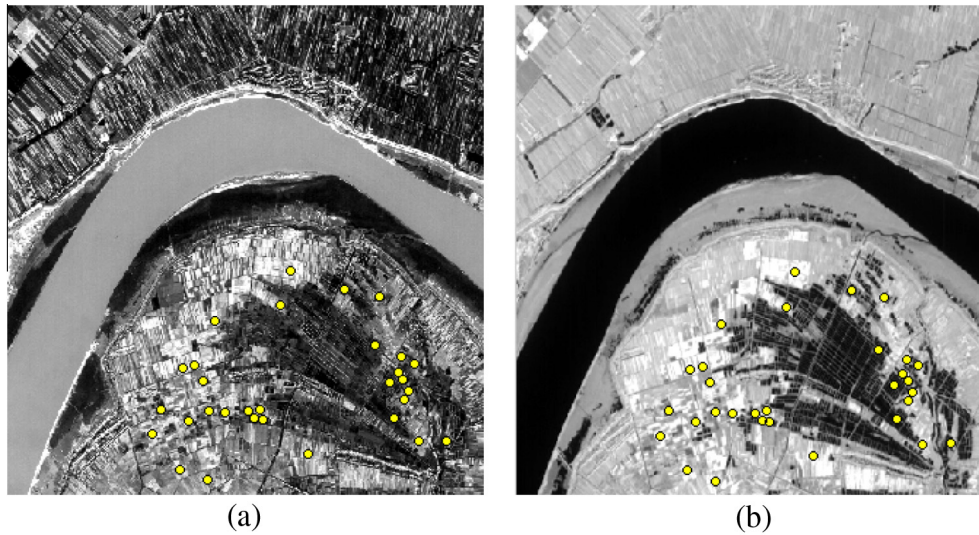


Fig. 3. The tie points detected using SR-SIFT between a SPOT 4 band 2 (a) and a SPOT4 band 4 image (b).

determine a small search area to detect the correspondences in the subsequent fine registration, which is carried out in accordance with the following steps: interest point extraction, tie-point detection by LSS, elimination of mismatched tie points, and image rectification.

2.2.1. Interest point extraction

In the fine registration, the first step is the extraction of the interest points in the pre-registered image. Schmid et al. (2000) evaluated different interest point detectors and found that the Harris corner detector (Harris and Stephens, 1988) performed best. This detector is therefore applied to extract the interest point in this step. However, the Harris detector often needs to set a fixed threshold to extract the interest points. This may make the interest points unevenly distributed over an image (Fig. 4a). To address this problem, a block-based scheme, which is free of threshold, is adopted in the extraction of interest points. First, the image is divided into $n \times n$ non-overlapping blocks, and the Harris value is computed for each block. Then, the Harris value in every block is

ranked from largest to smallest, and the strongest k points are selected as the interest points, where k is the number of interest points desired. An example of interest points extracted using this scheme is shown in Fig. 4b. The block-based scheme evenly distributes the interest points over the image.

2.2.2. Tie-point detection by LSS

LSS is a local feature descriptor that captures the internal geometric layouts of local self-similarities within images and represents the indirect local image property. As such, LSS can be used to match a textured region with other differently textured regions as long as they have similar layouts/shapes (Shechtman and Irani, 2007). This property is interesting for matching images because their local shape is similar but their intensity or texture is different.

Fig. 5 illustrates an example of the LSS descriptor generation of a local image region (typically 41×41 pixels). Within the local region, all surrounding image patches (typically 3×3 pixels) are compared with the patch centered at q using the sum of square

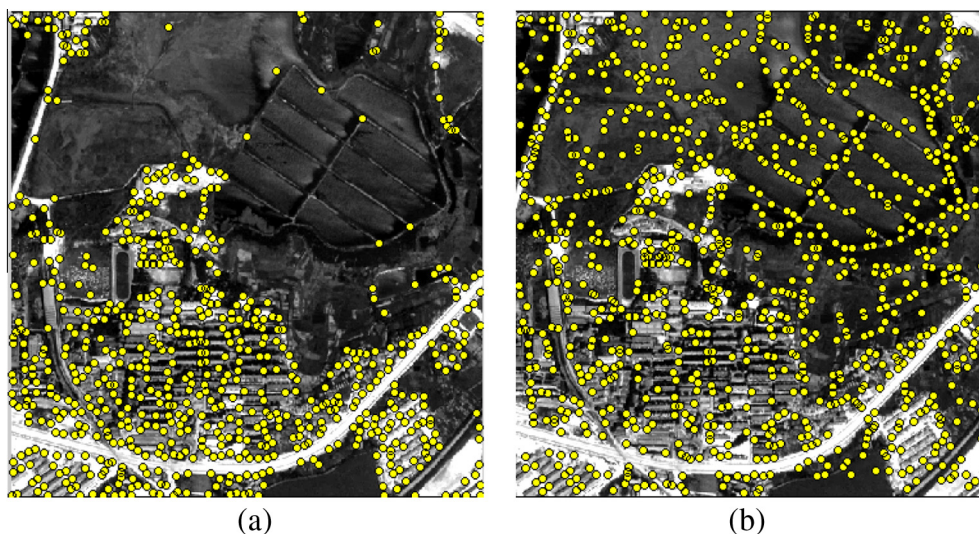


Fig. 4. The Harris interest points extracted by two different methods, namely, (a) fixed threshold and (b) block-based scheme where the image is divided into 10×10 blocks, and 10 interest points are extracted in each block.

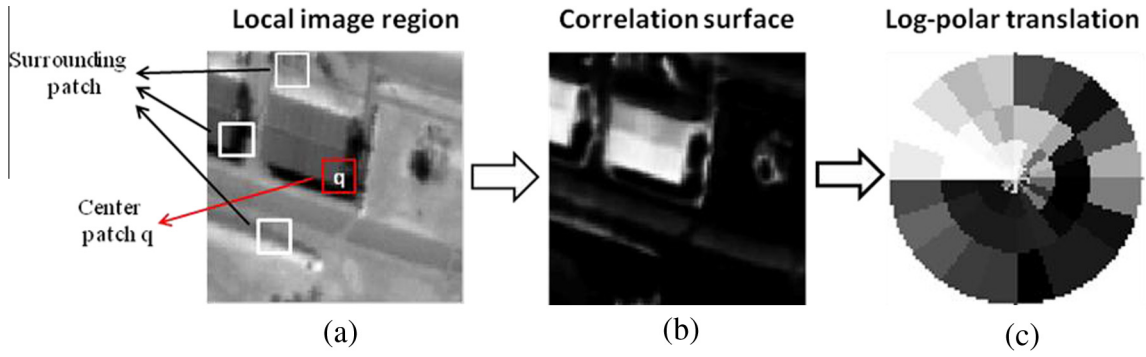


Fig. 5. An example of LSS descriptor generation. (a) local image region, (b) correlation surface of (a), (c) the log-polar translated LSS descriptor.

