ROBUST REGISTRATION FOR OPTICAL AND SAR IMAGES BASED ON SPATIAL GEOMETRIC CONSTRAINT AND STRUCTURE FEATURES

Mengmeng Wang¹, Yuanxin Ye^{1*}, Miaomiao Sun¹, Xin Tan¹, Liang Li²

¹Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610031, China

²Third Engineering Surveying and Mapping Academy in Sichuan Province, Chengdu 610500, China

Commission I, WG I/6

KEY WORDS: image registration, optical image, SAR image, spatial geometry constraint, structural feature, 3-D phase correlation

ABSTRACT:

Automatic registration of optical and synthetic aperture radar (SAR) images is a challenging task due to significant geometric deformation and radiometric differences between two images. To address this issue, this paper proposes an automatic registration method for optical and SAR images based on spatial geometric constraint and structure features. Firstly, the Harris detector with a block strategy is used to extract evenly distributed feature points in the images. Subsequently, a local geometric correction is performed by using the Rational Function Model, which eliminates the rotation and scale differences between optical and SAR images. Secondly, orientated gradient information of images is used to construct a geometric structural feature descriptor. Then, the feature descriptor is transformed into the frequency domain, and the three-dimensional (3-D) phase correlation is used as the similarity metric to achieve correspondences by employing a template matching scheme. Finally, mismatches are eliminated based on spatial geometric constraint relationship between images, followed by a process of geometric correction to achieve the image registration. Experimental results with multiple high-resolution optical and SAR images show that the proposed method can achieve reliable registration accuracy, and outperforms the state of the art methods.

1. INTROCTION

During the last decades, remote sensing systems and imaging techniques have undergone a rapid development, various multimodal remote sensing images (such as synthetic aperture radar (SAR), infrared, visible) can be acquired. These remote sensing images can be widely used in many applications, including change detection (Bruzzone and Bovolo, 2013), image fusion (Schmitt et al., 2017), agriculture and forestry monitoring (Rhee et al., 2010), and land-use analysis (Tuia et al., 2016). Combined utilization of multimodal remote sensing images can overcome some limitations of single-modal images. SAR and optical images are acquired by various sensors under different imaging mechanisms, i.e., the active SAR sensor and the passive optical sensor. SAR images are capable of delivering information in both day and night and under all-weather conditions because of its high penetration, but they have serious inherent speckle noise. Optical images have good interpretability, but they are easily affected by atmosphere condition such as clouds and fogs (Sui et al., 2015). Since the two types of images can provide complementary information, it is necessary to integrate SAR and optical images for following remote sensing applications. Therefore, automatic registration of optical and SAR images has still been a hot research spot.

In general, automatic image registration method can be generally classified into two categories: intensity-based and feature-based methods (Zitova and Flusser, 2003). Intensity-based methods estimate the optimal transformation parameter between images by control points (CPs), which are detected by using some similarity metrics. Commonly adopted similarity metrics include the mutual information (MI) (Suri and Reinartz, 2010) and the normalized cross correlation (NCC), NCC is sensitive to nonlinear radiometric difference between optical and SAR

images, which limits its application in optical and SAR image registration. Compare with NCC, MI can adapt to nonlinear radiometric difference between images, but MI has the disadvantage of large computation. Feature-based method estimate the geometric transformation by matching distinct features between images. These features include points, lines and contours (Ma et al., 2017; Li et al., 1995). Recently, local invariant features, such as scale-invariant feature transform (SIFT) and speeded up robust features (SURF), have been widely used for remote sensing image registration due to its invariance to image scale and rotation changes (Sedaghat et al., 2011; Patel et al., 2016). Although these features are robust to illumination changes and geometric distortions between optical images, they are vulnerable from the nonlinear radiometric differences between optical images and SAR images (Dare and Dowman, 2001).

