A NOVEL FINE REGISTRATION TECHNIQUE FOR VERY HIGH RESOLUTION REMOTE SENSING IMAGES

*Xianzhang Zhu[1], Yongjun Zhang[*1], Hui Cao[1], Kai Tan[2] and Xiao Ling[1]*

[1]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[2]Wuhan Research Institute of HUAWEI Technology Co., Ltd., Wuhan 430200, China

ABSTRACT

This paper presents a novel registration noise (RN) estimation technique for fine registration of very high resolution (VHR) images. This is accomplished by using a two-step strategy to estimate and mitigate residual local misalignments in standardly registered VHR images. The first step takes advantages of the superpixel segmentation and frequency filtering to generate sparse superpixels as the basic objects for RN estimation. Then local rectification is employed for fine registration of the input image under the aid of RN information. More factors are taken into consideration in order to enhance the RN estimation performance. The proposed approach is designed in a fine registration strategy, which can effectively improve the pre-registration result. The experimental results obtained with real datasets confirm the effectiveness of the proposed method.

Index Terms—Fine registration, VHR image, superpixel segmentation, sparse representation, local rectification.

1. INTRODUCTION

Image registration is the process of overlaying two or more images of the same geographical area taken at different times, viewpoints, and modalities or by different sensors. Accurate registration of images is a prerequisite of many remote sensing applications, such as non-supervision change detection, image mosaicking, and image fusion. In the past decades, intensive inventions on image registration methods have greatly improved the registration accuracy. Feature-based methods have been widely used to extract salient features and conduct feature-matching. Paul *et al.* [1] proposed a modified uniform robust SIFT to obtain uniformly distributed matched features and maximize the number of matches. Ma *et al.* [2] proposed a two-step non-rigid automatic registration scheme by using SIFT and NCC. In addition, methods based on image segmentation techniques have been developed gradually in recent years. Goncalves *et al.* [3] developed an effective method by combining image segmentation and SIFT to eliminate areas that are not suitable for control point detection and object level spatial attribute extraction. Ling *et al.* [4] presented a novel image matching method for multi-sources satellite images, which integrates the global Shuttle Radar Topography Mission (SRTM) data and image segmentation method to achieve robust and numerous correspondences.

However, it is difficult to completely eliminate the control point errors even with the support of gross error elimination technique because of the complexity of terrains in remote sensing images. A large number of false matching between the input image and reference image influenced the construction of transformation model and lead to local nonlinear geometric distortions. After the completion of the image registration process, visual interpretation and rolling shutter tools were always used to check the quality of registration results. Bruzzone *et al.* [5] proposed a change vector analysis (CVA) based approach to reduce the effects of RN in unsupervised change detection. Han *at al.* [6] further studied the reduction of local residual misalignment and proposed a segmentation-based fine registration (SBFR) approach. In their method, multitemporal images were assumed to be standardly registered, in which residual misalignments still existed. They estimated the geometric distortion of RN and used a piecewise linear function to achieve accurate and precise geometric alignment. However, they applied SBFR only to the homologous data sets and in small regions with rich texture. However, it is difficult to detect exact object representative points in repetitive or homogeneous textural images [6], which further influenced the estimation of residual local misalignment.

In this paper, we propose a novel fine registration approach that uses superpixel RN estimation in VHR multitemporal and multisensor images, which are also assumed to have been standardly registered. In this approach, the RN estimation is more efficient and reliable by using sparse represented superpixels and coefficient balance. What's more, it obtains accurate control points by optimizing the pre-registration matching result, where the gross error information and correlation coefficients between the input image and the reference image are fully considered. Finally, the newly generated control points are used to calculate the local transformation model and then employed to fine register the pre-registered VHR image.

The mathematical strategy of the proposed approach is described in the next section, followed by experimental demonstration of the proposed approach by using two data sets generated from Chinese GF2, GF1, and ZY3 satellites.

