Progress In Electromagnetics Research, Vol. 163, 89–106, 2018

Coarse-to-Fine Accurate Registration for Airborne SAR Images Using SAR-FAST and DSP-LATCH

Huai Yu[1, 2], Wen Yang[1, 2, *], and Yan Liu[1]

Abstract—Synthetic Aperture Radar (SAR) image registration is to establish reliable correspondences among the images of the same scene. It is a challenging problem to register the airborne SAR images for the instability of airborne SAR systems and the lack of appropriate geo-reference data. Besides, techniques for registering satellite-based SAR images relying on rigorous SAR geocoding cannot be directly applied to airborne SAR images. To address this problem, we present a coarse-to-fine registration method for airborne SAR images by combining SAR-FAST (Features from Accelerated Segment Test) feature detector and DSP-LATCH (Domain-Size Pooling of Learned Arrangements of Three patCH) feature descriptor, which only relies on the gray level intensity of SAR data. More precisely, we first apply SAR-FAST, which is an adapted version of FAST for analyzing SAR images, to detect corners with high accuracy and low computational complexity. To reduce the disturbance of speckle noise as well as to achieve efficient and discriminative feature description, we further propose an improved descriptor named DSP-LATCH to describe the features, which combines the Domain-size Pooling scheme of DSP-SIFT (Scale-Invariant Feature Transform) and the idea of comparing triplets of patches rather than individual pixel values of LATCH. Finally, we conduct a coarse-to-fine strategy for SAR image registration by employing binary feature matching and the Powell algorithm. Compared with the existing feature based SAR image registration methods, e.g., SIFT and its variants, our method yields more reliable matched feature points and achieves higher registration accuracy. The experimental results on different scenes of airborne SAR images demonstrate the superiority of the proposed method in terms of robustness and accuracy.

1. INTRODUCTION

Synthetic Aperture Radar (SAR) image registration is usually a prerequisite for image fusion [1] and change detection [2], which means a process of aligning two or more SAR images in geometry. The reference and object images can be acquired by the same or different sensors at different times, even from different viewpoints, which makes the registration task a big challenge. Notably, airborne SAR images are in lack of accurate geo-reference data, like appropriate Digital Terrain Model (DTM) and orbiting data. Besides, techniques for registering satellite-based SAR images cannot be directly applied to airborne SAR images [3]. The resolution of airborne SAR images is often in decimeter, which is at least one order of magnitude higher than that of satellite-based systems [4]. The higher resolution leads to more speckles as well as a need for more accurate pixel-level matching.

Generally, the registration methods of airborne SAR images can be grouped into two categories, i.e., intensity based methods and feature based methods. The intensity based methods often need a lot of processing time, and the registration results may be trapped into a local optimum. While the feature based methods do not require any prior information, and the globally optimal registration results can

Received 8 July 2018, Accepted 3 September 2018, Scheduled 16 September 2018

* Corresponding author: Wen Yang (yangwen@whu.edu.cn).

[1] School of Electronic Information, Wuhan University, Wuhan 430072, China. [2] State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China.

be easily found [5]. Furthermore, for the airborne SAR images of different bands and different imaging prospectives, the derived orbiting data and DTM data are often unreliable. To handle this problem, in this paper, we mainly analyze the feature based methods, which usually consist of three steps, i.e., feature detection, feature matching, and the precise registration according to the geometrical relations of matched points.

The feature detection methods for airborne SAR images are usually derived from those for optical images [1], among which the most widely used method relies on the SIFT feature [6]. It detects feature points among the image Gaussian pyramid [7], which results in the output feature being scale invariant. However, unlike optical images characterized by neat and uniform features, SAR images are formed by coherent interaction of the transmitted microwave with the targets. Hence, it suffers from the effects of speckle noise which arises from the coherent summation of the signals scattered from scatterers distributed randomly within each pixel. Speckle confers to SAR images a granular aspect with random spatial variations. These multiple interactions produce echoes which interfere with each other in either a constructive or destructive manner. Constructive interference results in a strong return signal and a bright pixel. Destructive interference results in a weak return signal and dark pixel. Thus the existence of speckle noise in SAR images [8, 9] makes it a tough work to extract stable feature points. Schwind et al. [10] proposed an algorithm by skipping the first scale-space OCTave (SIFT-OCT). Nevertheless, the SIFT-OCT method leads to a loss of the information details. Wang et al. [11] proposed an optimized method for SIFT feature based on Voronoi diagram which obtains a better feature distribution and higher registration accuracies. The Voronoi diagram is a partitioning of a plane into regions based on distance to points in a specific subset of the plane[†]. In [12], Dellinger et al. presented an improvement of Harris corner detection method [13] for SAR image, which replaces the differential gradient with the ratio gradient. However, none of these methods works well in heterogeneous areas with strong speckle noise. To overcome this problem, we use an adapted version of the Features from Accelerated Segment Test (SAR-FAST) method to detect corner points. Similar to the original FAST method, SAR-FAST considers the center point of the region which is distinguished from the neighborhood areas as a corner point [14]. To handle the disturbance of speckle noise, SAR-FAST first reduces the tiny textures caused by speckle noise using the rolling guidance filter (RGF) [15]. Consequently, the significant structures and edge information are preserved. Furthermore, SAR-FAST expands the detection windows and extract the gradient changes to increase the robustness .

Feature point matching can be achieved by comparing the similarity of feature descriptors. Descriptors are unique strings to describe the information of feature point. The simplest descriptor is the intensity information of images. However, this descriptor is not unique enough to distinguish different feature points. Two classical methods using the intensity gradient are the SIFT descriptor and the Speed Up Robust Features (SURF) [16] descriptor, the latter is an improvement of the SIFT descriptor with a higher computational efficiency. These two descriptors work very well in optical images, but they suffer from great performance degradations in SAR images due to the speckle noise. Considering the inherent property of SAR images, SAR-SIFT was proposed dedicated to SAR image [12, 17], which calculates the gradient by ratio instead of differential. Liu et al. [18] proposed the kernel affine invariant SIFT (KA-SIFT) which combined the kernel space and SIFT. Thus the descriptor of KA-SIFT is both affine invariant and linearly independent. In [19], a variant of SIFT is proposed for SAR images, which utilized the bilateral filter instead of the Gaussian filter to construct the spatial scale of anisotropy, resulting in better corner matching results of SAR images. In [20], the DSP-SIFT was presented to describe every feature by mean pooling of SIFT descriptors in multiple scales. Except for the SIFT based methods, other feature-based registration methods are also well exploited. Ye et al. [21] proposed a feature descriptor named the histogram of orientated phase congruency (HOPC) based on the image structural properties. Rui et al. [22] proposed a shape context descriptor based on shape information. The shape context descriptor is used to compare the contours of objects and detect invariant control points. Xiang et al. [23] presented a rotation invariant descriptor without assigning a dominant orientation, which showed robust rotation invariance and a superior matching performance.

In addition, binary descriptors, which can be matched by Hamming distance with high efficiency by bit operations, are designed to minimize the computation complexity and the storage cost. In the literature, many binary descriptors have been proposed, such as BRIEF [24], BRISK [25], FREAK [26]

[†] https://en.wikipedia.org/wiki/Voronoi_diagram.

and LATCH [27]. Although the binary descriptors require lower computational costs, the performance is usually inferior to the classical float descriptors (e.g., SIFT, SURF). To preserve the computation efficiency of binary descriptors as well as to obtain the high description precision of SIFT descriptor and its variants, we propose an improved descriptor, i.e., DSP-LATCH, which requires less calculation costs and works better for SAR image registration. In our method, we first build the scale space of SAR images and extract SAR-FAST in each scale. The direction of each corner is obtained according to the gradient statistics. Thus the scale and rotation invariant SAR-FAST feature points can be obtained. Then, the multi-scales LATCH descriptors are computed for each corner point. Finally, we apply the mean pooling to the multi-scale LATCH descriptors to generate the final DSP-LATCH descriptor.

The third step of feature based registration method is to get the fine transform parameters according to the geometrical relationship of matched points. For multi-temporal images, the various changes between images often result in poor registration accuracy in the rough registration stage. Therefore, fine registration is an important step for multi-temporal SAR image registration. Based on the matched points, Chui et al. proposed an automatic outlier suppression mechanism in the coherent point drift (CPD) algorithm [30] to solve the problem that outliers are always involved in the operation until the CPD converges [28, 29]. In [31], a novel dual-graph-based point matching method was proposed for multi-spectral image registration, in which Delaunay triangulation matching and Voronoi derived from Delaunay graph are employed to remove the outliers and recover the inliers. In [32], an overview of mutual information (MI) based fine registration methods was presented for medical images. In our method, we use the mutual information based method to find the fine registration results. By initializing the transform parameters using the feature matching results, the fine registration results are obtained by employing Powell algorithm [33].