Recently, Ye et al. proposed a robust feature descriptor based on the structural properties of images, which is named the histogram of orientated phase congruency (HOPC) (Ye et al., 2017). Next, Ye et al. proposed a novel pixel feature representation by using orientated gradients of images, which is named channel feature of orientated gradients (CFOG) (Ye et al., 2019). These feature descriptors showed a robust ability to handle the significant nonlinear radiometric differences between optical and SAR images. Nevertheless, these feature descriptors are high-dimensional features with large volume data, and not invariant for scale and rotation differences. In fact, scale and rotation differences between optical and SAR images also are the factors that affect the registration performance.

Currently, deep learning-based methods have been used to multimodal remote sensing image registration. Merkle et al., used a Siamese network to learn the spatial shift between optical and

* Corresponding author: Yuanxin Ye, yeyuanxin@swjtu.edu.cn

SAR image patches (Merkle et al., 2017). Girard et al., proposed a deep learning method to improve the registration performance by training a multi-task learning (Girard et al., 2018). Subsequently, Bürgmann designed a descriptor learned by a deep neural network named HardNet, which can automatically detect CPs between SAR and optical images (Bürgmann et al., 2019). In general, optical and SAR image registration by deep learning-based has become an active research direction in recent years. However, one drawback of the current methods is the restriction of large training datasets, and making the training datasets is a tedious task.

In order to address the above-mentioned issues, this paper proposes a robust and automatic optical-to-SAR remote sensing image registration method based on spatial geometry constraint and structural features. First, local scale and rotation differences between optical and SAR images are eliminated by using Rational Function Model (RFM) of remote sensing images. Second, the matching region is predicted based on the space geometric constraint, and a structure feature descriptor is built by using the orientated gradient information of images. Then, a fast matching similarity metric based on the structure feature and three-dimensional (3-D) phase correlation is established in the frequency domain, and a template matching manner is employed to detect CPs between optical and SAR images. Finally, mismatches are removed based on the spatial geometric constraint relationship, and the transform parameters for image registration are obtained by using the correct CPs.

2. METHODOLOGY

The flowchart of the proposed method is shown in Figure 1, which consists of: 1) feature point extraction; 2) local geometric correction; 3) structural feature extraction; 4) Similarity metric based on 3-D phase correlation; 5) mismatch elimination and geometric transformation.

Figure 1. Flowchart of proposed method

2.1 Feature Point Extraction

In order to extract the evenly distributed feature points in the optical image, we use the block-based Harris corner detector (Ye

and Shan, 2014) to extract feature points. Firstly, the optical image is divided into $M \times M$ non-overlapping blocks. Then, the Harris value of each pixel is computed and ranked from largest to smallest in every block, and the point with the largest value is selected as the feature point. Figure 2 shows the feature points extracted by the traditional Harris corner detector and the block-based Harris corner detector. It can be clearly observed that the block-based Harris corner detector can make the feature points evenly distribute over the image.

Figure 2. The Harris feature points extraction results by (a) fixed threshold on image and (b) non-overlap block-based scheme

Figure 3. (a) Process of local geometric correction. (b) Comparison of the optical image before and after correction. Optical image before correction (left), Optical image after correction (middle), SAR image (right)

2.2 Local Geometric Correction

In order to eliminate the rotation and scale differences between the optical and SAR images, a local geometric correction is performed for an image patch centered on a feature point by using RFM (Grodecki and Dial, 2003; Wang et al., 2017). The RFM relates 3D object-space ($D_{lon}, D_{lat}, D_{hei}$) ground coordinates to 2D image-space (Line, Sample) through a ratio of cubic polynomials. In order to improve the numerical stability, the object coordinates and image coordinates are normalized to $(-1, +1)$. The equation of RFM is expressed as follows:

$$\begin{cases} r = \frac{p_1(X_n, Y_n, Z_n)}{p_2(X_n, Y_n, Z_n)} \\ c = \frac{p_3(X_n, Y_n, Z_n)}{p_4(X_n, Y_n, Z_n)} \end{cases} \quad (1)$$

Where (r, c) is the normalized image coordinates, and (X_n, Y_n, Z_n) is the normalized ground coordinates. p_1, p_2, p_3 and p_4 are the terms of third-order polynomials of (X_n, Y_n, Z_n) .