differences (SSD) between the patch intensity. Then, the $SSD_q(x, y)$ is normalized and transformed into a “correlation surface” $S_q(x, y)$.

$$S_q(x, y) = \exp\left(-\frac{SSD_q(x, y)}{\max(var_{noise}, var_{auto}(q))}\right) \quad (4)$$

where var_{noise} is a constant that corresponds to the intensity variations (in illumination or due to noise). $var_{auto}(q)$ is the maximal SSD of all patches within a very small neighborhood of q (of radius 1) relative to the center patch q , expected to account for the patch contrast and its pattern structure.

To account for local affine deformation, the correlation surface $S_q(x, y)$ is transformed into a log-polar representation and is partitioned into bins (e.g., 20 angles, 4 radial intervals). The maximal correlation value of each bin is selected to generate the LSS descriptor vector associated with the pixel q . Finally, the LSS descriptor is linearly stretched to the range of [0..1] to achieve invariance to the intensity variations of different patches in the local image region. Fig. 6 shows the LSS descriptors computed from

the corner, edge, and homogenous regions of the visible and infrared images in the same scene. The LSS descriptors are quite similar despite the large intensity differences between the two images.

Theoretically, LSS should be resistant to non-linear intensity differences because it captures the internal geometric layouts of the image and represents the local shape property. The NCC of LSS (hereafter assigned as LSCC) is used as the similarity metric for multispectral image registration. The LSCC is defined as

$$LSCC = \frac{\sum_{k=1}^N (S_{q_1}(k) - \bar{S}_{q_1})(S_{q_2}(k) - \bar{S}_{q_2})}{\sqrt{\sum_{k=1}^N (S_{q_1}(k) - \bar{S}_{q_1})^2 \sum_{k=1}^N (S_{q_2}(k) - \bar{S}_{q_2})^2}} \quad (5)$$

where q_1 and q_2 are the image pixels in the reference and the sensed images, respectively. $S_{q_1}(k)$ and $S_{q_2}(k)$ are the LSS descriptors of a surrounding image region (template window) centered at q_1 and q_2 , respectively. \bar{S}_{q_1} and \bar{S}_{q_2} represent the means of the LSS descriptors within the template window. N is the dimension of LSS descriptor.

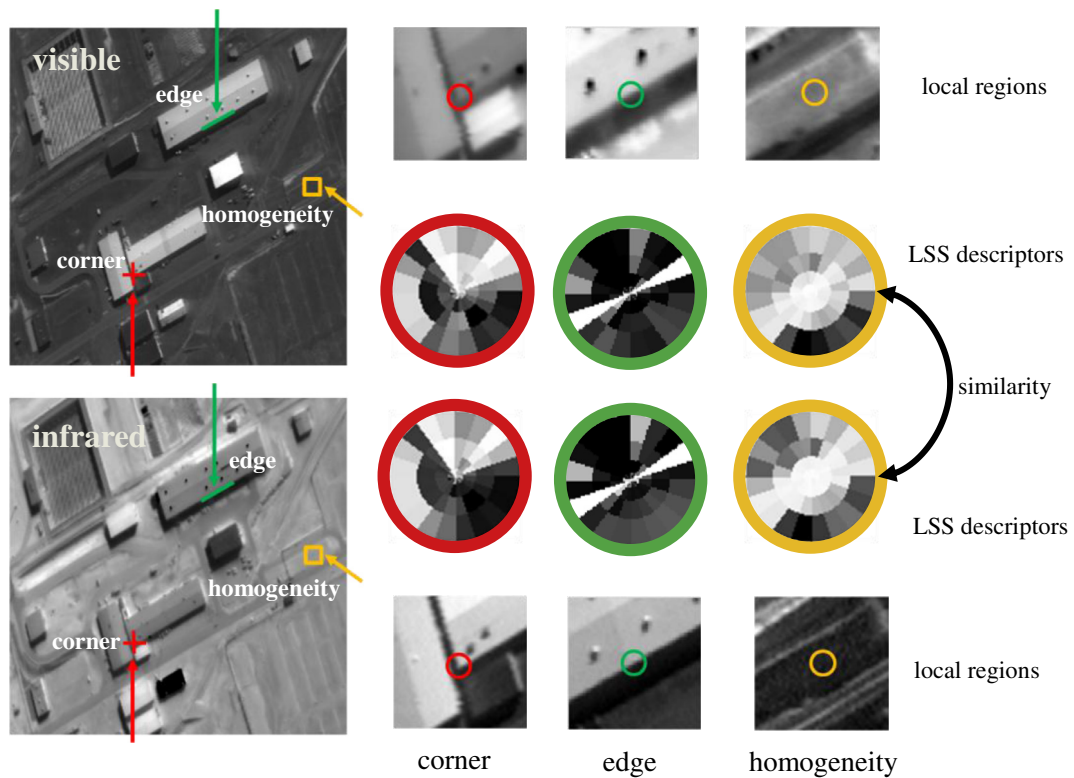


Fig. 6. The LSS descriptors of the visible and infrared images in the corner, edge, and homogenous regions.