To summarize, inspired by the existing image registration methods, we propose a coarse-to-fine accurate registration method for airborne SAR images. In the coarse step, we present an efficient feature based rough registration method by combining SAR-FAST feature detector and DSP-LATCH feature descriptor. The SAR-FAST feature point enjoys very good robustness to speckle noise in SAR images, and the DSP scheme of LATCH description further promotes its robustness and efficiency. In the fine step, we use the geometrical relations of matched points as the initialization of Powell algorithm to maximize the mutual information, which avoids the registration result being trapped into a local optimum. Although there are many feature based registration methods for natural image, very few of them consider the characteristic of SAR images, e.g., the abundant speckle noise. Based on our previous work of SAR-FAST feature points extraction [14], we further proposed the DSP-LATCH descriptor, which is specially designed for SAR images and shows good discriminability. Unlike the common satellite-based SAR image registration framework, the proposed method does not rely on any orbiting data and DTM data, but directly uses the gray level characteristics of airborne SAR images. The main contributions of this paper are two folds:

- The proposed DSP-LATCH descriptor enjoys both computational efficiency of binary descriptors and high description accuracy of SIFT based descriptors, which is more suitable for the registration of airborne SAR images.
- A coarse-to-fine feature based framework is presented for the registration of airborne SAR images. Compared with SIFT-like feature based methods, the proposed approach yields more reliable matched feature points and achieves higher registration accuracy than the state-of-the-arts.

The rest of this paper is organized as follows: The proposed method is presented in Section 2, and the experimental results and comparisons are illustrated in Section 3. Finally, concluding remarks are given in Section 4.

2. AIRBORNE SAR IMAGE REGISTRATION BASED ON SAR-FAST AND DSP-LATCH

In this section, we present our feature based registration method of airborne SAR image in detail. The proposed registration framework is shown in Fig. 1. The RGF method [15] is first used to reduce the speckle noise. Then, SAR-FAST is utilized to detect corners. Next, DSP-LATCH is applied to describe corner points. By using the RANdom SAmple Consensus (RANSAC) algorithm, we can get the initial

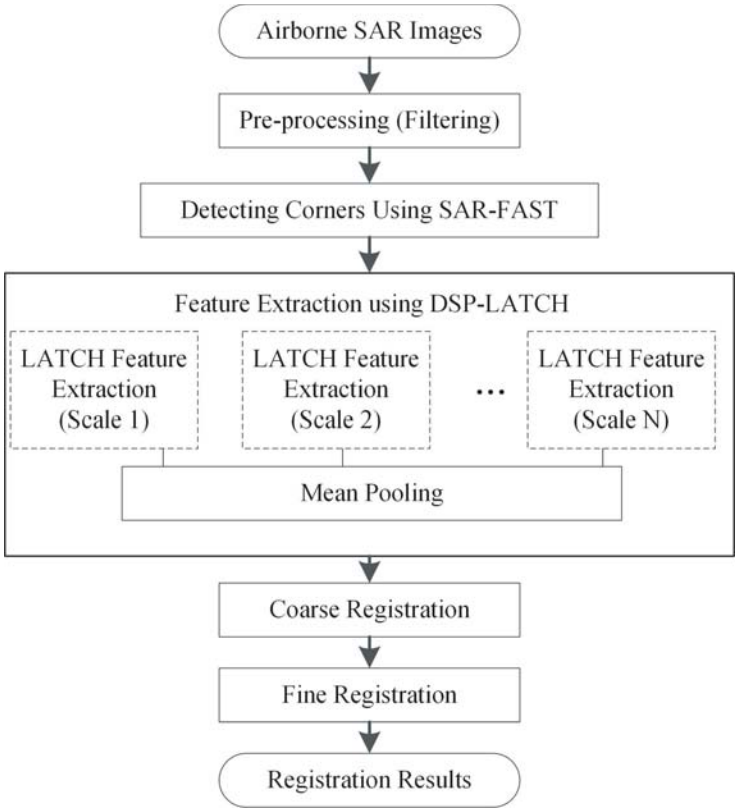


Figure 1. SAR image registration framework of the proposed method.

affine transformation parameters between two SAR images. Finally, we apply the fine registration method based on mutual information to obtain final precise registration results.

2.1. Detecting Corners Using SAR-FAST

Based on our previous work [14], we use SAR-FAST to detect corner points. SAR-FAST is inspired by the FAST corner detection method in optical image processing. To better capture the gradient change information, SAR-FAST expands the detection windows and radius. Fig. 2 shows the detection window of SAR-FAST. SAR-FAST consists of four steps, i.e., image preprocessing, corner detection, false detection point elimination, and non-maximum suppression.

2.1.1. Image Preprocessing

In airborne SAR images, the presence of speckle noise is inevitable. Basic textures as they appear, are affected by multiplicative speckle noise and should be preserved by despeckling algorithms [34, 35]. In intensity image, the speckle noise presents as adjacent intensity saltation. When we use the intensity image to detect corner points, the tiny texture caused by speckle noises would lead to false detections. Thus, we use the RGF method to remove the tiny texture while preserving the significant structures and edge information [15]. RGF method includes two steps, i.e., elimination of tiny textures and enhancement of significant textures. First, the Gaussian blurring is applied to remove the tiny textures. Since the original image contains significant structures and edge information, we use the structure of the original image as the guidance to enhance the significant textures of the blurred image. The process is repeated until tiny textures are removed while the significant structures are preserved.

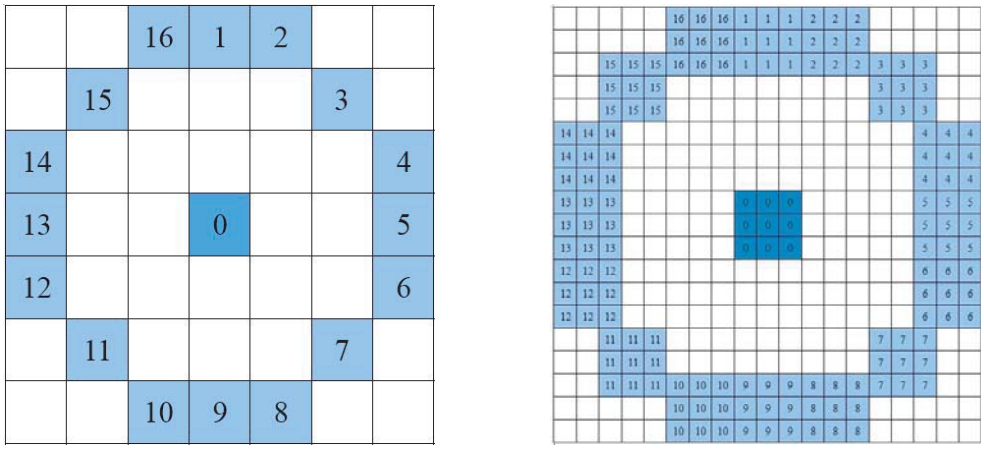


Figure 2. The feature detection windows of FAST and SAR-FAST. (a) FAST detection window. (b) SAR-FAST detection window.

2.1.2. Corner Detection

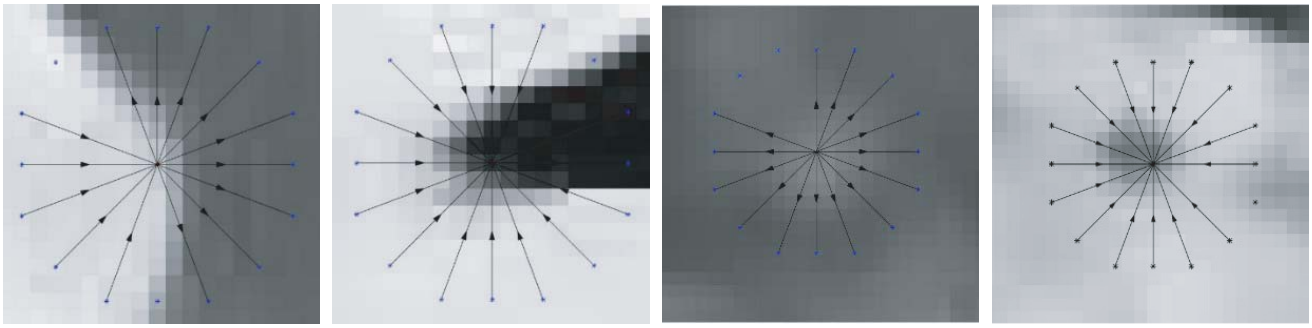
Being different from our previous work [14], the SAR-FAST corners are extracted on the Gaussian pyramid to make the extended SAR-FAST robust to scale changes [36]. To enhance the robustness of the detector, SAR-FAST changes the detection radius and window size of the original design of FAST [37]. The original detection window schematic diagram of FAST is shown in Fig. 2(a) with one pixel's center and 3 pixels' radius. As to SAR-FAST, when calculating the similarity between pixels, a small detection window with a size of 3×3 pixels is applied. With radius of 9 pixels, there exist 16 small windows to be evaluated, which can be seen in Fig. 2(b). We assume the figure is a patch of intensity map of SAR image. For every single 3×3 patch, we formulate the 19×19 patch kernel and detect if the center patch is different from the 16 surrounding patches.