In the processing of local geometric correction, an image patch centered on a feature point is first determined in the optical image, and a digital elevation model (DEM) is used as the elevation datum. Then, the RFM is used to perform the geometric rectification for the image patch. Figure 3 shows the above process. We can see that the local correction can effectively eliminate the rotation and scale differences between optical and SAR images.

2.3 Structural Feature Extraction

As mentioned above, RFM can remove obvious geometric distortions (such as rotation and scale differences) between optical and SAR images. Accordingly, the main difficulty of image registration is significant nonlinear radiometric differences between optical and SAR image. Considering the similar geometric structural properties between them, we extract the structural properties of images by using a robust structural feature descriptor named CFOG. The CFOG feature descriptor is an extension of histogram of orientated (HOG) (Dalal and Triggs, 2005), which can capture geometric structure features of images, and CFOG is robust to nonlinear radiometric differences between images. Figure 4 shows the construction process of CFOG, which mainly includes two parts: orientated gradient channels and 3-D Gaussian convolution.

Figure 4. Process chain of CFOG

2.3.1 Orientated Gradient Channel: Given an inputted template image, we first compute m orientated gradient channels of the image, which are called as $g_i, 1 \leq i \leq m$. $g_0(x, y)$ is the gradient magnitude of quantized orientation θ at (x, y) of each orientated gradient channel, which is larger than zero, otherwise its value is zero. Then, the orientated gradient channel is defined as:

$$g_o = \left\lfloor \frac{\partial I}{\partial \theta} \right\rfloor \quad (2)$$

Where I is the image, θ is the orientation of the derivative, and $\lfloor \cdot \rfloor$ denotes that the enclosed quantity is equal to itself when its value is positive or zero otherwise. In the actual process, we are

not necessary to separately compute g_o for each orientation, the horizontal and vertical gradients (named g_x and g_y , respectively) can be used to compute the magnitude of each orientated gradient channel. as follows:

$$g_o = \text{abs}(\cos \theta * g_x + \sin \theta * g_y) \quad (3)$$

Where θ is the quantized gradient orientation, g_x and g_y are separately computed by the filters $[-1, 0, 1]$ and $[1, 0, 1]^T$ with the two 1-D convolutions. abs represents the absolute value, which can handle the intensity inverse between multimodal remote sensing images by modifying the gradient orientation to the range of $[0^\circ, 180^\circ)$ (Yi et al., 2008).

2.3.2 3-D Gaussian Convolution: Once the orientated gradient channel is generated, the convolved feature channel can be obtained by a convolution with a 3-D Gaussian-like kernel g_σ , as shown below:

$$g_o^\sigma = g_o * \left[\frac{\partial I}{\partial \theta} \right] \quad (4)$$

Where σ represents the standard deviation of the Gaussian convolution kernel. Strictly speaking, the Gaussian convolution kernel is not a 3-D Gaussian function in the 3-D space, but a two-dimension Gaussian kernel in X- and Y-directions and a kernel of $[1, 2, 1]^T$ in the gradient orientation direction (hereinafter referred to as Z-direction). The gradient in the orientation direction is smoothed by convolution in the Z-direction, which can reduce the influence of orientation distortions caused by local geometric deformation and radiometric variation between optical and SAR images. Finally, the g_o^σ is normalized to form the orientated gradient channel feature. From figure 5, we can see that the two visualized 3-D CFOG representation of optical and SAR images, which have the high similarity despite significant radiometric differences between the optical and SAR images.

Figure 5. (a) optical image. (b) SAR image. (c) 3-D structural feature representation of optical image. (d) 3-D structural feature representation of SAR image

2.4 Similarity Metric Based on 3-D Phase Correlation

Since CFOG is a 3-D structural feature descriptor with a large-volume data. It is time consuming for image matching by using traditional similarity metrics (such as NCC) in the space domain. Accordingly, we build a fast similarity metric based on CFOG in

the frequency domain because the dot product operation in the frequency domain is equal to the convolution operation in the space domain.