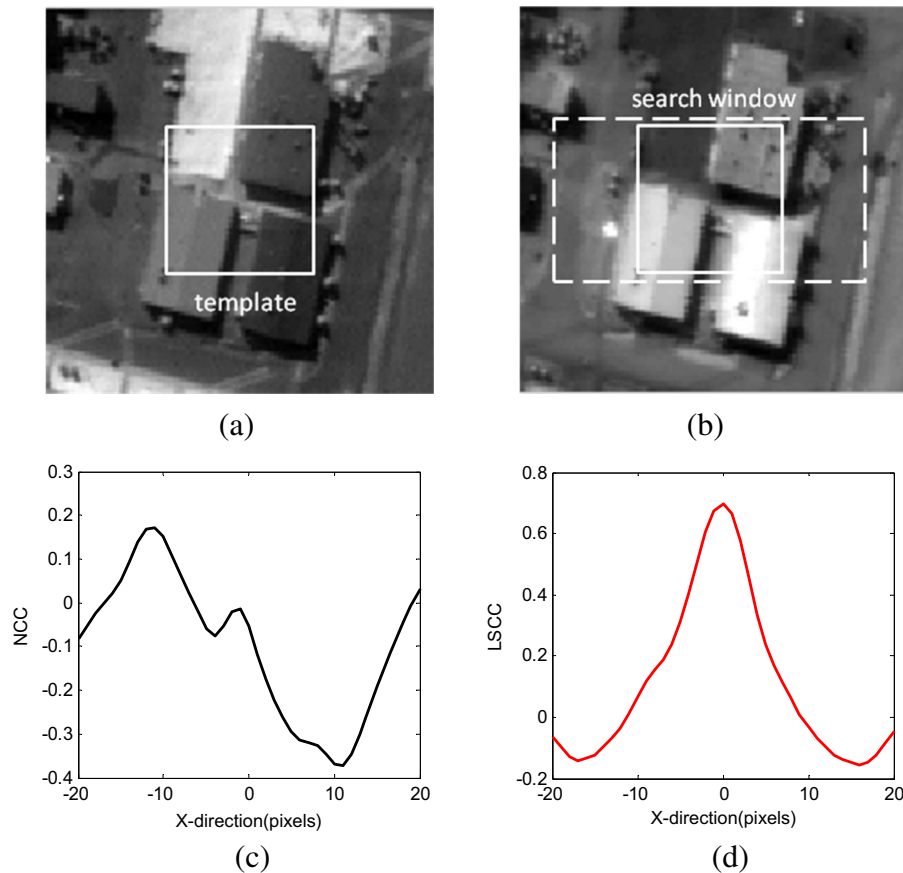


Fig. 7. The comparison of the similarity curve of LSCC and NCC. (a) The visible image, (b) the infrared image, (c) the NCC similarity curve, (d) the LSCC similarity curve.

LSCC is compared with the conventional NCC (computed from the grey intensity) by the similarity curve to illustrate its advantages to match the multispectral images. A template window (41×41 pixels) is first selected from a visible image. Then, we respectively calculate the LSCC and NCC for x -direction translations (-20 to 20 pixels³) within a search window of an infrared image. The LSCC and NCC similarity curves are shown in Fig. 7. It can be clearly seen that LSCC achieves the correct match at the maximum, but NCC fails. This example preliminarily indicates that LSCC is more robust than NCC to the non-linear intensity differences. The more comparative analysis for the performance of LSCC and NCC is presented in Section 3.4.

After a set of the interest points are extracted in the pre-registered image, a small search window in the reference image can be determined because the images have been coarsely aligned. Then, LSCC is used to detect the tie points using a template matching strategy. In the matching process, a conventional way to avoid some unreliable tie points is by setting a LSCC threshold. However, initial experiments show that the threshold is related to the template size and image scene, which results in being difficult to define a proper threshold for all the multispectral images. It often occurs that a LSCC threshold does not successfully distinguish good matches from bad ones. We instead apply a bidirectional matching technique (Di Stefano et al., 2004) which includes two steps, namely, forward matching and backward matching. In the forward step, for an interest point p_1 in the pre-registered image, its match point p_2 is found by the maximum of LSCC between the template window in the pre-registered image and the search window in the reference image. In the backward step, the match point of p_2

should be found in the pre-registered image using the same method. If the two matching procedures obtain a consistent result, the matched point pair (p_1, p_2) is regarded as the tie points.

2.2.3. Elimination of mismatched tie points

Given several uncertainty factors such as shadow and occlusion, the above resultant tie points are inevitably not error-free. The large tie-point errors are eliminated using a global consistency check method based on a global transform. The choice of transform is crucial for the consistency check, and it depends on the types of relative geometric deformations between the images. As the actual deformations between the images from different sensors are usually unknown, we select the projective model to handle common global transform, such as translation, rotation, and scale. It is suitable for the low- and high-resolution images covering the plain areas because few complex deformations exist among these images. However, for the images covering the mountain and urban areas, some researches find that the third-order polynomial model is often better than other global transforms such as projective and second-order polynomial models for prefitting the non-rigid deformations among such images (Hong and Zhang, 2008; Ma et al., 2010). The similar results are also presented in our experiments (Section 3.3). Therefore, in our global consistency check, the projective model is used for the low- and medium-resolution images with lowly undulating terrain, while the third-order polynomial model is selected for images with significant terrain relief.

The residual error and the root mean-square error (RMSE) are usually used to assess the registration accuracy. They are defined as

$$residual = \sqrt{(x_r - T(x_p))^2 + (y_r - T(y_p))^2} \quad (6)$$

³ This refers to the translation range of the center pixel of template window.

Table 1
Specifications of the test data.

Test	Reference image				Sensed image			
	Sensor	Spectrum (μm)	Size(pixels) resolution	Date	Sensor	Spectrum (μm)	Size(pixels) resolution	Date
Test 1	TM Band 3	0.63 ~ 0.49 (red)	700 × 825 30 m	2004/9	ETM + Band 7	2.09 ~ 2.35 (infrared)	743 × 847 30 m	2002/11
Test 2	ASTER Band 1	0.52 ~ 0.60 (green)	1099 × 1236 15 m	2002/7	TM Band 5	1.55 ~ 1.75 (infrared)	531 × 602 30 m	2002/7
Test 3	Worldview Band 8	0.86 ~ 1.04 (near-infrared)	865 × 901 2.0 m	2011/10	QuickBird Band 1	0.45 ~ 0.52 (blue)	719 × 755 2.4 m	2007/11

$$RMSE = \sqrt{\frac{\sum_{n=1}^N \text{residual}_n^2}{N}} \quad (7)$$

where (x_r, y_r) and (x_p, y_p) are the coordinates in the reference and pre-registered images, respectively. T represents the transform used in the global consistency check, and N is the total number of tie points.

The elimination of mismatched tie points (i.e. errors) is an iterative refining procedure. A transform (projective or third-order polynomial model) is first set up using the least squares method with all the tie points. Then, the residuals and RMSE of tie points are calculated. Subsequently, the tie point with the largest residual is removed. The aforementioned process is repeated until RMSE is less than a given threshold (e.g. 1 pixel).