When calculating the similarity of pixels, the smallest intensity value is used as the reference value. Suppose that the mean value of the center small window is P, and the threshold is $Th (Th > 0)$. We use WP_i^j to denote the jth ($j \in [1, 9]$, $j \in N$) pixel value of the ith window W_i. For the ith window $W_i\,(1 \leq i \leq 16)$, the similarity flag K_i is calculated by

$$K_i = \begin{cases} 1, & \forall j \in [1,9], \quad WP_i^j - P \geq Th; \\ 0, & \exists j \in [1,9], \quad |WP_i^j - P| < Th; \\ -1, & \forall j \in [1,9], \quad WP_i^j - P \leq -Th. \end{cases} \tag{1}$$

When the differences of all the 9 pixel values WP_i^j to the mean center value P are larger than the threshold Th, we consider the i-th window as not similar to the center window, i.e., $K_i = 1$. Then if all the 9 differences are smaller than the threshold $-Th$, we also consider the i-th window as not similar to the center window, i.e., $K_i = -1$. Otherwise K_i is equal to 0. We use AI (arrow in) to present the gradient towards the center point ($K_i = 1$), namely the arrow points inside to the center point. Similarly, we use AO (arrow out) to present the gradient from the center point ($K_i = -1$), namely the arrow points outside to the center point. If more than M (empirical $M = 8$) consecutive K_i are all 1 or all -1, then we determine the center point as corner point. Considering the fact that the corner point is of a certain angle, and the total gradient number is 16, if the center point belongs to an edge rather than a corner, there will be equal numbers of AO and AI, i.e., 8. Thus if the center point is a corner, one of number of AO and AI must be larger than 8. This is the reason why the empirical threshold M is set as 8. By introducing the above standards, we can preliminarily determine the corner points.

The possible situation of feature detection in real SAR image is shown in Fig. 3. From which we can observe that there are at least 9 consecutive arrows point to same direction, e.g., 9 consecutive arrows point outside in Fig. 3(a) and 13 consecutive arrows point inside in Fig. 3(b). According to the feature detection criterion, these two detection outputs are regard as good features. While in other



(a) (b) (c) (d)

Figure 3. The gradient distribution of corners. (a) Good detection. (b) Good detection. (c) Error detection. (d) Error detection.

cases, homogeneous areas with strong speckle noise would be detected as corner points as well. Two typical gradient distributions, i.e., all the gradients toward the neighborhood (Fig. 3(c)) and all the gradients toward the center points (Fig. 3(d)), would cause error detection. Namely, error detection points only exist in the areas affected by speckle disturbance with gradient variation only in one direction. Considering this, the erroneous corner detection can be eliminated.

When the detected corner points satisfying the aforementioned conditions are obtained, non-maximum suppression (NMS) is applied to avoid repeated detection in a local area, which follows the NMS scheme of original FAST [37]. Finally the detected corner points in each scale of the pyramid are reverted to the original image to generate the scale transformation invariant feature points.

2.2. Generating DSP-LATCH Descriptor

After the feature points are detected by SAR-FAST, we describe the feature points in a compact way. Traditional methods directly use the SIFT and its variant SAR-SIFT as descriptors. However, both the aforementioned descriptors are of low efficiency. Here we resort to the efficient binary descriptor. Since we could not thoroughly remove speckle noise, the descriptor must be robust to speckle noise.

2.2.1. LATCH Descriptor

Usually, binary descriptors are generated by intensity values. The intensity-value based method is simple and efficient, but sensitive to noises. To overcome the common drawbacks of binary descriptors, Levi and Hassner [27] proposed the LATCH descriptors which is a fast and compact binary descriptor. The core idea of LATCH is matching patches with patches instead of comparing the values of pixel pairs, which enhance the robustness of the binary descriptors while preserving the efficiency.

We set the feature point as center of the neighborhood area, which is remarked as the yellow box in Fig. 4. Instead of two pixels, LATCH selects three patch triplets as a group to generate descriptors to describe the neighborhood areas of a feature point. We denote the patch set as $\hat{S} = \{\hat{s}_t\}_{t=1,\ldots,T} = \{P_{t,a}, P_{t,1}, P_{t,2}\}_{t=1,\ldots,T}$, where T is the number of sets, $P_{t,a}, P_{t,1}, P_{t,2}$ are the $k \times k$ patches selected. We set $P_{t,a}$ as the anchor patch, $P_{t,1}$ and $P_{t,2}$ as the auxiliary patches. Considering that only a small number of bits are typically required, we must consider which of the many possible triplet arrangements should be employed. The selection of values for S is performed beforehand, either randomly (e.g., BRIEF [24]), manually (BRISK [25]), or is automatically learned from training data (ORB [38] and FREAK [26]). LATCH selects patches by learning from the dataset used in [39]. The local image windows in this dataset are partitioned into 500k pairs of patches labeled as being same (the two windows present the same physical scene point, viewed from different viewpoints or viewing conditions) and "not same". Then 56k patch triplets are randomly selected. The pixel coordinates $P_{t,a}$, $P_{t,1}$ and $P_{t,2}$ are selected randomly from all the triplets' combinations within the neighborhood windows S of the feature point. Then we define the quality of an arrangement by summing the number of times

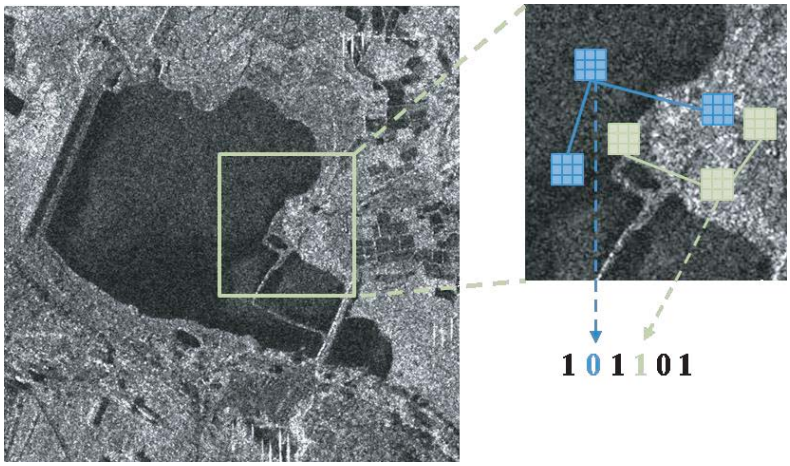


Figure 4. The LATCH descriptor generation.

that it correctly yields the same binary value for same labeled pairs and different values for not same labeled pairs.

To prevent highly correlated arrangements being selected, only the triple combination whose absolute correlation with all previously selected combinations is smaller than the threshold τ is selected. In our experiments, this value is set to $\tau = 0.2$ and left unchanged just as the original LATCH descriptor set. The binary descriptors are generated by calculating the Frobenius norm between the anchor patch and its auxiliary patches [27].

$$g(S, \hat{s}_t) = \begin{cases} 1, & \text{if } \|P_{t,a} - P_{t,1}\|_F^2 > \|P_{t,a} - P_{t,2}\|_F^2; \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and S denotes the neighborhood window of feature point. The Frobenius norm of matrix A is $\|A\|_F = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}|a_{i,j}|^2}$, with $a_{i,j}$ denoting the (i, j) element of A. The similarity between reference patch and auxiliary patch is calculated by the Frobenius norm of their intensity differences. By exploiting all the similarities of neighborhood patches, we obtain the binary descriptors of feature points, which can be denoted as [27]

$$b_S = \sum_{1 \le t \le T} 2^t g(S, \hat{s}_t). \tag{3}$$

In Fig. 4 we show how to generate the LATCH descriptors. We set the feature point as center of the neighborhood area. In the neighborhood area we use two triplet patches (yellow and blue) to generate two bits of the binary descriptor by computing the similarities between them. For example, by calculating the Frobenius norm between the anchor patch and its auxiliary patch of the yellow triplet, it yields one bit $g(S, \hat{s}_t) = 1$. For a binary descriptor of 32 bits, the patch size is 7×7, i.e., $k = 7$, and the neighborhood size is 48×48.

2.2.2. DSP-LATCH Descriptor

LATCH descriptor achieves a good description for feature point in SAR images. However, without considering the information loss after filtering, the description capability of LATCH is limited. Inspired by DSP-SIFT [20], we apply the pooling strategy to the descriptors in different scales to remove the redundant information while preserving significant information. The DSP-LATCH descriptor is more distinguishable than DSP-SIFT since it also retains the advantages of binary descriptor.

