*Progress In Electromagnetics Research M, Vol. 18, 259–269, 2011*

# SAR IMAGE MATCHING METHOD BASED ON IMPROVED SIFT FOR NAVIGATION SYSTEM

**S. Ren [1, *], W. Chang [1], and X. Liu [2]**

[1]School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, P. R. China

[2]Navy Submarine Academy, Qingdao 266071, P. R. China

**Abstract**—In order to ensure that SAR scene matching aided navigation system can acquire the position errors and yawing errors simultaneously, we propose an image matching algorithm based on Scale Invariant Feature Transform (SIFT). However, the SIFT is proposed for optical image, and its performance degrades when used in SAR image. To enhance the adaptability of SIFT, two ways are employed. One is the application of a preprocessing on image pairs before matching. The other is the establishment of a scale and rotation restriction criteria on tie-points after SIFT matching. Compared with other matching methods, experiment results show that the proposed method is much more suitable for SAR image and successes in matching performance improvement. Furthermore, the method can meet the real-time requirement.

## 1. INTRODUCTION

In the SAR scene matching aided navigation system, the navigation error is corrected by the fusion of current navigation information and position update information [1], where the latter is obtained through matching of SAR real-time image and a geo-referenced image, i.e., reference image. Therefore, image matching method is one of the key components of this system, which must satisfy three requirements: the real-time performance, the sub-pixel accuracy and the robustness.

In order to acquire the position errors and yawing errors simultaneously, matching methods based on point feature are regarded as a better choice. Sheng and Alsdorf [2] give an example of the regular matching process based on point feature, and indicates that a

*Received 27 April 2011, Accepted 15 June 2011, Scheduled 22 June 2011*

* Corresponding author: Sanhai Ren (sanhairen@163.com).

good point descriptor is in favor of improving matching performance. At present, many point descriptors have been developed [3], such as Moravec descriptor, Harris descriptor, Schmid descriptor, etc. By evaluating the performance of descriptors computed for local interest regions, Mikolajzyk and Schmid found that the SIFT-based descriptor performs better than others [4]. Scale Invariant Feature Transform (SIFT) feature is not only invariant to image rotation and scale change, but also partially invariant to illumination change and some affine distortion [5]. In order to improve the real-time performance of SIFT, integral image [6] and Principal Component Analysis (PCA) [7] methods are also added into the method.

SIFT is very powerful for optical image, and has exhibited great success in some optical image processing applications [5], such as view matching for 3D reconstruction, object recognition, image mosaic, duplicate detection, motion tracking and segmentation, navigation and guidance, and some others that require identification of matching locations between images. However, SAR image is quite different from optical image, even for high-resolution image and filter-processed image, as objects in SAR image are dimmer and their details are more ambiguous. In addition, SAR image has severe speckle noise, which differs from the noise of optical image [8]. Therefore, the performance degrades when SIFT is used in SAR image matching, such as the tie-points detected by SIFT are too less, or the ratio of false matches to total matches is too high.

The methods to eliminate false matches in optical image are exploited in two independent studies [9, 10], which are called SR-SIFT in this letter. Both of these methods use an eliminate-after-matching approach, hence they intend to reduce false matches in the SIFT output. Yi et al. [9] form a histogram of scale difference and define a window around the peak of this histogram. The matches with scale difference outside this window are rejected. A limitation in the method is that only images pairs with approximately the same scale are considered. Bastanlar et al. [10] make an improvement on the method proposed in [9]. They preprocess image pairs using scale ratio to adjust their scales, and then eliminate false matches as [9] does. Compared with the method proposed in [9], the method in [10] can not only eliminate false matches, but also increase the number of correct matches. However, the computation time is too long to be adopted in the navigation system. Furthermore, both of them are proposed for optical image, and the performance will degrade when used in SAR image.

To make SIFT more suitable for SAR image, several improvements were made on it, which is called ISIFT in this letter. Section 2

introduces the SIFT method and analyzes the problems when it is used in SAR image. Section 3 describes in detail of the improvements on SIFT. Section 4 completes some experiments and compares the results with others from recent published methods. Section 5 gives the conclusion.

## 2. PROBLEMS IN SIFT-BASED MATCHING

### 2.1. SIFT-based Matching

In the navigation system, the SIFT-based matching method needs to complete the SIFT feature extraction, the SIFT feature matching and the parameter calculation. The process is described here:

(i). SIFT feature detection: in this step, a Difference-of-Gaussian (DOG) scale space is firstly built by Gaussian filter and Gaussian pyramid down algorithm [5]. Then the key-points are detected as the minima or maxima compared with its 8 neighbors in the same scale and 9 corresponding neighbors at neighboring scale in this scale space. The points located on the unstable edge or with low contrast should be eliminated. With the nearby information of location, scale and principal curvatures, the sub-pixel location and precise scale of key-point can be acquired.

(ii). Orientation assignment: one or more orientations are assigned to each key-point location based on local image gradient directions. All future operations are performed on image data that has been transformed relating to the assigned orientation, scale and location for each feature, thereby providing invariance to these transformations.