2.2.4. Image rectification

After the removal of tie points with large errors, determining a geometric transform is necessary to rectify the pre-registered image. Considering the limitation of orthorectification that requires accurate sensor geometry and high-precision DEM, we employ a PL transform to fit the local distortions caused by terrain relief. PL transform is a local deformation model based on an image triangulation from a Delaunay TIN constructed by using tie points. The Delaunay TIN is first constructed with all the tie points by the incremental insertion method (Tsai, 1993), and then the parameters of an affine transform (Eq. (8)) is calculated for each triangle using the coordinates of its three vertices. Finally, the rectification of the pre-registered image is performed on each triangle region.

$$\begin{aligned} x_r &= a_0 + a_1 x_p + a_2 y_p \\ y_r &= b_0 + b_1 y_p + b_2 y_p \end{aligned} \quad (8)$$

where (x_r, y_r) and (x_p, y_p) are the coordinates in the reference and pre-registered images, respectively.

3. Experiments and evaluation

To validate the proposed method, it was tested using three sets of multispectral remote sensing images from different sensors. The test sets were divided into two categories: medium-resolution multispectral image and high-resolution multispectral image. The details of test sets are listed in Table 1. The images of test sets are shown in Fig. 8, where the first, second and third rows correspond to test 1, test 2, and test 3, respectively. The image pair of each set was composed of a reference and a sensed image, which were respectively acquired from different spectra resulting in significant non-linear intensity differences. In addition to these, other characteristics of each set are described below.

- (1) The test set 1 consisted of the medium-resolution multispectral images (from TM and ETM+) covering a plain area located at the eastern part of Wuhan, China. The resolutions of these two images are both 30 m/pixel. The maximum elevation difference⁴ is about 65 m in the covered area of the

images. The two images have almost no obvious local distortions due to the low resolution and the small terrain relief. However, rotation difference (about 16°) can be observed between the images.

- (2) The test set 2 consisted of the medium-resolution multispectral images (from ASTER and TM) covering a mountain area located at the northern part of Denver, USA. The resolutions of the ASTER image and the TM image are respectively 15 m/pixel and 30 m/pixel. The two images have differences in scale (2 times) and rotation (about 12°). Moreover, the ground elevation range varies between 1573 m and 2622 m in the covered area of the images. The large terrain relief results in the local distortions between the images.
- (3) The test set 3 consisted of the high-resolution multispectral images (from Worldview and QuickBird) covering an urban region located in San Francisco, USA. The Worldview image has a resolution of 2.0 m/pixel, and the QuickBird image has a resolution of 2.4 m/pixel. Scale difference (1.2 times) and rotation difference (about 18°) can be observed between the images. Due to being captured by different sensor models, these two high-resolution images contain the significant local distortions caused by the relief displacement of the building. Additionally, given the temporal difference of almost four years (Table 1) between the images, some buildings have changed (left part of the image) during this period. All these differences in this test make it very difficult to register the two images.

3.1. Parameter setting

For all the test sets, the d_{ratio} was set to 0.6 for matching the SR-SIFT keypoints in the pre-registration stage. In the fine registration stage, the pre-registered image was divided into 10×10 blocks, and 15 Harris interest points were extracted using the window of 3×3 pixels in each block, reaching a total of 1500 interest points. To calculate LSS, the surrounding patch and the template window size were set to 3×3 pixels and 41×41 pixels, respectively. In the global consistency check method, due to the few local distortions between the images, the projective model was used to check the mismatches for test 1. For test 2 and test 3, the third-order polynomial model was applied for error elimination because the images of the two tests have the more significant local distortions. The RMSE threshold is set to 1 pixel for all the tests.

3.2. Registration process

In the pre-registration stage, The SR-SIFT algorithm was first used to achieve the matches between the reference and the sensed image. Then, the outliers were removed using RANSAC, thereby obtaining the tie points between the images. Finally, the projective transform was used to rectify the sensed image. Table 2 lists the number of tie points in the pre-registration stage. We can observe from the table that the SR-SIFT algorithm only detects a small number of tie points. This means that the SR-SIFT algorithm is very

⁴ The elevation information in this paper is obtained from the ASTER Global DEM.