According to the shift property of Fourier transform, the shift in the spatial domain of two functions corresponds to the linear phase difference in the frequency domain. 3-D phase correlation is used as the similarity metric because the CFOG descriptor is a 3-D feature volume. Let $f_1(x, y, z)$ and $f_2(x, y, z)$ be two functions of the optical and SAR image feature volumes respectively. Since the scale and rotation differences has been eliminated by the local geometric correction, there is only a shift relationship between two feature volumes, which is $f_1(x, y, z) = f_2(x - x_0, y - y_0, z)$. According to the shift property of Fourier transform, the function can be written as:

$$F_1(u, v, w) = F_2(u, v, w) \exp(-2\pi i(ux_0 + vy_0)) \quad (5)$$

Where F_1 and F_2 are the 3-D Fourier transform of f_1 and f_2 . Then, by computing the normalized cross-power spectrum, we have:

$$Q(u, v, w) = \frac{F_1(u, v, w)F_2(u, v, w)^*}{|F_1(u, v, w)F_2(u, v, w)^*|} = \exp(-2\pi i(ux_0 + vy_0))\vec{k} \quad (6)$$

Where $*$ denote the complex conjugate, and $\vec{k}$ is a 3-D unit vector. Noting that the inverse Fourier transform of $\exp(-2\pi i(ux_0 + vy_0))$ is a Dirichlet function at (x_0, y_0) , we can determine the (x_0, y_0) by finding the pulse peak in the inverse Fourier transform (Figure 6). Therefore, we can obtain the translation shift between the two feature representation volumes.

3. EXPERIMENTS

In this section, we validate the performance of the proposed method on three sets of optical and SAR images (see Figure 7) by comparing the other two state-of-the-art intensity-based

methods (i.e., NCC and MI). In order to make a fair comparison, both NCC and MI are applied to image registration with a process of local geometric correction described in Section 2.2. The details of test sets are listed in Table 1. In our experiments, the SAR images are used as the reference image because they have been ortho-rectified, and the optical images are used as the sensed image.

Figure 6. The cross-power spectrum matrix of feature

3.1 Evaluation Criteria and Parameter Setting

We analyse the performance of the proposed method by four indices: the number of the correct matches (NCMs), the correct matching ratio (CMR), RMSE and the running time (Fan et al., 2018). CMR denotes the matching precision, which can be written as $CMR = NCM/NM$. NCM and NM denote the number of correct matches points and the number of matches points, respectively. RMSE is calculated from the manually selected check points between the images. In the template matching, the sensed image was first divided into 25×20 non-overlapping blocks. Then one Harris feature point was extracted in each block, for a total of 500 feature points. The template window sizes and search region sizes are set to 121×121 and 200×200 pixels respectively. The experiment is performed on a computer with an Intel(R) Xeon(R) E5-1603 2.8-GHz processor and 8.0 GB of physical memory, using Microsoft Visual C++ with Geospatial Data Abstraction Library (GDAL).

Figure 7. Experiment Data. (a) test1 data. (b) test2 data. (c) test3 data

Test data	Data description		
	SAR image	Optical image	Study area description
Dataset 1	Sensor: GF-3 Resolution: 5m Date: Mar., 2018 Size: 7751*7393	Sensor: GF-2 Resolution: 4m Date: Apr., 2018 Size: 7300*6908	The study area is with Chengdu City, Sichuan Province, China. some farmland exists in the images.
Dataset 2	Sensor: GF-3 Resolution: 5m Date: Mar., 2018 Size: 12169*12518	Sensor: ZY-3 Resolution: 2.1m Date: Jan., 2017 Size: 24516*24576	The study area is with Wuhan City, Hubei Province, China. The scale difference between two images are approximately 2.5 times and the time interval is 14 months.
Dataset 3	Sensor: GF-3 Resolution: 5m Date: Jul., 2018 Size: 7738*6384	Sensor: GF-2 Resolution: 4m Date: Aug., 2018 Size: 7300*6908	The study area is with Meishan City, Sichuan Province, China. The city surrounded by mountains. The large terrain relief results in the significant local distortions.