(iii). SIFT feature descriptor: for each key-point, rotate its neighborhood block to its orientation, cut the block into $4 \times 4$ slices and compute the Histogram of Oriented Gradient (HOG) for each slice. The $4 \times 4$ HOGs form its descriptor.

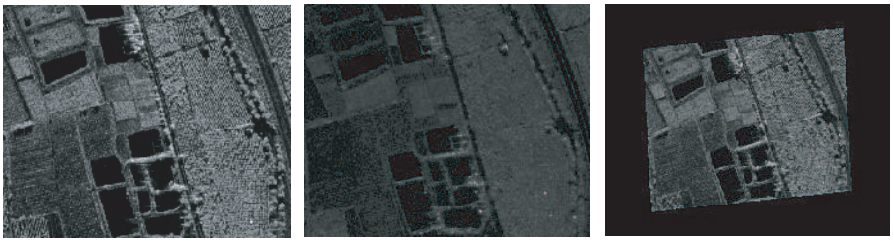
(iv). SIFT feature matching: the best candidate match for each key-point in real-time image is found by identifying its nearest neighbor in the database of key-points from reference image. The nearest neighbor is defined as the key-point with minimum Euclidean distance for the key-point descriptor.

(v). Parameter calculation: a set of tie-points can be acquired from (i)–(iv), then the translation and rotation angle between real-time image and reference image, which will be transformed into position error and yawing error of INS, can be calculated through the least-squares solution.

### 2.2. Problems in SIFT Matching

SIFT detects local extrema in DOG scale space, making it robust against additive noise. However, noise in SAR image follows multiplicative noise model, and cannot be mitigated by DOG processing. For instance, for the given image pair in Figures 1(a) and 1(b), the original SIFT achieves only 4 matches, which are not sufficient for parameter calculation, because at least 3–4 times of the least matches (6–8 matches) are needed for the reliability.

SIFT assigns a 128-element feature vector to each key-point as its descriptor. For each key-point in real-time image, we calculate its Euclidean distance with all key-points of reference image, and find the tie-point by judging whether the Euclidean distance of the closest neighbor to that of the second-closest neighbor is less than a threshold [5]. This is very effective for optical image. However, the performance degrades when it is used in SAR image. For example, there are 40 false matches out of 110 matches in Figures 1(a) and 1(c), where 12 of these false matches have a scale ratio (defined in formula (4a)) less than 0.59 or greater than 0.805, and 35 of these false matches have a rotation difference (defined in formula (4b)) less than 2° or lager than 6°, which are beyond the ranges for scale ratio and rotation difference as defined in the Step 4 of Section 3.



(a) (b) (c)

**Figure 1.** Image pairs. (a) is a SAR image of Ku band. (b) is a SAR image of L band. (c) is a transformed image from (a), where the scale factor is 0.7 and rotation angle is 4°.

Therefore, when SIFT is used in SAR image pairs, it might be not able to detect sufficient tie-points, and the ratio of false matches to total matches is too high. In order to make SIFT more suitable for SAR image, several improvements are made on it as described in the next section.

## 3. THE IMPROVEMENTS ON SIFT MATCHING

According to the character of SIFT, we preprocess SAR image pairs to transform their noise model, and then eliminate false matches using the dominant scale ratio and rotation difference, from which a significant improvement on SIFT matching is observed.

If we use $Y$ to denote the SAR image magnitude, and $X$ is the speckle-free value of magnitude of the electric field of the backscattered signal, then $Y$ is related to $X$ by the multiplicative model $Y = F \cdot X$. For a single-look image ($L = 1$), the normalized fading random variable $F$ obeys the Rayleigh distribution [11], as (1) shows.

$$P(F) = \frac{\pi F}{2} \exp\left(-\frac{\pi F^2}{4}\right) \quad F \geq 0 \tag{1}$$

In optical image, $Y$ is related to $X$ by the additive model $Y = F + X$, and $F$ follows the Gaussian distribution. In order to make SIFT more suitable for SAR image, logarithmic transformation is employed to convert a multiplicative speckle model to an additive noise model.

After a natural logarithmic transformation, SAR image can be converted into

$$\ln(Y) = \ln(F) + \ln(X) = \hat{F} + \hat{X} \tag{2}$$

where $\hat{F}$ follows the double exponential or Fisher-Tipper distribution, as (3) shows.

$$P(\hat{F}) = \frac{\pi e^{2\hat{F}}}{2} \exp\left(-\frac{\pi e^{2\hat{F}}}{4}\right) \quad L = 1 \tag{3}$$

It has been established that as the number of looks increase, the speckle random variable approaches a Gaussian distribution. In [11], a distance between cumulative distributions is computed to measure the deviation of the log-transformed speckle from Gaussianity. It is clearly demonstrated that for amplitude image, the log-transformed speckle noise is already statistically very close to the Gaussian pdf. A conclusion was drawn that the Gaussian approximation is relatively good when $L \geq 3$.