2. PROPOSED METHOD FOR FINE REGISTRATION IN VHR IMAGES

As shown in Fig.1, the proposed approach consists of two steps: 1) estimating local RN distribution and 2) re-rectification of the registered image. For the first step, the superpixels are segmented from the input pre-registered VHR image in order to calculate a frequency map of the input image. The sparse superpixels are generated by filtering each superpixel in the frequency domain, which are used to estimate the superpixel RN distribution with spatial correlation contrast. For the second step, superpixel RN distribution is employed to optimize the control points obtained in the pre-registration process for building the local transformation model. And the final fine registered VHR image is generated by local rectification.

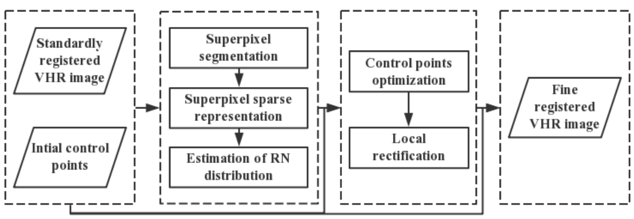

Fig. 1. Flowchart of the proposed fine registration approach

2.1. Estimation of local RN distribution

The control point error is random in the global range of image, which is influenced by the image distortion and complexity of ground objects. Thus, the deviation between the registered image and the reference image shows different intensity and different direction in the local area. In this paper, simple linear iterative clustering (SLIC) [7] is employed to extract superpixels as effective basic-objects to calculate the local misalignments between the two images, because it is computationally efficient and can generate superpixels compactly with uniform size and well adhered region boundaries. The setting of superpixel size has a very important impact on the accuracy and efficiency of the estimation process. It is determined by the initial width S that is assigned in advance of the segmentation process. The segmentation results of different S values are shown in Fig. 2. As one can see, when the S value is too small, the superpixel may contain only one kind of ground object, which can lead to a lack of saliency and thus affect the rough location recognition. On the other hand, it is difficult to obtain precise position due to the large number of pixels and the complex changes of ground objects when the S value is too big. To this end, the deviations between the registered image and the reference image are distinguished by visual interpretation. The superpixel width was set at 400 in this paper because this number is more likely to have enough significant features to support the spatial correlation analysis.

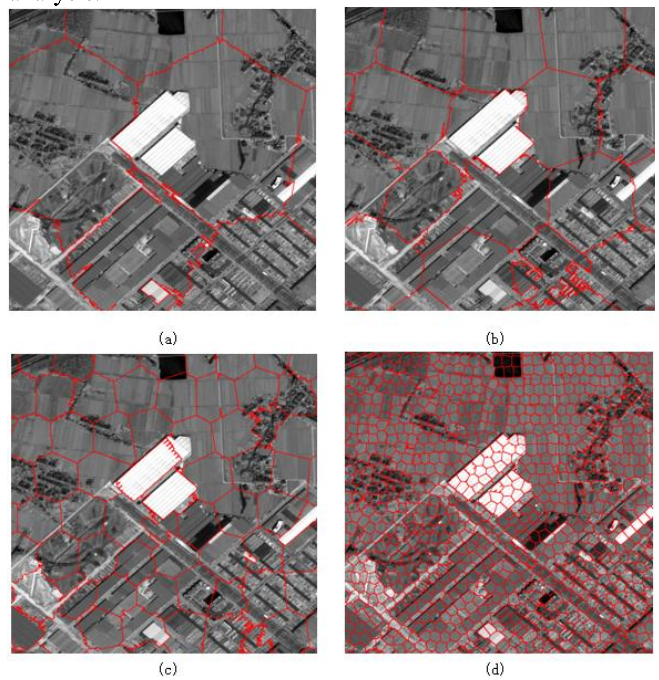

Fig. 2. Different superpixel segmentation results when S is set at 800 (a), 400 (b), 100 (c) and 50 (d) pixels, respectively.