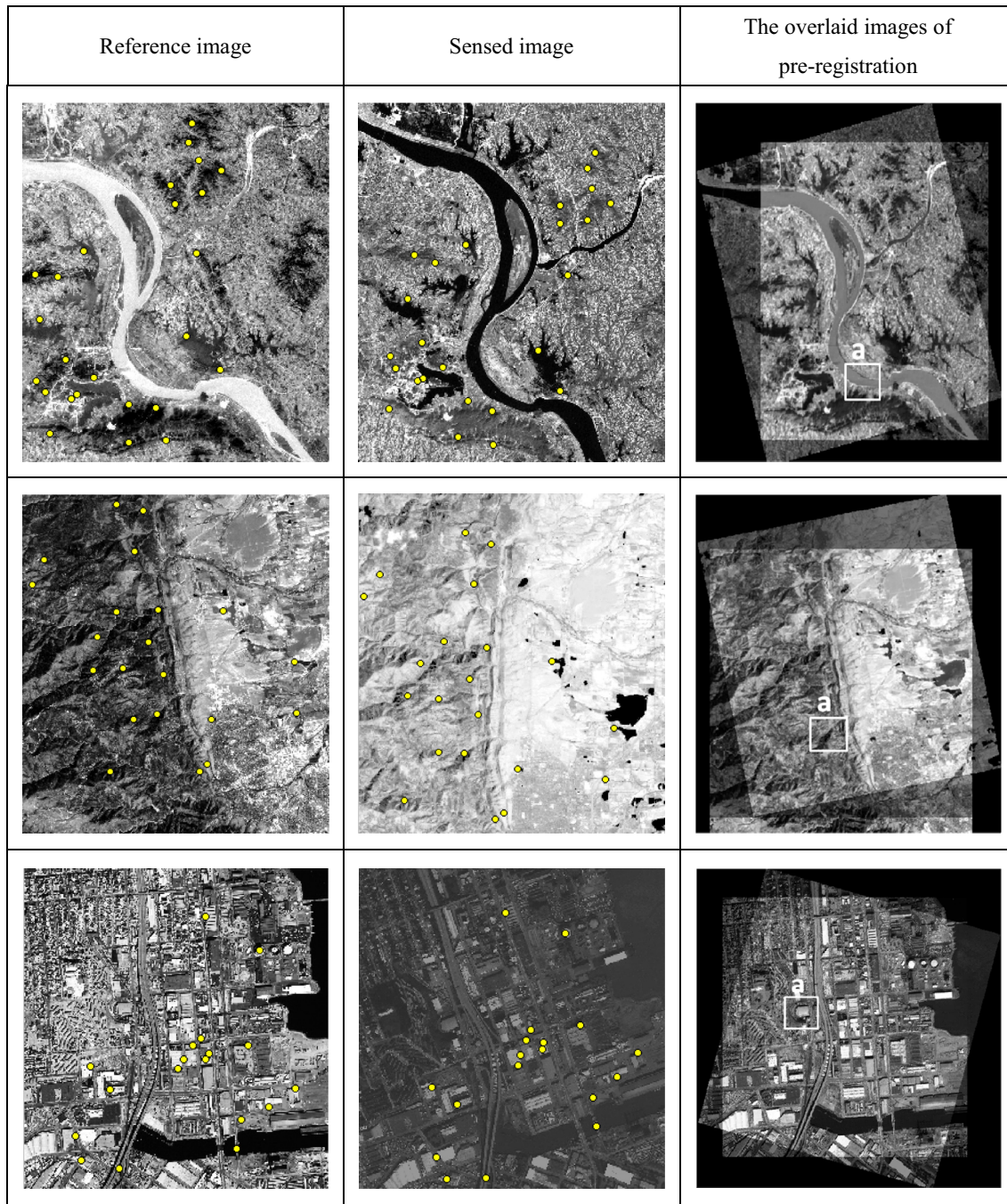


Fig. 8. The pre-registration results of all the test sets. The first, second and third rows correspond to test 1, test 2, and test 3, respectively.

sensitive to non-linear intensity differences between the images. However, the tie points detected by SR-SIFT are enough for the estimation of the projective transform, which can coarsely rectify the sensed image. The results of the pre-registration are shown in Fig. 8. One can see that the pre-registered image and the reference image have been coarsely aligned. No obvious translation, rotation, and scale differences can be observed between

Table 2
The number of tie points for pre- and fine registration.

Test	Pre-registration		Fine registration	
	SR-SIFT matches	SR-SIFT tie points	LSS matches	LSS tie points
Test 1	35	25	920	793
Test 2	33	21	831	720
Test 3	26	18	627	472

the images. Nevertheless, in the enlarged sub-images of pre-registration (Fig. 9), it can be observed that some misalignments still exist between the pre-registered image and the reference image for test 2 and test 3. This is because the images of the two tests contain the substantial local distortions caused by terrain relief, which cannot be handled very well by the projective transform. These distortions would be further dealt with in the following fine registration.

In the fine registration stage, 1500 even-distributed interest points were first extracted by the block-based Harris detector in the pre-registered image, and a small search window (20×20 pixels⁵) was determined in the reference image, where the matches

⁵ This refers to the moving range of the center pixel of template window.

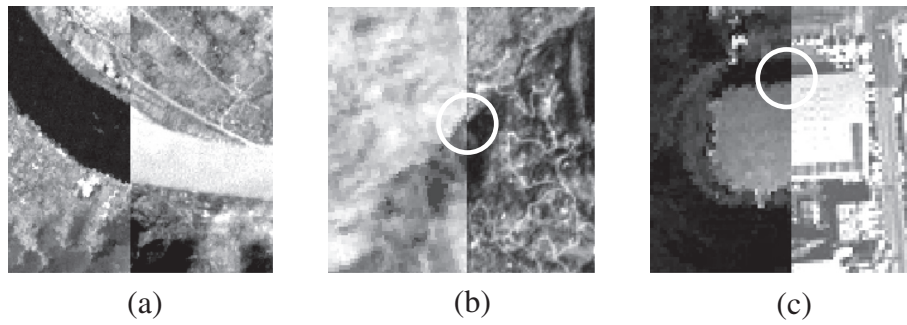


Fig. 9. The enlarged sub-images of the pre-registration for all the test sets. (a) test 1, (b) test 2, (c) test 3. These sub-images correspond to the boxes shown in “the overlaid images of pre-registration (Fig. 8)”, respectively. The misalignments between the pre-registered image and the reference image are marked by the circles for test 2 and test 3.

Reference image	Pre-registered image	The overlaid images of fine registration

Fig. 10. The fine registration results of all the test sets. The first, second and third rows correspond to test 1, test 2, and test 3, respectively.

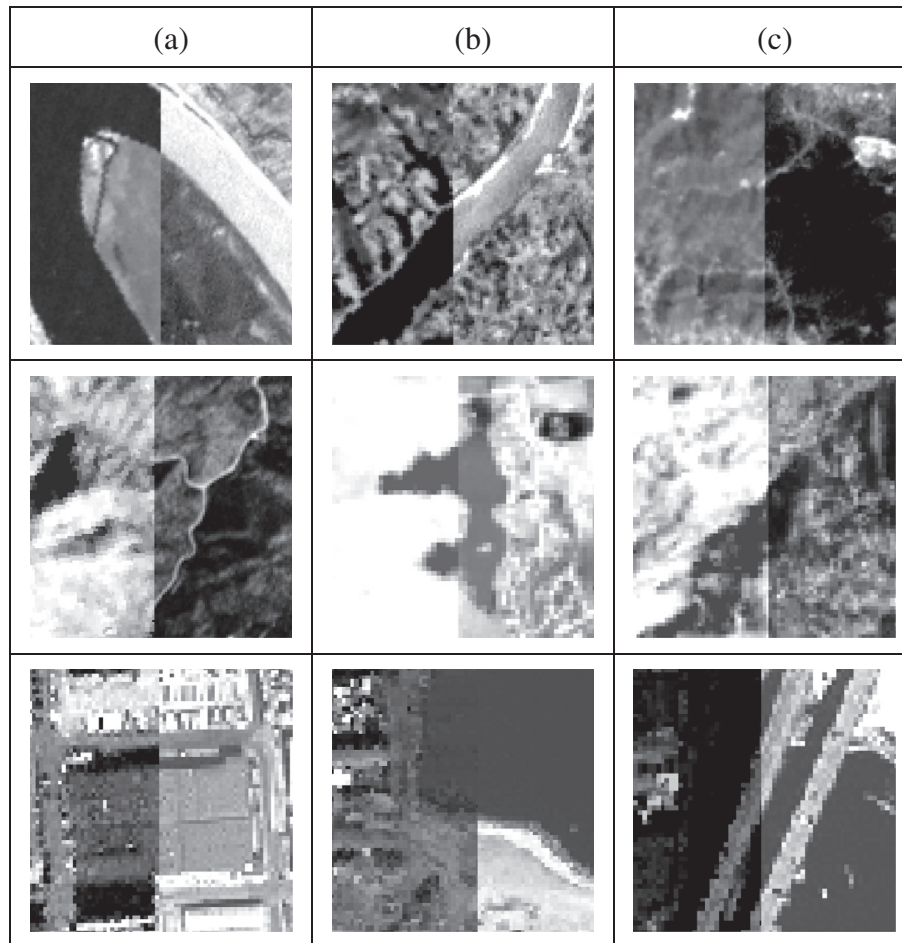


Fig. 11. The enlarged sub-images of all the test sets in the fine registration. The first, second and third rows correspond to test 1, test 2, and test 3, respectively. The first, second and third columns correspond to a, b, and c boxes shown in “the overlaid images of fine registration (Fig. 10)”, respectively.