The procedure of DSP-LATCH can be described as follows. First we set SAR-FAST corner point as the center of current scale to build the scale pyramid. Then we calculate the gradient directions of each feature point to get the direction information of feature points. Patches are then normalized into the same scale, with LATCH descriptors in different scales extracted. Finally we apply mean pooling to all the descriptors and the DSP-LATCH descriptors are obtained.

2.3. SAR Image Registration

Mutual information [40] is commonly used in registration methods, which considers the intensity values as random variables to measure the correlation between images. In mutual information based image registration methods, two images to be aligned are regarded as two random variables, whose statistical correlation is calculated by entropy of the variables. Suppose the two images to be aligned are I and J. According to the definition of mutual information, the mutual information of I and J can be written by

$$MI(I, J) = H(I) + H(J) - H(I, J). \tag{4}$$

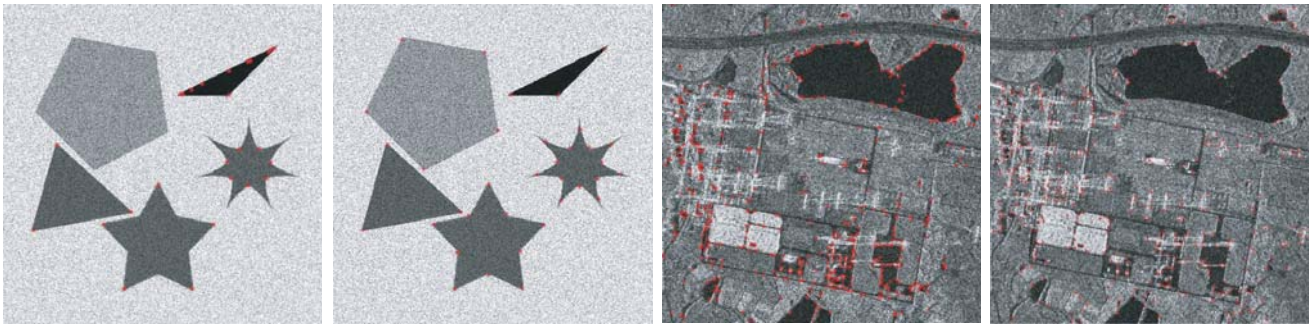
where $H(I)$ and $H(J)$ denote the entropies of the two image respectively, and $H(I, J)$ denotes the joint entropy of the two images.

The Powell algorithm is used to maximize the mutual information and calculate the fine registration results, with the coarse registration results as the initial values [41]. Powell algorithm is a local optimization method with high efficiency. However, the Powell algorithm might be easily trapped into a local optimum and is reliable to initial values. After coarse registration, we initialize the transformation parameters to reduce the search domain, which avoids the solution being trapped into a local optimum. By doing so, the Powell algorithm guarantees the validity and reliability, and it helps to obtain accurate results.

3. EXPERIMENTAL RESULTS

3.1. Corner Detection Results

In this section, two sets of images are used for corner detection. The first dataset is a synthetic image generated by optical image with additive speckle noise of Gamma distribution. The second dataset is a real SAR image. This SAR image is a intensity data with 800×800 pixels and 3 m spatial resolution. The experimental platform consists of an Intel Core i7-4790 CPU, a 32G RAM, a 64 bit Windows 7 operating system with OpenCV 3.3 and Matlab 2017a. In the experiment, the similarity measure threshold Th is set to 20 empirically. SAR-Harris is implemented for performance comparison. The experimental results are shown in Fig. 5.



(a) (b) (c) (d)

Figure 5. The detected corners (in red) of SAR-Harris and SAR-FAST. (a) Synthetic SAR (SAR-Harris). (b) Synthetic SAR (SAR-FAST). (c) Real SAR (SAR-Harris). (d) Real SAR (SAR-FAST).

Comparing Figs. 5(a) and (b), we learn that SAR-FAST accurately detects all corners in the synthetic SAR image, while SAR-Harris produces some false positives in the triangle structure and some missed positives in the pentagon structure. SAR-Harris yields 18 true positives, 5 false positives and 17 missed positives, but SAR-FAST detects 35 true positives, 0 false positive and 0 missed positives. In real SAR image, the distribution of corners is more complex. In Fig. 5(d) we see the SAR-FAST also detects corners correctly in most cases. Overall, compared with SAR-Harris, SAR-FAST leads to less false negatives and false positives, which indicates better detection accuracy.

To further test SAR-FAST, we use the repeatability, which refers to whether the algorithm can detect the corresponding corners in the same location of different images, as the index to evaluate the algorithm. For image I and image J, we denote the detected corner set as P_I and P_J respectively. The repeatability is calculated by

$$k = |P_I \cap P_J|_{L_0}, \tag{5}$$

where k denotes the number of the intersection of two feature point sets. For the i-th feature point in image I and the j-th feature point in image J, we define the spatial intersection as:

$$P_{I_i} \cap P_{J_j} = \begin{cases} 1, & \text{if } D(P_{I_i}, P_{J_j}) \leq d,\ 0 \leq i \leq M_I,\ 0 \leq j \leq M_J; \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

where D is the Euclidean distance between two corner points; d denotes the distance error; M_I, M_J are the number of detected SAR-FAST points respectively. We choose 5 sets of fine aligned Bi-temporal images as the dataset to find out the repeatability of SAR-FAST. The range of d is set to be $[0, 5]$ and the stride is set to be 0.2. To measure the repeatability percentage, we use the number of matched pairs satisfying $d \leq 10$ as the fraction base. The repeatability percentages under different values of d are shown in Fig. 6. From the curve in Fig. 6, we can observe that SAR-FAST achieves much better repeatability than SAR-Harris in the same error range d. Thus our SAR-FAST has more correct and repeatable corner detection results. In our previous work, we demonstrated that SAR-FAST also works better than SAR-Harris in rotation invariance and robustness [14].

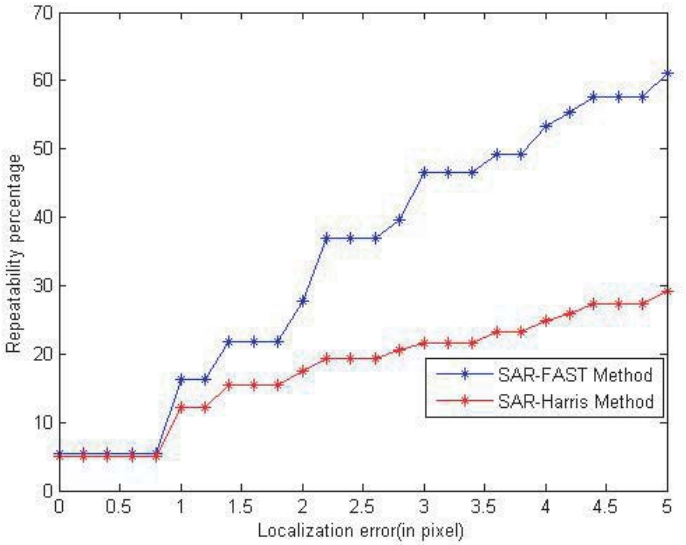


Figure 6. The repeatability comparison of SAR-Harris and SAR-FAST.

3.2. Registration Results

In our work, we choose three sets of Bi-temporal SAR intensity images by the high-resolution airborne SAR system as the experimental datasets for coarse-to-fine registration. Both the images in the first set are acquired in the X-band, HH polarization with 600×600 pixels over Linshui City in Hainan Province of China. Its acquisition date is January 31st, 2010. The resolution of the data is 0.50 m. The distance between platform and the closest scene edge is about 7 km. For the second and third datasets, there are two images with 800×800 pixels captured over Zunhua City, Hebei Province, China, in 0.50 m spatial resolution. The images are in the P-band, VV polarization at time 1 (T1) and in the L-band, VV polarization at time 2 (T2), respectively. The timestamps of the images are September 26th, 2013 (T1) and October 12th, 2008 (T2), respectively. The flight height is about 4 km. The first dataset is of same band and captured in the same time, which is mainly used to validate the effectiveness of the proposed