Then, we exploit the difference-between-original-and-prediction (DBOP) filter, which is a step of the Laplacian pyramid, to generate a bandpass frequency map of the pre-registered image while avoiding scrambling frequencies. The bandpass frequency map generated by DBOP filter is defined as

$$X_F = |X - \mathrm{M}_{\uparrow}(\mathrm{M}_{\downarrow}(X \otimes g_{5\times5})) \otimes g_{5\times5}| \quad (1)$$

where X represents the input registered image, $g_{5\times5}$ represents the 5×5 window Gaussian filter kernel, and M is the sampling matrix, which is used to obtain the prediction image by downsampling and upsampling steps. Each superpixel specifies at least 50% of the pixels by calculating an adaptive frequency threshold, which is defined as follows:

$$T_F^n = \min\{t | \textstyle\sum_{i=1}^{t} P_{X_F}^n(i) > 0.5,\ t \in [0, X_{F_{max}}]\} \quad (2)$$

where i is the value in X_F, n is the label of the superpixel, t is the current cumulative X_F value, and $P_{X_F}^n(i)$ denotes the frequency ratio of i calculated by the histogram statistic. Then, threshold segmentation is employed to eliminate pixels that have a value of less than T_F^n in each superpixel. Letting (x,y) be the spatial positions of the samples, the sparse superpixel is defined as follows:

$$SS^n(x,y) = \begin{cases} X(x,y) & if\ X_F(x,y) \geq T_F^n \\ \emptyset & else \end{cases} \quad (3)$$

Since RN represents the drift of homonymous pixels between multitemporal images, a spatial correlation measure can be employed to estimate it. Here we use a sparse superpixel as the basic-object and set a $w \times h$ search window for supporting the multiple displacement analysis in the reference image, where w and h are the width and height of the displacement range, respectively. In addition, the corresponding projection pixel value on the reference image is obtained by bilinear interpolation resampling. Supposing that the number of sparse superpixels in the registered image is N, for each object sparse superpixel $SS^n(n = 1, \ldots, N)$, the maximum correlation location of the reference image is searched. This can be regarded as the actual homonymous location and has two components $\{u^n, v^n\}$ within the search window coordinate system. However, it can be further improved to the subpixel level by coefficient balancing.

2.2. Re-rectification of the registered image

As mentioned in the previous registration process, the control points are obtained by the image matching algorithm introduced in [4], and their gross errors are calculated by RANSAC-based mismatch detection. For each RN estimated superpixel, the vector of misalignment is calculated. This information can be used to adjust the coordinates of the control points that are within this superpixel range. For each control point pair, adjustment is applied to the transformed control point coordinates on the pre-registered image while the control point coordinates on the reference image are maintained. By combining the initial control points and the superpixel RNs, the detailed optimization procedure is listed as follows.

a) Selection of estimated noise superpixels: Since the optimization is aimed at the poorly registered area and the RN vector is difficult to accurately estimate in the regions with low correlation, not all the superpixels are available for optimization. RN intensity threshold T_ρ and maximum correlation coefficient threshold T_c are set to classify the superpixels into the non-selected and selected groups with the following formula:

$$SS_U^n = \begin{cases} 0 & if\ \rho^n < T_\rho \ \&\&\ C^n < T_c \\ 1 & else \end{cases} \tag{4}$$

where ρ^n is the intensity of the superpixel RN, and C^n denotes the maximum correlation coefficient.

b) Selection of effective control points: For the control points that fall into the selected superpixels, some are matched correctly, and those whose errors are greater than the index threshold are selected as the effective control points for optimization.

c) Adjustment of control points: For the selected control points on the pre-registered image, the new coordinates are obtained according to the estimated RN information of the superpixel in which they are located. These coordinates are considered as new control points that still correspond to the maintained control points on the reference image.

Finally, local affine transformation based on triangulated irregular network (TIN) is employed to mitigate the local residual misalignments in the pre-registered image. For each of the triangles, the triangulated corresponding control points in the two images are used to calculate a warping function. Then, affine transformation is applied to the local rectification from the registered image to the reference image. The fine registered image is generated after all the triangles are rectified.