were found by LSCC combined with the bidirectional matching technique. Then the tie points between the images were obtained after the mismatches were removed using the global consistency check. Finally, these tie points were used to construct the TIN, and the fine registration was achieved using the PL model. From “Fine registration” column in Table 2, one can be seen that LSCC detects a large number of tie points for all the test sets. The first two tests achieved the more tie points compared with test 3. The reason may be that the temporal differences of almost four years result in the changes of some areas between the images of test 3. However, test 3 achieved a considerable number of tie points for image registration. Fig. 10 shows that the tie points of each test are densely and evenly distributed over the images, which is very useful for the PL model to handle the local distortions caused by terrain relief. Based on the obtained results for three test sets in the fine registration stage, it has proven that LSCC is robust to the non-linear intensity differences between multispectral remote sensing images.

3.3. Accuracy analysis

Two different methods have been used to evaluate the registration accuracy of the proposed method. One is visual inspection which checks the overlay of the reference and the registered image. The other method is based on RMSE calculated from the manually selected check points between the images.

Fig. 11 shows that the three sub-images of each test are enlarged for visual inspection. The sub-images cover different surface scenes, such as the river bank, mountain, bridge, and building.

Generally speaking, the registered image fits very well with the reference image despite the significant non-linear intensity differences. However, it needs to be noted that the slight parallaxes around the building (the first sub-image of the third row in Fig. 11) are still observed due to the relief displacement of the buildings. This is a tough problem for high-resolution image registration, and it cannot be solved perfectly through an image-to-image registration method, unless a true orthorectification is applied (Hong and Zhang, 2008)

On the other hand, a set of 20 check points was manually selected between the reference image and the registered image for the three test data. The registration accuracy represented by RMSE is shown in Table 3.

Table 3 shows that the proposed method is successful in registering all the tests. However, they achieved different registration accuracies due to the differences in the resolutions and terrains. The registration accuracies of the first two tests are higher than that of test 3. The reason for such difference is that the data of the first two tests contain smaller local distortions caused by lower resolutions. Compared with test 1 covering a plain area, test 2 covering a mountain area achieves relatively lower accuracy because it has a larger terrain relief. As for test 3, its registration accuracy is the lowest among all the test sets. This is because of the test data being high-resolution images covering an urban area, where the relief displacement of the buildings leads to the significant local distortions. However, the RMSE of test 3 is 1.52 pixels, which is acceptable accuracy for a registration of high-resolution images covering an urban area. Overall, these results demonstrate that

Table 3

The registration accuracies of the three tests.

Test	No. of check points	RMSE (pixels)	Characteristics
Test 1	20	0.65	Medium-resolution plain area
Test 2	20	0.88	Medium-resolution mountain area
Test 3	20	1.52	High-resolution urban area

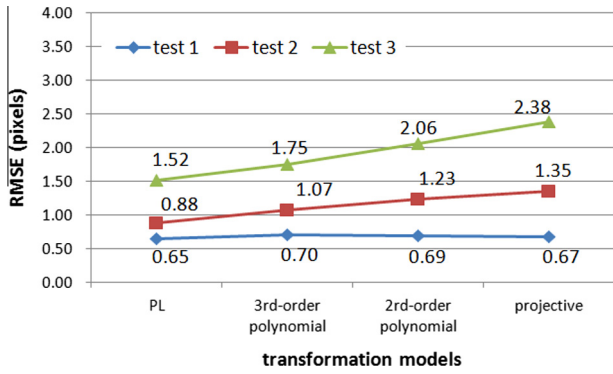


Fig. 12. The registration accuracy comparison of different transforms. The PL represents the piecewise linear transform.

the proposed method can obtain reliable registration accuracies for the multispectral remote sensing images with significant non-linear intensity differences.

To analyze the performance of the transforms, the registration accuracy using the PL model was compared with those of projective and polynomial models. The comparison results are shown in Fig. 12. We can find that the registration accuracy of the PL model is higher than that of others, especially for test 2 and test 3. This is because the PL model can handle the local distortions well when the tie points are densely and evenly distributed over the whole image (Arévalo and González, 2008). Additionally, it can be also seen that the projective model achieves the slightly higher accuracy than the second-order or third-order polynomial models for test 1, while the third-order polynomial model performs better than the projective and 2-order polynomial models for test 2 and test 3. This indicates that a projective model is more beneficial for medium-resolution images covering the plain area, whereas a 3-order polynomial is more suitable for images with terrain relief in the global consistency check.

3.4. Comparison of LSCC with NCC

As LSCC was used as a new similarity metric for tie-point detection in the proposed method, it was compared with NCC to verify

Table 4

The correct matches, the total matches, and the correct match rates of NCC and LSCC in different template sizes for all the three tests. CM: correct matches, TM: total matches, CMR: correct match rate.