Table 1. Specifications of the test data

Method	Dataset 1	Dataset 2	Dataset 3
Proposed Method	NCM	454	429
	CMR (%)	94.98	91.86
	RMSE	0.979	0.985
	Running time (s)	76.465	73.793
MI + local geometric correction	NCM	213	192
	CMR (%)	44.56	41.11
	RMSE	1.423	1.493
	Running time (s)	4295.936	4099.389
NCC + local geometric correction	NCM	111	89
	CMR (%)	23.22	19.06
	RMSE	1.858	1.982
	Running time (s)	992.971	946.614

Table 2. Comparison of the proposed method with the MI and NCC methods

Figure 8. registration checkboard result of dataset1

Figure 9. Registration checkboard result of dataset2

Figure 10. Registration checkboard result of dataset3

3.2 Result and Analysis

The registration results of the three datasets by the proposed method are shown in Figure 8, 9 and 10, and Table 2 lists the detail performance of different methods for the three datasets. It can be clearly observed that NCC fails in the matching for the three image datasets, and the most higher CMR is only 23.22% among all three datasets. The main reason is that NCC is vulnerable to nonlinear radiometric differences between SAR and optical images (Ye and Shan, 2014). Compared with NCC, the performance of MI improves to some extent. Although its RMSEs and CMRs are acceptable, its running time are more than 4000 seconds that is quite time consuming. From Table 2, it is obvious that the proposed method outperforms the two method on the three datasets in terms of NCM, CMR, RMSE and the running time. This is because that the proposed method builds the descriptor by using the structural features of SAR and optical images, which makes it more robust against nonlinear radiometric difference. Moreover, the proposed method speeds up image matching by using the 3-D phase correlation in the frequency domain. On the other hand, due to the different characteristic of different datasets, the CMRs of dataset 2 and dataset 3 decline compared with dataset 1. The reason may be that the temporal difference of dataset 2 is fourteen months, which has a significant impact on the NM and RMSE due to the change of ground objects. Since the images of dataset3 have much mountain areas, significant local distortions caused by large terrain relief not only make it difficult to image matching, and the RFM is also difficult to accurately fit the local distortions, resulting in the registration performance degrades. In summary, the proposed method performs better than MI and NCC for the registration of SAR and optical images.

4. CONCLUSIONS

In order to address the difficulties caused by significant nonlinear radiometric differences and geometric deformation between optical and SAR images, this paper proposes an automatic registration method based on spatial geometry constraint and structure features. In the proposed method, a local geometric correction is carried out by using the RFM, and a geometric structural feature representation is constructed by using the CFOG descriptor. Then, the feature descriptor is transformed into the frequency domain, and 3-D phase correlation is used as the similarity metric for image matching. We have evaluated the proposed method by using multiple high resolution optical and SAR images in various scenarios. Compared with the state-of-the-art intensity-based method(i.e., MI and NCC), the proposed method presents more satisfactory performance in the NCM, the CMR and the computational efficiency. It indicates that the proposed method can effectively handle the radiometric differences and geometric deformation between optical and SAR images.

Since the proposed method is based on geometric spatial constraint relationship and structural features of images, there are some limitations that should be considered. The precondition of the proposed method is that images requires a local geometric correction based on RFM, which can eliminate rotation and scale difference between optical and SAR images. In addition, the registration performance of the proposed method may degrade when geometric features of images are lacking.

ACKNOWLEDGEMENTS

This paper is supported by the National Natural Science Foundation of China (No.41971281).