3. EXPERIMENTAL RESULTS

The proposed method was tested on two multi-temporal VHR remote sensing image data sets. The first data set (DS1), acquired over the region of Wuxi City, China, contained two temporally different panchromatic images taken by GF-2 with a spatial resolution of 0.8 m. The reference image of DS1 was taken in April 2015, whereas the input image was taken in March 2016. A subset of 6000*6000 pixels that contains buildings, roads, farmland and water area was considered in the experiment. The second data set (DS2) is a set of multi-source panchromatic images that are located at Huizhou City, China. The reference image of DS2 was taken by ZY-3 with a spatial resolution of 2.1m in August 2015, whereas the input image was taken by GF-1 with a spatial resolution of 2m in January 2017. Similarly, the above used subsets were both 6000*6000 pixels and contain mountains, vegetation, urban areas, and water areas. After completing the proposed fine registration process, superpixel RN estimation was performed on the fine registered image once again.

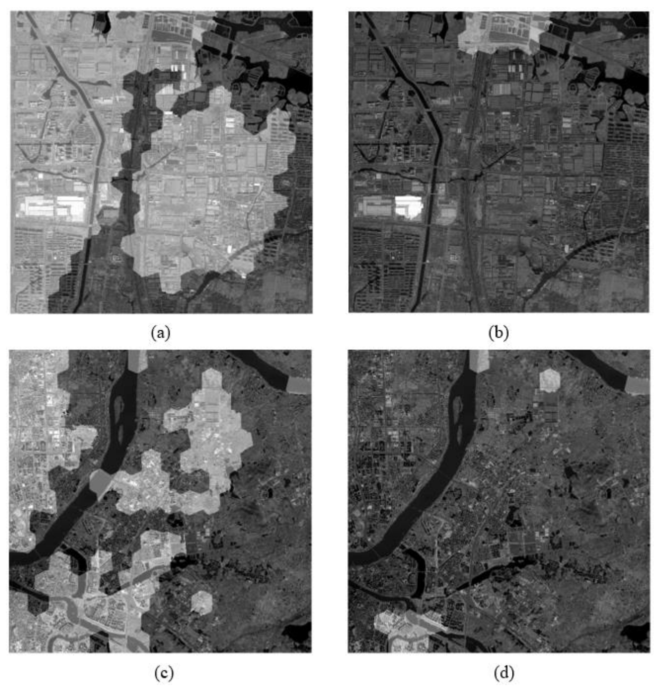

Fig.3. The RN distribution superposition diagram of two data sets. (a) and (b) indicate the RN distributions of the GF-2 image before (a) and after (b) applying the proposed approach; (c) and (d) indicate the RN distributions of the GF-1 image before (c) and after (d) applying the proposed approach.

The RN distribution superposition diagram generated before and after applying the proposed fine registration on the two data sets is shown in Fig. 3. The estimated noise superpixels correspond to the light regions, whereas the non-noise superpixels correspond to the dark regions. It can be seen that the sizes of the noise regions after applying the proposed approach [Fig.3 (b) and (d)] were less than the pre-registration results [Fig. 3 (a) and (c)]. As expected, the overall registration accuracy in the two data sets was effectively improved. For further visual comparison, Fig. 4 displays four groups of detailed checkerboard images generated from the two experimental data sets. The red circles in Fig. 4 highlight some of the edge objects that are easy to evaluate the registration results. The misalignments generated from the pre-registration process [Fig. 4 (a-d)] were effectively eliminated in the fine registered images generated by the proposed method [Fig. 4 (e-h)], and thus the edge objects were precisely aligned.