Test	Methods	The CM, TM, and CMR(%) in different template sizes(pixels)									
		21 × 21		31 × 31		41 × 41		51 × 51		61 × 61	
		CM/TM	CMR	CM/TM	CMR	CM/TM	CMR	CM/TM	CMR	CM/TM	CMR
Test 1	NCC	449/739	60.8	462/747	61.8	506/777	65.1	512/778	65.8	514/780	65.9
	LSCC	569/782	72.8	707/868	81.6	793/920	86.2	829/938	88.4	844/947	89.1
Test 2	NCC	398/662	60.1	439/695	63.2	450/704	63.9	477/713	66.9	481/716	67.2
	LSCC	537/718	74.8	649/789	82.3	720/831	86.6	748/847	88.3	763/855	89.2
Test 3	NCC	261/493	52.9	301/524	57.4	327/549	59.6	334/554	60.3	337/557	60.5
	LSCC	345/528	65.3	422/585	72.1	472/627	75.3	500/638	78.4	509/643	79.2

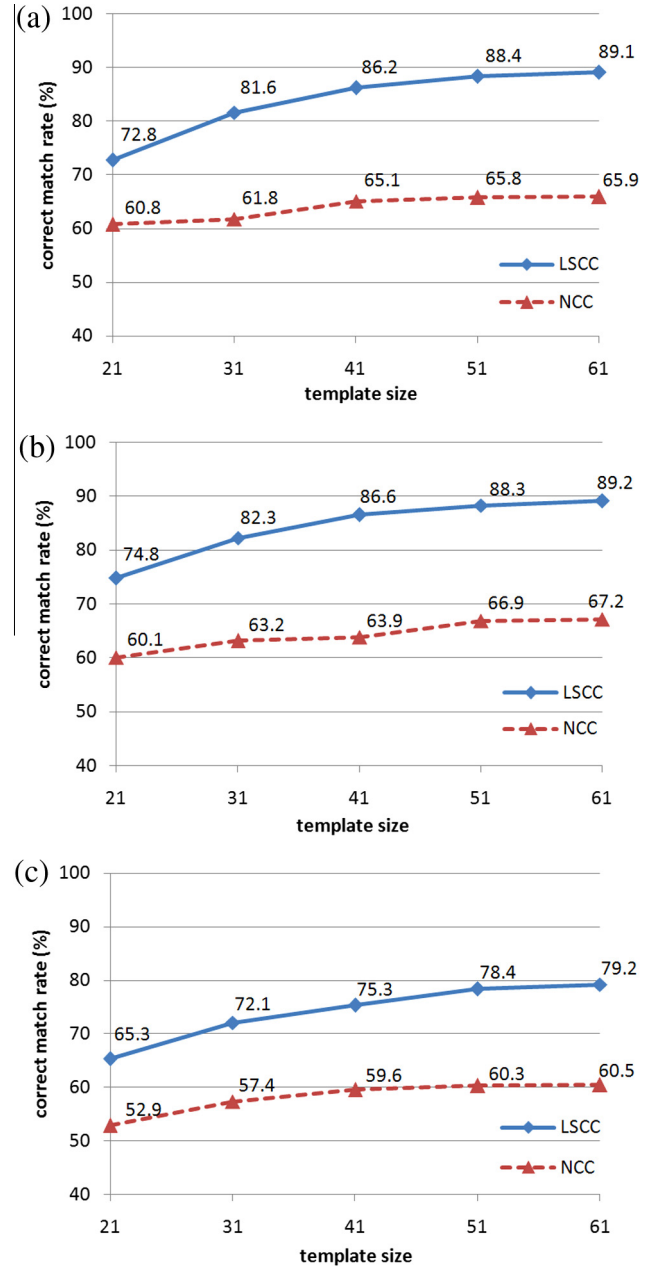


Fig. 13. The correct match rates of LSCC and NCC in the different template sizes for all the tests. (a) The correct match rate of test 1, (b) the correct match rate of test 2, (c) the correct match rate of test 3.

its ability to match multispectral remote sensing images. The comparison experiment was performed by detecting tie points in the fine registration of the three previous image pairs. First, 1500

Harris interest points were extracted in the pre-registration image. Then, the matches were obtained within a search window (20×20 pixels) of the reference image using LSCC and NCC combined with the bidirectional matching technique, respectively. To analyze the sensibilities of the two similarity metrics with respect to the template window size (hereafter referred to as template size), different template sizes for matching were used in the above procedure. The mismatches were removed by the global consistency check method. Table 4 lists the correct matches, the total matches, and the correct match rates⁶ of the two similarity metrics for all the three test sets. We use the correct match rate to evaluate the two similarity metrics because it directly reflects the matching performance. Fig. 13 shows the correct match rates of the two similarity metrics.

Fig. 13 shows that the correct match rates of LSCC are higher than that of NCC in all template sizes for the three tests. This finding indicates that LSCC is more discriminative than NCC in detecting tie points between the three multispectral image pairs. The reason is that NCC is only invariant to linear intensity changes and cannot handle complex intensity differences between images very well (Hel-Or et al., 2011). On the contrary, LSCC is based on the LSS descriptor, which captures the local shape property of images and is less sensitive to non-linear intensity differences. We can also observe from Fig. 13 that the correct match rates of these two similarity metrics rise with the increase of the template size. Compared with that of NCC, the correct match rate of LSCC grows at a faster ratio when the template size increases, especially between 21×21 pixels and 41×41 pixels. This finding may be attributed to the fact that a larger template size can contain more shape properties, which may improve the robustness of the similarity metric to some extent. However, although a large template size can improve the correct match rate of LSCC, the correct match rate has no significant improvement when the template size is beyond a certain range, such as 51×51 pixels. Furthermore, a large template size will increase the computational expense. Considering the correct match rate and the computational efficiency comprehensively, a proper template size for LSCC should be between 41×41 and 51×51 pixels for practical application. Overall, the above results demonstrate that compared with NCC, LSCC is a more reliable similarity metric for matching the multispectral remote sensing images with non-linear intensity differences.

4. Conclusions

Image registration is necessary step for combined utilization of the remote sensing images acquired at different spectra. Automatic registration of multispectral remote sensing images could be a challenging task due to the significant non-linear intensity differences. Conventional registration techniques have limitations for solving this problem. This paper presented an automatic image registration method for multispectral remote sensing images by using various techniques including the SR-SIFT algorithm, the Harris corner detector, the LSS descriptor, and the PL model. The proposed method involves two stages, namely, pre-registration and fine registration. In the first stage, the pre-registration is performed using SR-SIFT and a projective transform to eliminate the obvious translation, rotation, and scale differences between the reference and the sensed images. In the second stage, a set of Harris interest points are first extracted in the pre-registered image. Then, LSS is used as a new similarity metric (named LSCC) for tie-point detection by the bidirectional matching technique. Finally, the pre-registered image is rectified using the PL model.