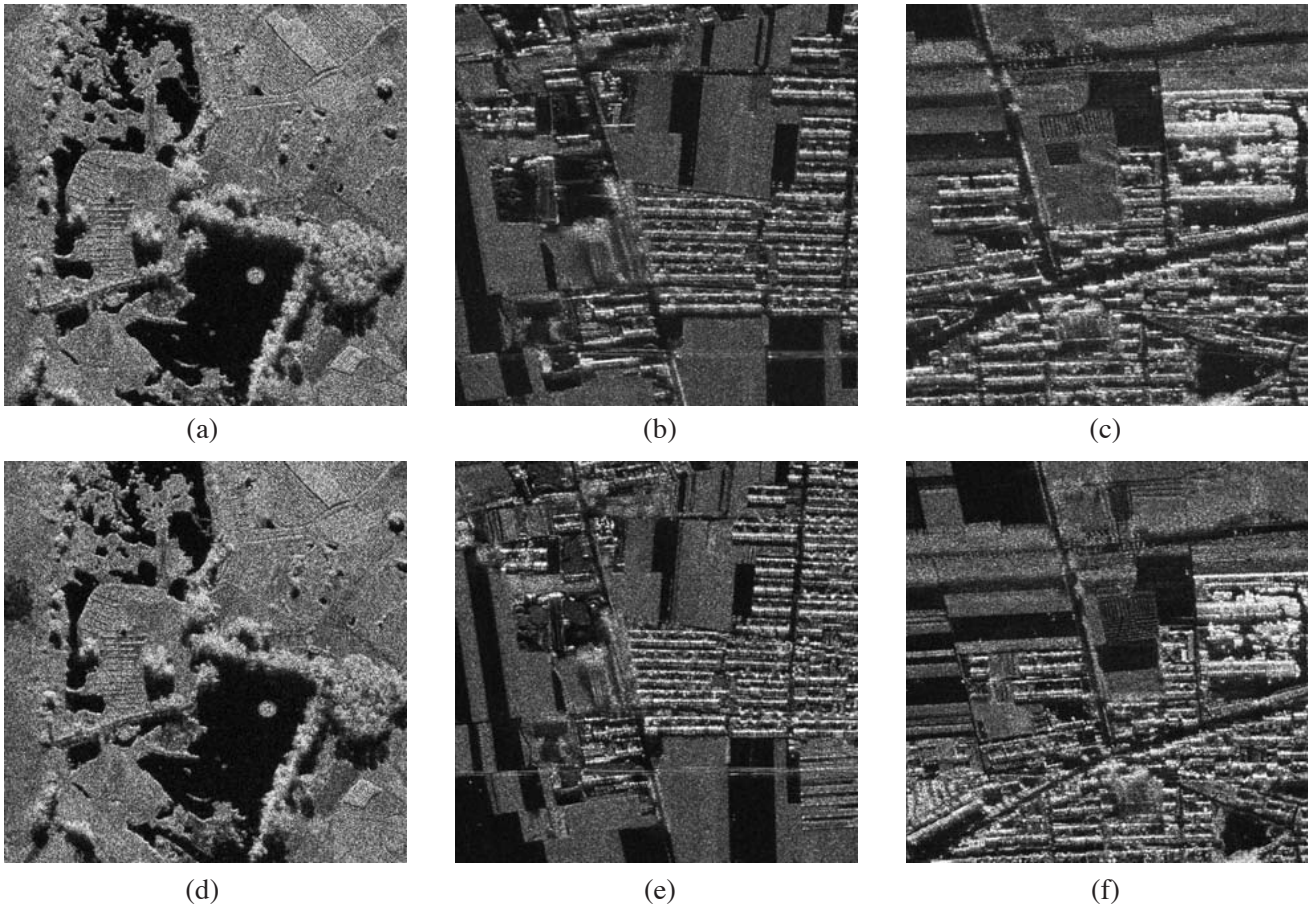


Figure 7. SAR images of dataset 1, dataset 2 and dataset 3. (a) SAR image of dataset 1. (T1) (b) SAR image of dataset 1. (T2) (c) SAR image of dataset 2 (T1). (d) SAR image of dataset 2. (T2) (e) SAR image of dataset 3 (T1). (f) SAR image of dataset 3 (T2).

registration method. The second and third datasets are more challenging to test the superiority of the proposed method than other features, since the resolution is different, and there exist many changes due to the band differences and time differences. The three datasets are shown in Fig. 7.

We use four descriptors, i.e., SIFT, DSP-SIFT, SAR-SIFT, and LATCH, to compare with the proposed DSP-LATCH method. First the nearest neighbor matching algorithm is applied based on distances between descriptors to get the candidate match results. Then, RANSAC is used to refine the matching pairs. Finally, we calculate the transformation parameters by the least square algorithm. In this section, we analyze and evaluate the five descriptors qualitatively and quantitatively. In the section of quantitative analysis, we use the matching quantity and the matching error as the indexes. For the labeled pairwise matching point set in two images $P = (x, y)$, $P' = (x', y')$, the matching error is characterized by Root Mean Squared Error (RMSE), which can be derived by

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left((x_i - T(x_i'))^2 + (y_i - T(y_i'))^2\right)} \tag{7}$$

where N denotes the number of matching pairs, and $T(\cdot)$ denotes matrix transpose. RMSE measures the re-projection distance error after the points are transposed. By using the same kind of labeled pairwise matching point set, the final matched pairs re-projection error directly reflects the registration accuracy. Theoretically, smaller RMSE indicates better matching result. For all the competitors, they use the same set of matched point pairs as the control points to calculate the RMSE. The quantitative comparison is shown in Tables 1, 2, and 3.

Table 1. The matching quantity and error of the 5 descriptors (on dataset 1).

Descriptor	SIFT	SAR-SIFT	DSP-SIFT	LATCH	DSP-LATCH
Quantity	169	422	124	413	**543**
RMSE	2.34	2.16	**1.98**	2.16	2.07
Time (s)	0.75	0.96	4.85	0.69	3.89

Table 2. The matching quantity and error of the 5 descriptors (on dataset 2).

Descriptor	SIFT	SAR-SIFT	DSP-SIFT	LATCH	DSP-LATCH
Quantity	9	31	25	55	**191**
RMSE	545.65	3.10	2.30	19.88	**2.16**
Time (s)	0.76	1.59	9.77	1.08	15.15

Table 3. The matching quantity and error of the 5 descriptors (on dataset 3).

Descriptor	SIFT	SAR-SIFT	DSP-SIFT	LATCH	DSP-LATCH
Quantity	8	39	20	223	**276**
RMSE	307.20	2.70	3.44	4.59	**2.49**
Time (s)	0.74	1.40	8.73	0.99	14.85

From Tables 1, 2, and 3 we learn that DSP-LATCH achieves larger number of matched pairs than other competitors. For dataset 1, DSP-SIFT reaches the smallest matching error, and then DSP-LATCH. For dataset 2 and dataset 3, DSP-LATCH obtains the smallest matching error. Then, about the time costs, by using the Domain-size pooling (DSP) strategy, DSP-SIFT and DSP-LATCH are slower than SIFT and LATCH. However, the registration results are improved substantially for the challenging datasets (i.e., dataset 2 and dataset 3). Since SIFT feature even fails in registering the two data. The additional time costs are negligible when its applications demand very precise registration results.

To further demonstrate the effectiveness of DSP-LATCH compared with DSP-SIFT, we show the matching results of DSP-SIFT and DSP-LATCH for dataset 1 in Figs. 8 and 9, respectively. By comparison, we learn that the matching pairs of DSP-SIFT are of small quantity and the matching distribution is concentrated. Consequently, the probability of matched pairs corresponding to the same transformation model is greater, and the matching error is much smaller. Although the RMSE of DSP-LATCH is a little larger than DSP-SIFT, the number of matched pairs is much larger with a more uniform distribution. Overall, DSP-LATCH works better than the competitors in matching corners. Compared to LATCH, DSP-LATCH preserves more significant information, leading to better matching results.

From the aforementioned analysis, we learn that DSP-LATCH achieves the best matching results in terms of matching quantity, matching distribution and matching error in coarse registration. The matching of feature points is the first step for the registration of airborne SAR images, which is followed by fine registration step. The fine registration results using mutual information of three datasets are shown in Fig. 10. The matching results of SAR-FAST and DSP-LATCH are set as the initial value. Then, three sets of images are aligned using mutual information by Powell algorithm. To show the registration results more clearly, we use checkerboard to show the registration results of the intensity images in mosaic fashion. In the boundary of mosaic blocks we can observe that straight lines and curves link to the adjacent part smoothly, without any evident breaking and bending occurred.

In addition, we compare the fine registration based on mutual information with the thin-plate spline for robust point matching (TPS-RPM) [28]. The fine registration results using TPS-RPM are

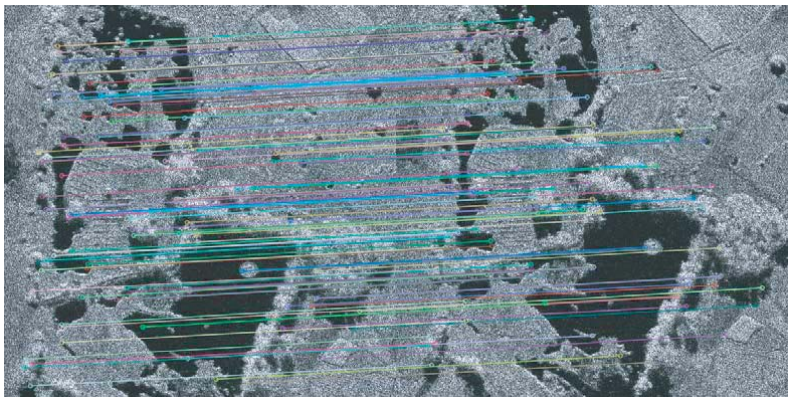


Figure 8. The matching results of DSP-SIFT descriptors (on dataset 1).