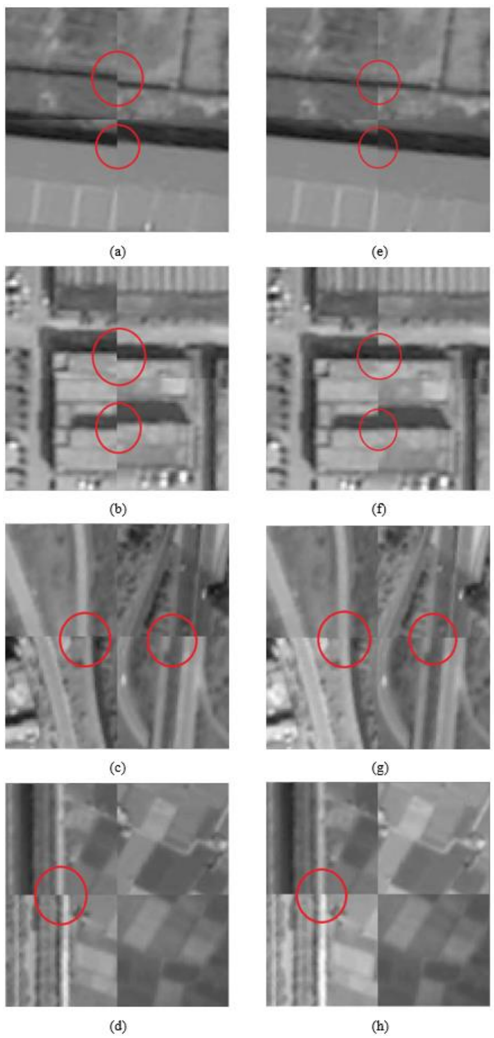

Fig.4. Detailed checkerboard images generated from two data sets. (a),(b) (from DS1) and (c),(d) (from DS2) are the pre-registered local subscene, whereas (e),(f) and (g),(h) are the corresponding subscene after applying the proposed fine registration approach.

4. CONCLUSION

This paper proposed a novel fine registration approach for VHR images, which is capable of estimating and mitigating the local residual misalignments by local re-registration. Based on the assumption that the input image is standardly registered, the proposed method uses a two-step strategy to fine register the pre-registered image. The proposed method firstly employs superpixel segmentation and designed frequency filtering to generate sparsely represented superpixels to robustly preserve the edge pixels and improve the RN estimation process. Then, coefficient balancing is performed to further improve the precision of the RN value, which provides an optimization reference for the initial control points to improve their quality. Finally, the fine-registered VHR image is generated by local rectification transformation, which is calculated from the newly optimized control points. The preliminary results showed that the superpixel RN estimation based fine registration can be effectively used to mitigate local residual misalignments in VHR images. How to generate more robust estimation objects to adapt more image scenes will be investigated in the future.

5. ACKNOLOWDGMENT

This work was supported in part by the Natural Science Foundation of China under grant 41571434 and 41322010.

6. REFERENCES

[1] S. Paul and U. C. Pati, "Remote Sensing Optical Image Registration Using Modified Uniform Robust SIFT," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 9, pp. 1300-1304, 2016.
[2] J. Ma, J. C. Chan, and F. Canters, "Fully Automatic Subpixel Image Registration of Multiangle CHRIS/Proba Data," *IEEE Transactions on geoscience and remote sensing*, vol. 48, no. 7, pp. 2829–2839, 2010.
[3] H. Goncalves, L. Corte-Real and J. A. Goncalves, "Automatic Image Registration Through Image Segmentation and SIFT," *IEEE Transactions on Geoscience and Re€mote Sensing*, vol. 49, no. 7, pp. 2589–2600, 2011.
[4] X. Ling, Y. Zhang, J. Xiong, X. Huang, and Z. Chen, "An Image Matching Algorithm Integrating Global SRTM and Image Segmentation for Multi-Source Satellite Imagery," *Remote Sensing*, vol. 8, no. 8, pp. 672, 2016.
[5] L. Bruzzone, and R. Cossu, "An Adaptive Approach to Reducing Registration Noise Effects in Unsupervised Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 11, pp. 2455-2465, 2003.
[6] Y. Han, F. Bovolo and L. Bruzzone, "Segmentation-Based Fine Registration of Very High Resolution Multitemporal Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2884-2897, 2017.
[7] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC Superpixels Compared to State-of-the-art Superpixel Methods," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274-2282, 2012.