REFERENCE

- Bruzzone, L., Bovolo, F., 2013. A Novel Framework for the Design of Change-Detection Systems for Very-High-Resolution Remote Sensing Images. *Proceedings of the IEEE*, 101(3), 609-630.
- Schmitt, M., Tupin, F., Zhu, X., 2017. Fusion of SAR and optical remote sensing data – Challenges and recent trends. *IEEE International Geoscience and Remote Sensing Symposium*, 5458-5461.
- Rhee, J., Im, J., Carbone, G., 2010. Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. *Remote Sensing of Environment*, 114(12), 2875-2887.
- Tuia, D., Marcos, D., Camps, G., 2016. Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization. *ISPRS Journal of Photogrammetry and Remote Sensing*, 120, 1-12.
- Sui, H., Xu, C., Liu, J., Hua, F., 2015. Automatic Optical-to-SAR Image Registration by Iterative Line Extraction and Voronoi Integrated Spectral Point Matching. *IEEE Transactions on Geoscience and Remote Sensing*, 53(11), 6058-6072.
- Zitova, B., Flusser, J., 2003. Image registration method: a survey. *Image and Vision Computing*, 21, 977-1000.
- Suri, S., Reinartz, P., 2010. Mutual-Information-Based Registration of TerraSAR-X and Ikonos Imagery in Urban Areas. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2), 939-949.
- Ma, W., Wen, Z., Wu, Yue., Jiao, L., Gong, M., Zheng, Y., Liu, L., 2017. Remote Sensing Image Registration with Modified SIFT and Enhanced Feature Matching. *IEEE Geoscience and Remote Sensing Letters*, 14(1), 3-7.
- Li, H., Manjunath, B., Mitra, S., 1995. A Contour-Based Approach to Multisensor Image Registration. *IEEE Transactions on Image Processing*, 4(3), 320-334.
- Sedaghat, A., Mokhtarzade, M., Ebadi, H., 2011. Uniform Robust Scale-Invariant Feature Matching for Optical Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11), 4516-4527.
- Patel, M., Thakar, V., Shah, S., 2016. Image Registration of Satellite Images with varying illumination level using HOG descriptor based on SURF. *Procedia Computer Science*, 93, 382-388.
- Dare, P., Dowman, I., 2001. An improved model for automatic feature-based registration of SAR and SPOT images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56, 13-28.
- Ye, Y., Shan, J., Bruzzone, L., Shen, L., 2017. Robust Registration of Multimodal Remote Sensing Images Based on Structural Similarity. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5), 2941-2958.
- Ye, Y., Bruzzone, L., Shan, J., Bovolo, F., Zhu, Q., 2019. Fast and Robust Matching for Multimodal Remote Sensing Image

Registration. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11), 9059-9070.

Merkle, N., Luo, W., Auer, S., 2017. Exploiting Deep Matching and SAR Data for the Geo-Localization Accuracy Improvement of Optical Satellite Images. *Remote Sensing*, 9(6), 586.

Girard, N., Charpiat, G., Tarabalka, Y., 2018. Aligning and Updating Cadaster Maps with Aerial Images by Multi-Task, Multi-Resolution Deep Learning. *ACCV 2018: Computer Vision*. 675-690.

Bürgmann, T., Koppe, W., Schmitt, M., 2019. Matching of TerraSAR-X derived ground control points to optical image patches using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. 158, 241-248.

Ye, Y., Shan, J., 2014. A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences. *ISPRS Journal of Photogrammetry and Remote Sensing*, 90, 83-95.

Grodecki, J., Dial, G., 2003. Block Adjustment of High-Resolution Satellite Images Described by Rational Polynomials. *Photogrammetric Engineering and Remote Sensing*, 69(1), 59-68.

Wang, T., Zhang, G., Yu, L., Zhao, R., Deng, M., Xu, K., 2017. Multi-Mode GF-3 Satellite Image Geometric Accuracy Verification Using the RPC Model. *Sensors*, 17(9).

Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 886-893.

Yi, Z., Cao, Z., Yang, X., 2008. Multi-spectral remote image registration based on SIFT. *Electronics Letters*, 44(2), 107-108.

Fan, J., Wu, Y., Li, M., Liang, W., Cao, Y., 2018. SAR and Optical Image Registration Using Nonlinear Diffusion and Phase Congruency Structural Descriptor. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9), 5368-5379.