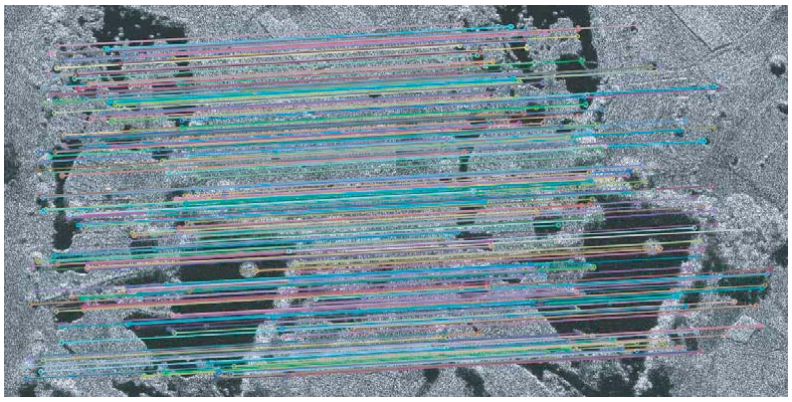
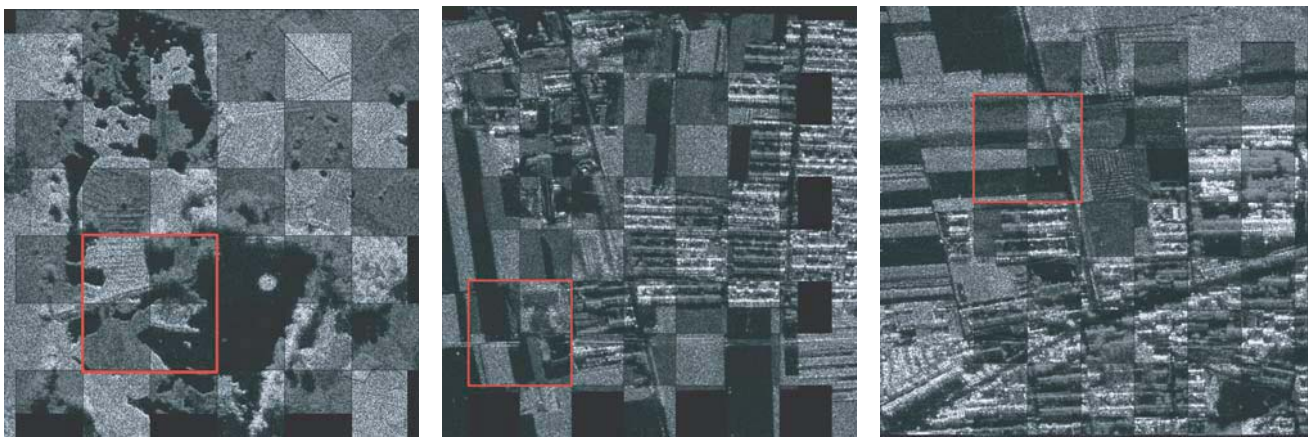


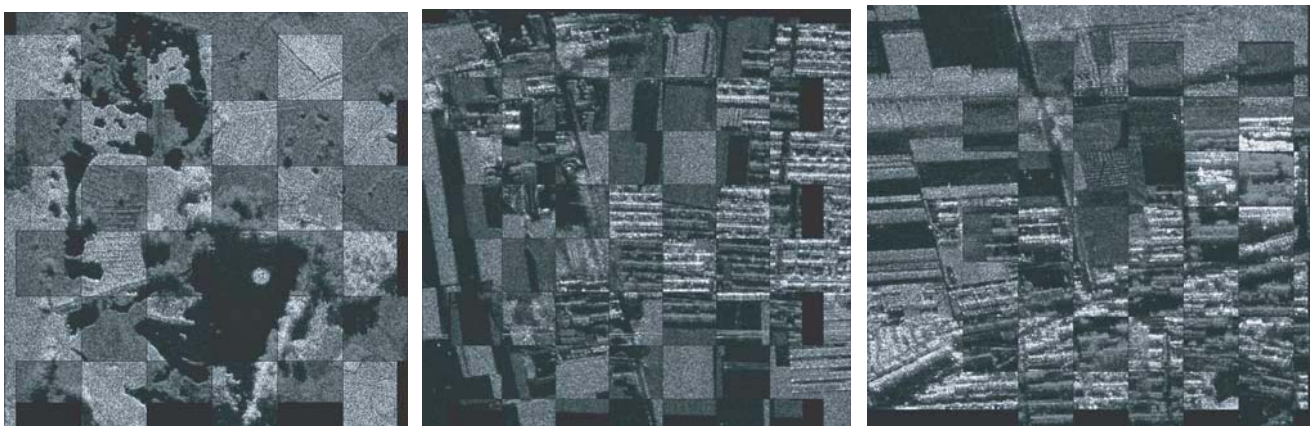
Figure 9. The matching results of DSP-LATCH descriptors (on dataset 1).



(a) (b) (c)

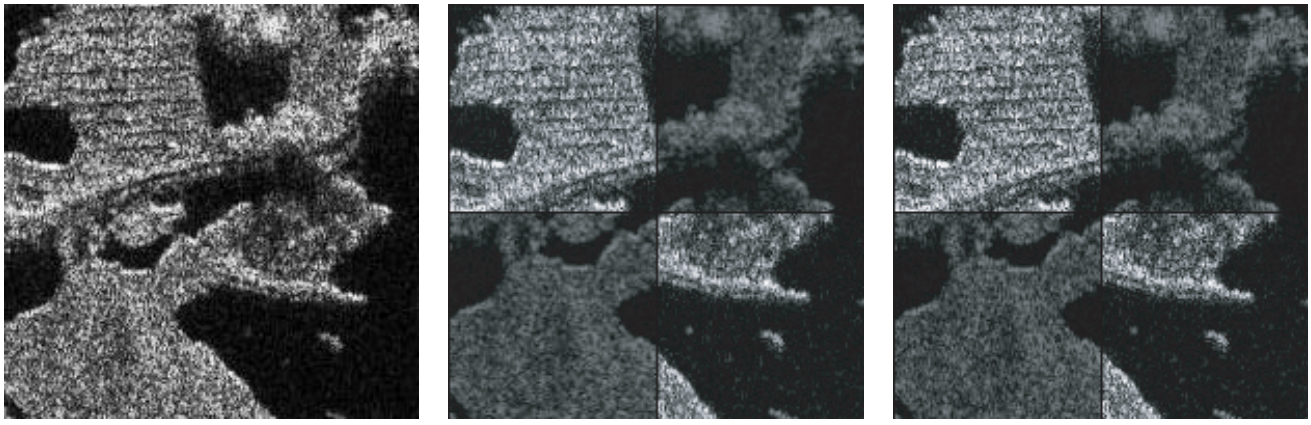
Figure 10. The fine registration results based on mutual information. (a) Fine registration result of dataset 1. (b) Fine registration result of dataset 2. (c) Fine registration result of dataset 3.

shown in Fig. 11. To give a detailed comparison of the two methods, three selected regions (marked with red squares in Fig. 10) of three dataset are shown in Figs. 12, 13 and 14, respectively. From which we can observe that the structural contours of our method are linked more smoothly. So the results using mutual information have relatively smaller registration error. To analyze the registration results



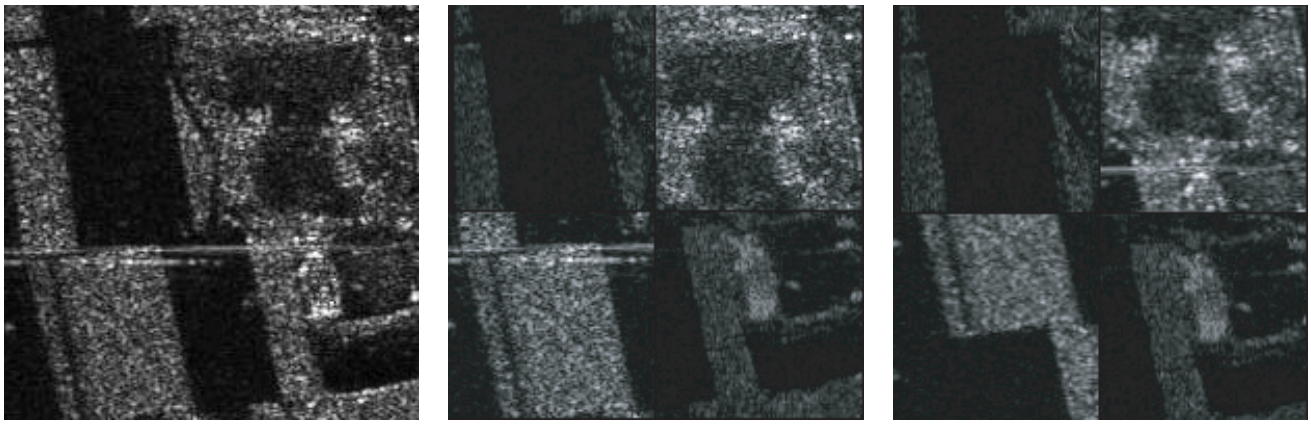
(a) (b) (c)

Figure 11. The fine registration results based on TPS-RPM. (a) Fine registration result of dataset 1. (b) Fine registration result of dataset 2. (c) Fine registration result of dataset 3.