The proposed method has been evaluated using three pairs of multispectral remote sensing images from different sensors, each

having significant non-linear intensity differences and geometric distortions. The experimental results demonstrate that the proposed method can achieve reliable registration accuracies for all the test data. However, there are some problems that need to be addressed further. It can be observed that SR-SIFT only achieves a small number of tie points in the pre-registration. The main reason is that SR-SIFT uses the local gradient information to build the feature descriptor, and the local gradient information is sensitive to non-linear intensity differences (Tsai et al., 2010). Further research could introduce the global structure information to enrich the feature descriptor for improving the matching performance. In the fine registration, though the PL model can handle the local distortions well with as many tie points as possible, it cannot effectively rectify the relief displacement from buildings for high-resolution images covering an urban region. A true orthorectification may be an effective approach to solve this problem. In addition, LSCC is compared with NCC in terms of the performance of the correct match rate. LSCC is found to perform better than NCC for the test data and is robust to non-linear intensity differences among multispectral remote sensing images. Nevertheless, we should note that the performance of LSCC may decline if the images do not include enough shape or structure information because LSCC depends on the LSS descriptor, which represents the local shape property of the image. In this case, an image enhancement technique can be applied to enhance the shape or edge feature, which may be of some help to image matching. In future work, a more detailed analysis for LSCC will be carried out using various multispectral remote sensing images.

Acknowledgement

This paper is supported by the National Basic Research Program (973 Program) of China (Nos. 2012CB719901 and 2012CB719904) and Southwest Jiaotong University Discipline Development Project. We are very grateful also for the constructive comments of the anonymous reviewers and members of the editorial team.

References

- Abdel-Hakim, A.E., Farag A.A., 2006. CSIFT: a SIFT descriptor with color invariant characteristics. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1978–1983.
- Arévalo, V., González, J., 2008. An experimental evaluation of non-rigid registration techniques on Quickbird satellite imagery. *Int. J. Remote Sens.* 29 (2), 513–527.
- Cole-Rhodes, A.A., Johnson, K.L., LeMoigne, J., Zavorin, I., 2003. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. *IEEE Trans. Image Process.* 12 (12), 1495–1511.
- Dai, X., Khorram, S., 1999. A feature-based image registration algorithm using improved chain-code representation combined with invariant moments. *IEEE Trans. Geosci. Remote Sens.* 37 (5), 2351–2362.
- Dare, P., Dowman, I., 2001. An improved model for automatic feature-based registration of SAR and SPOT images. *ISPRS J. Photogram. Remote Sens.* 56 (1), 13–28.
- Di Stefano, L., Marchionni, M., Mattocchia, S., 2004. A fast area-based stereo matching algorithm. *Image Vis. Comput.* 22 (12), 983–1005.
- Fan, X.F., Rhody, H., Saber, E., 2010. A spatial-feature-enhanced MMI algorithm for multimodal airborne image registration. *IEEE Trans. Geosci. Remote Sens.* 48 (6), 2580–2589.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24 (6), 381–395.
- Goshtasby, A., 1986. Piecewise linear mapping functions for image registration. *Pattern Recogn.* 19 (6), 459–466.
- Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: Proceedings 4th Alvey Vision Conference, Manchester, pp. 147–151.
- Hel-Or, Y., Hel-Or, H., David, E., 2011. Fast template-matching in non-linear tone-mapped images. In: Proceedings IEEE International Conference on Computer Vision, Barcelona, Spain, pp. 1355–1362.
- Hong, G., Zhang, Y., 2008. Wavelet-based image registration technique for high-resolution remote sensing images. *Comput. Geosci.* 34 (12), 1708–1720.
- Kelman, A., Sofka, M., Stewart, C.V., 2007. Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, pp. 3257–3263.

⁶ The correct match rate = the correct matches/the total matches.

- Kern, J.P., Pattichis, M.S., 2007. Robust multispectral image registration using mutual-information models. *IEEE Trans. Geosci. Remote Sens.* 45 (5), 1494–1505.
- Li, H., Manjunath, B., Mitra, S.K., 1995. A contour-based approach to multisensor image registration. *IEEE Trans. Image Process.* 4 (3), 320–334.
- Li, X., Zheng, L., Hu, Z., 2006. SIFT based automatic registration of remotely-sensed imagery. *J. Remote Sens.* 10 (6), 829–835.
- Li, Q.L., Zhang, H.S., Wang, T.F., 2011. Multispectral image matching using rotation-invariant distance. *IEEE Geosci. Remote Sens. Lett.* 8 (3), 406–410.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60 (2), 91–110.
- Ma, J.L., Chan, J.C.W., Canters, F., 2010. Fully automatic subpixel image registration of multiangle CHRIS/Proba data. *IEEE Trans. Geosci. Remote Sens.* 48 (7), 2829–2839.
- Morel, J.M., Yu, G., 2009. ASIFT: a new framework for fully affine invariant image comparison. *SIAM J. Imag. Sci.* 2 (2), 438–469.
- Schmid, C., Mohr, R., Bauckhage, C., 2000. Evaluation of interest point detectors. *Int. J. Comput. Vision* 37 (2), 151–172.
- Sedaghat, A., Mokhtarzade, M., Ebadi, H., 2011. Uniform robust scale-invariant feature matching for optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 49 (11), 4516–4527.
- Shechtman, E., Irani, M., 2007. Matching local self-similarities across images and videos. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis, USA, pp. 1–8.
- Torabi, A., Bilodeau, G.A., 2012. Local self-similarity based registration of human ROIs in pairs of stereo thermal-visible videos. *Pattern Recogn.* 46 (2), 578–589.
- Tsai, C.L., Li, C.Y., Yang, G.H., Lin, K.S., 2010. The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence. *IEEE Trans. Med. Imag.* 29 (3), 636–649.
- Tsai, V.J.D., 1993. Delaunay triangulations in TIN creation: an overview and a linear-time algorithm. *Int. J. Geogr. Inform. Sci.* 7 (6), 501–524.
- Wong, A., Clausi, D.A., 2007. ARRSI: Automatic registration of remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* 45 (5), 1483–1493.
- Wong, A., Clausi, D.A., 2010. AISIR: Automated inter-sensor/inter-band satellite image registration using robust complex wavelet feature representations. *Pattern Recogn. Lett.* 31 (10), 1160–1167.
- Yi, Z., Cao, Z.G., Yang, X., 2008. Multi-spectral remote image registration based on SIFT. *Electron. Lett.* 44 (2), 107–108.
- Yu, L., Zhang, D.R., Holden, E.J., 2008. A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. *Comput. Geosci.* 34 (7), 838–848.
- Zitova, B., Flusser, J., 2003. Image registration methods: a survey. *Image Vis. Comput.* 21 (11), 977–1000.