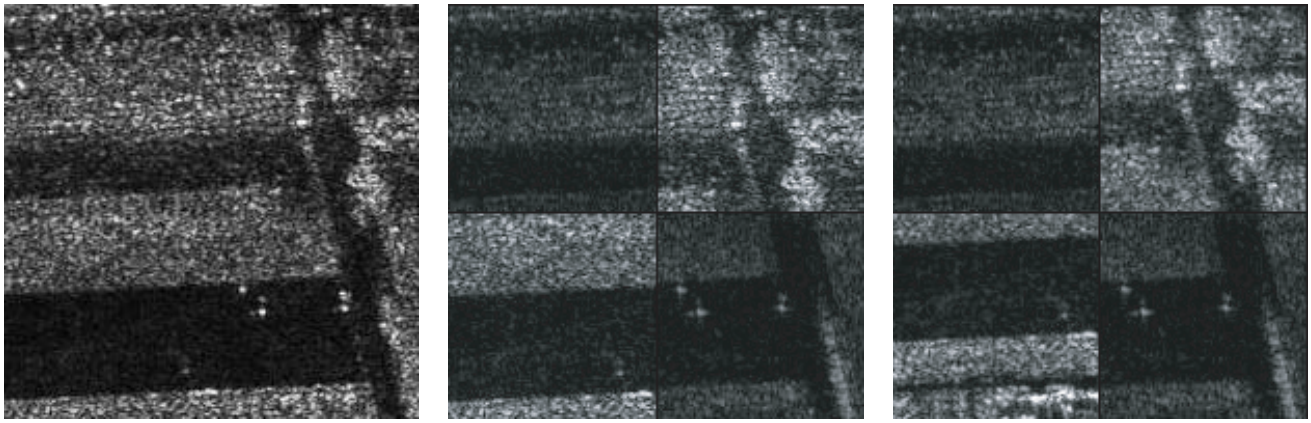
(a) (b) (c)

Figure 12. The fine registration results of selected regions in dataset 1. (a) Ground truth. (b) Mutual information. (c) TPS-RPM.



(a) (b) (c)

Figure 13. The fine registration results of selected regions in dataset 2. (a) Ground truth. (b) Mutual information. (c) TPS-RPM.



(a) (b) (c)

Figure 14. The fine registration results of selected regions in dataset 3. (a) Ground truth. (b) Mutual information. (c) TPS-RPM.

Table 4. The fine registration result comparison on three datasets.

Method	Dataset 1		Dataset 2		Dataset 1	
	CC	SSIM	CC	SSIM	CC	SSIM
Feature + MI	**0.59**	**0.63**	**0.34**	**0.31**	**0.48**	**0.40**
Feature + TPS-RPM	0.52	0.53	0.10	0.09	0.10	0.05

quantitatively, we use two indexes, i.e., correlation coefficient (CC) and structural similarity (SSIM), to measure the registration accuracy. For CC and SSIM, larger values mean better performance. The registration results for three dataset are shown in Table 4, respectively. For the dataset 1, because the two images are from the same band and the same time, both the methods based MI and TPS-RPM obtain good results. However, compared with TPS-RPM based method for the other two datasets, the results using MI are much better in terms of CC and SSIM. The reason is that the MI-based method utilizes all the matched sets in the rough matching step, while the TPS-RPM further removes some "outliers" of the matched sets. The intensity differences and time changes make the removing of "outliers" decreases the registration accuracy. Thus the registration error of the MI-based method is relatively smaller.

4. DISCUSSION

Airborne SAR image registration is a very important task in remote sensing applications. In this paper, we first analyze the every procedure of feature based registration methods and then we introduce a coarse to fine registration framework for SAR images. Our proposed registration method mainly focuses on the number and the precision of extracted feature points and the precision of final registration results. From Section 3.1, we can observe the number and the precision of reliable feature points extracted by SAR-FAST are much higher than SAR-Harris. From Section 3.2, we can find the quantity of the final matched points of our method, i.e., DSP-LATCH, is much higher than that of SIFT and its variants (SAR-SIFT and DSP-SIFT). Without using any geo-reference data, the proposed method produces good registration results on multi-band and multi-temporal airborne SAR images. Thus we conclude that the final registration results of the proposed method are much better.

Analyzing from the feature characteristics, SIFT yields a relatively smaller number of matches while producing relatively higher RMSE, which fully shows its inferiority to DSP-LATCH for SAR image registration. In terms of the features robustness to scale and rotation changes, we used three different pairwise airborne SAR images in our experiments. The first pair comes from the same band of

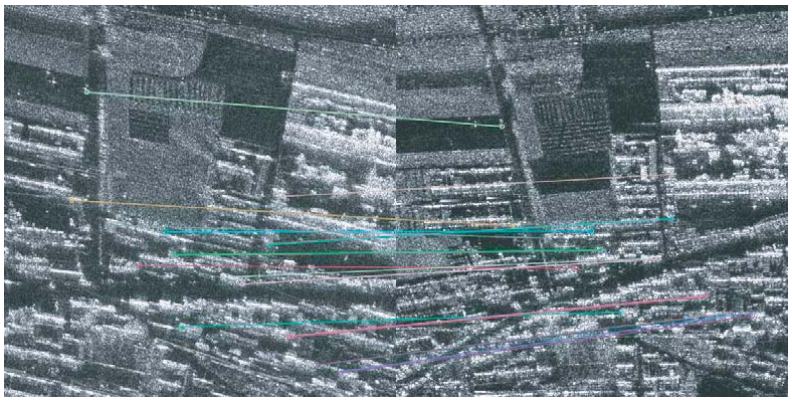


Figure 15. The matching results of SAR-SIFT descriptors on modified dataset 3.

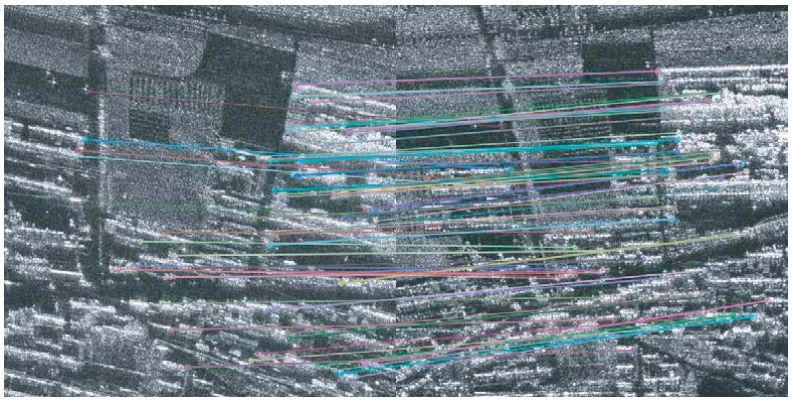


Figure 16. The matching results of DSP-LATCH descriptors on modified dataset 3.

two different antennas, which preliminarily validates the effectiveness of the proposed method. Then for the later two datasets, they are from different bands of two different time, which are very challenging for the gray level difference and time difference. In addition, we have changed the dataset 3 and re-conducted the experiments to show the superiority of DSP-LATCH to SIFT. The first image in dataset 3 is rotated by 15 degrees, and the second image is zoomed out for 75%. Because SIFT feature failed in obtaining the right matching points, the RMSE reached 207.20 pixels' error. Thus we compare the results with its variant, SAR-SIFT. The matching results of SAR-SIFT and DSP-LATCH are shown in Fig. 15 and Fig. 16. The quantitative results are shown in Table 5. The experiment using DSP-LATCH results in 135 pairs of matching points and 1.47 pixels RMSE, while SAR-SIFT yields 24 pairs of matching points and 2.93 pixels RMSE. Thus compared with SAR-SIFT feature, DSP-LATCH is also robust to rotation transformations.

In the experiments, we have used three sets of airborne SAR images both from the same band and different bands to demonstrate the effectiveness of the proposed coarse-to-fine registration method. It is worth to mention that our method can also be effective in handling the registration of multi-polarized and polarimetric SAR data. After the speckle noise reduction, the proposed SAR-FAST uses large

Table 5. The matching quantity and error of the 5 descriptors on modified dataset 3.

Descriptor	SIFT	LATCH	SAR-SIFT	DSP-SIFT	DSP-LATCH
Quantity	5	48	24	13	**135**
RMSE	207.20	2.02	2.93	3.20	**1.47**

detection window and radius to further decrease the possibility of detecting wrong feature corners. With the DSP scheme and patch matching in DSP-LATCH, the descriptor shares with great robustness and discrimination. By starting from the local invariant information in image itself, the feature based methods are particularly useful when coherence-based methods are limited by loss of coherence, i.e., for rapid and incoherent changes and large acquisition time intervals between the two SAR images. The framework can also work for multi-polarized and polarimetric SAR data.

5. CONCLUSION

In this paper, we propose a coarse-to-fine registration framework for airborne SAR images, which integrates SAR-FAST detector and DSP-LATCH descriptor. Without relying on any geocoding, the proposed method produces state-of-the-art results on challenging airborne SAR images. The SAR-FAST works well in terms of feature repeatability and rotation robustness. The DSP-LATCH descriptor outperforms the competitors in matching quantity and matching distribution, which results in better registration results. The experimental results on the airborne SAR image datasets demonstrate that the proposed method can work well based on the intensity data without any geo-reference data. In future work, we will exploit how to extend the distribution of matched points to further improve the registration accuracy and utilize our registration results to achieve automatic change detection.

ACKNOWLEDGMENT

The research was partially supported by the National Natural Science Foundation of China under Grant 61331016 and Grant 61771351, and the project for innovative research groups of the natural science foundation of Hubei Province (No. 2018CFA006). Sincere thanks are given to the anonymous reviewers and members of the editorial team for their comments and valuable contributions.

REFERENCES

1. Xiang, Y., F. Wang, and H. You, "OS-SIFT: A robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, No. 6, 3078–3090, 2018.
2. Marin, C., F. Bovolo, and L. Bruzzone, "Building change detection in multitemporal very high resolution SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53, No. 5, 2664–2682, 2015.
3. Hong, Z., W. Chao, Y. Tang, et al., "A new image registration method for multi-frequency airborne high-resolution SAR images," *IEEE International Geoscience and Remote Sensing Symposium*, 167–169, 2003.
4. Gao, G., X. Qin, and S. Zhou, "Modeling SAR images based on a generalized gamma distribution for texture component," *Progress In Electromagnetics Research*, Vol. 137, 669–685, 2013.
5. Wang, F., H. You, and X. Fu, "Adapted anisotropic Gaussian SIFT matching strategy for SAR registration," *IEEE Geoscience and Remote Sensing Letters*, Vol. 12, No. 1, 160–164, 2015.
6. Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, 91–110, 2004.
7. Lindeberg, T., "Scale-space theory: A basic tool for analyzing structures at different scales," *Journal of Applied Statistics*, Vol. 21, No. 1–2, 225–270, 1994.
8. Datcu, M., "Wavelet-based despeckling of SAR images using Gauss-Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45, No. 12, 4127–4143, 2007.
9. Yan, T., W. Yang, X. Yang, C. López-Matínez, H. C. Li, and M. Liao, "Polarimetric SAR despeckling by integrating stochastic sampling and contextual patch dissimilarity exploration," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 10, No. 6, 2738–2753, 2017.

10. Schwind, P., S. Suri, P. Reinartz, and A. Siebert, "Applicability of the SIFT operator to geometric SAR image registration," *Remote Sensing*, Vol. 31, No. 8, 1959–1980, 2010.
11. Wang, B., J. Zhang, L. Lu, G. Huang, and Z. Zhao, "A uniform SIFT-like algorithm for SAR image registration," *IEEE Geoscience and Remote Sensing Letters*, Vol. 12, No. 7, 1426–1430, 2015.
12. Dellinger, F., J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53, No. 1, 453–466, 2015.
13. Harris, C. and M. Stephens, "A combined corner and edge detector," *Alvey Vision Conference*, Vol. 3, 147–152, 1988.
14. Liu, Y., H. Yu, W. Yang, and L. Li, "SAR image registration using SAR-FAST corner detection," *Journal of Electronics and Information Technology*, Vol. 39, No. 2, 430–436, 2017.
15. Zhang, Q., X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," *European Conference on Computer Vision*, 815–830, 2014.
16. Bay, H., T. Tuytelaars, and L. van Gool, "SURF: Speeded up robust features," *European Conference on Computer Vision*, 404–417, 2006.
17. Dubois, C., A. Nascetti, A. Thiele, M. Crespi, and S. Hinz, "SAR-SIFT for matching multiple SAR images and radargrammetry," *PFG — Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, Vol. 85, No. 3, 149-158, 2017.
18. Liu, X., Z. Tian, Q. Lu, L. Yang, and C. Chai, "A new affine invariant descriptor framework in shearlets domain for SAR image multiscale registration," *AEU — International Journal of Electronics and Communications*, Vol. 67, No. 9, 743–753, 2013.
19. Fan, B., C. Huo, C. Pan, and Q. Kong, "Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT," *IEEE Geoscience and Remote Sensing Letters*, Vol. 10, No. 4, 657–661, 2013.
20. Dong, J. and S. Soatto, "Domain-size pooling in local descriptors: DSP-SIFT," *IEEE Conference on Computer Vision and Pattern Recognition*, 5097–5106, 2015.
21. Ye, Y., J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multi-modal remote sensing images based on structural similarity," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, No. 5, 2941–2958, 2017.
22. Rui, J., C. Wang, H. Zhang, and F. Jin, "Multi-sensor SAR image registration based on object shape," *Remote Sensing*, Vol. 8, 923, 2016.
23. Xiang, Y., F. Wang, L. Wan, and H. You, "An advanced rotation invariant descriptor for SAR image registration," *Remote Sensing*, Vol. 9, 686, 2017.
24. Calonder, M., V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," *European Conference on Computer Vision*, 778–792, 2010.
25. Leutenegger, S., M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," *IEEE International Conference on Computer Vision*, 2548–2555, 2011.
26. Alahi, A., R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," *IEEE Conference on Computer Vision and Pattern Recognition*, 510–517, 2012.
27. Levi, G. and T. Hassner, "LATCH: Learned arrangements of three patch codes," *IEEE Winter Conference on Applications of Computer Vision*, 1–9, 2016.
28. Chui, H. and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding*, Vol. 89, No. 2, 114–141, 2003.
29. Liu, S., G. Sun, Z. Niu, N. Li, and Z. Chen, "Robust rigid coherent point drift algorithm based on outlier suppression and its application in image matching," *Journal of Applied Remote Sensing*, Vol. 9, 095085, 2005.
30. Zhang, H., W. Ni, W. Yan, J. Wu, and S. Li, "Robust SAR image registration based on edge matching and refined coherent point drift," *IEEE Geoscience and Remote Sensing Letters*, Vol. 12, No. 10, 2115–2119, 2015.
31. Zhao, M., B. An, Y. Wu, B. Chen, and S. Sun, "A robust delaunay triangulation matching for multispectral/multidate remote sensing image registration," *IEEE Geoscience and Remote Sensing*

Letters, Vol. 12, No. 4, 711–715, 2015.

32. Pluim, J. P., J. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Transactions on Medical Imaging*, Vol. 22, No. 8, 986–1004, 2003.
33. Powell, M. J. D., "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The Computer Journal*, Vol. 7, No. 2, 155–162, 1964.
34. Liu, G., W. Yang, G.-S. Xia, and M. Liao, "Structure preserving SAR image despeckling via L0-minimization," *Progress In Electromagnetics Research*, Vol. 141, 347–367, 2013.
35. Walessa, M. and M. Datcu, "Model-based despeckling and information extraction from SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, No. 5, 2258–2269, 2015.
36. Fan, J., Y. Wu, M. Li, W. Liang, and Q. Zhang, "SAR image registration using multiscale image patch features with sparse representation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 10, No. 4, 1483–1493, 2017.
37. Rosten, E., R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 1, 105–119, 2010.
38. Rublee, E., V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," *IEEE International Conference on Computer Vision*, 2564–2571, 2011.
39. Brown, M., G. Hua, and S. Winder, "Discriminative learning of local image descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 1, 43–57, 2011.
40. Cover, T. M. and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 2006.
41. Gong, M., S. Zhao, L. Jiao, D. Tian, and S. Wang, "A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52, No. 7, 4328–4338, 2014.