Consequently, the noise in the log-transformed SAR image can be considered as additive noise, and its Probability Density Function (PDF) is more similar to Gaussian distribution with the increasing number of SAR looks. One issue related to the logarithmic transformation arises from the fact that the mean of the log-transformed speckle noise is not zero, whereas a significant set of techniques assume Gaussian white noise with zero mean. Therefore,

we should subtract the mean value from the log-transformed image to make the noise with zero mean [12].

Each key-point detected by SIFT is expressed by a vector ($x_i$, $y_i$, $\sigma_i$, $\theta_i$). Elements of the vector denote location ($x_i$, $y_i$), scale $\sigma_i$ and orientation $\theta_i$. The definition of scale ratio $S_R$ and rotation difference $R_D$ of tie-points $P_1$ ($x_{1i}$, $y_{1i}$, $\sigma_{1i}$, $\theta_{1i}$) and $P_2$ ($x_{2i}$, $y_{2i}$, $\sigma_{2i}$, $\theta_{2i}$) are:

$$S_{Ri}(P_1, P_2) = \sigma_{1i}/\sigma_{2i} \quad i = 1, 2, \ldots, k \tag{4a}$$

$$R_{Di}(P_1, P_2) = \theta_{1i} - \theta_{2i} \quad i = 1, 2, \ldots, k \tag{4b}$$

where $k$ is the number of correct matches.

For correct matches, $S_R$ and $R_D$ should be similar or equal to the real value, as SIFT scale ratio and rotation difference reveals the scale ratio and rotation difference of features between image pairs. Therefore, the matches with $S_R$ and $R_D$ deviating largely from the real value should be eliminated.

In order to estimate the approximate scale ratio and rotation difference, histograms of $S_R$ and $R_D$ are formed respectively. For each histogram, the peak and two sub-peaks which are closest to it are gotten. Then a parabola is fitted to the three histogram values to interpolate the peak position for better accuracy.

The improved SIFT-based matching method contains five steps, summarized as follows:

Step 1: Preprocess the image pairs using a logarithmic transformation to convert the noise model from a multiplicative model to an additive model.

Step 2: Perform SIFT matching on the preprocessed image pairs, and plot histograms of $S_R$ and $R_D$ of tie-points.

Step 3: Extract the peaks of each histogram, and take them as the approximate scale ratio $P_s$ and rotation difference $P_r$ of the images to be matched. Then, take the $P_s$ and $P_r$ as the real value of the scale ratio and rotation difference of the matches.

Step 4: Eliminate the false matches whose $S_R$ are beyond the scope $[0.8 \times P_s, 1.2 \times P_s]$, or $R_D$ are beyond the scope $[P_r - 2^\circ, P_r + 2^\circ]$, where the scopes are acquired from statistical experiments [9].

Step 5: Calculate the translation and rotation angle between real-time image and reference image through the least-squares solution.

## 4. EXPERIMENTAL COMPARISON

From the above statements, two major improvements can be observed when ISIFT-based matching method is applied to SAR image:

1st. The preprocessing of image pairs increases the number of correct matches;

2nd. The application of scale and rotation restriction criteria decreases the ratio of false matches to total matches. Furthermore, it improves the accuracy of matching position and rotation angle.

Therefore, two experiments are used for validating the robustness of the proposed method.

### 4.1. Experiment 1

The first experiment is used for testifying the validity of the preprocessing method. Various kinds of SAR image pairs (128 × 128 pixels) are selected to do the experiment, as shown in Figures 1(a) and (b), and their results are shown in Table 1, where the left number of the '/' is the number of correct matches, and the right number of the '/' is the number of matches.

In the original SIFT method, there is no further processes to eliminate the false matches, so the number of correct matches and that of total matches is the same. In the SR-SIFT method (the method proposed in [10]), a filter is employed to the detected matches to eliminate the false ones, so the number of total matches is the same as that of the SIFT method, but the number of correct matches is less than that of the SIFT. A preprocessing method and a restriction criteria is employed in the ISIFT method, which makes both the number of total matches and that of correct matches increase greatly. From Table 1, we can also get that neither SIFT nor SR-SIFT can detect sufficient matches (6–8 matches) to calculate transformation parameters for some kinds of SAR image pairs.

**Table 1.** The number of correct matches and total matches.

| | farmland | mountain | bridge | city |
|---|---|---|---|---|
| SIFT | 4/4 | 23/23 | 4/4 | 2/2 |
| SR-SIFT | 3/4 | 18/23 | 3/4 | 2/2 |
| ISIFT | 40/65 | 100/140 | 22/40 | 10/15 |

### 4.2. Experiment 2

The second experiment is used for testifying the validity of scale, rotation restriction criteria. In the scene matching aided navigation system, the matching method is mainly used for calculating the translation and rotation angle between real-time image and reference

image, so the accuracy of matching position and rotation angle can be regarded as an index of the method's validity.

In order to perform simulation experiments on computer, a large number of image pairs are required. If image pairs already exist, serial segments of real-time image could be produced directly from them. However, sometimes there are not enough image pairs available but only a single image. Thus segments of real-time images can be created by cutting casually from the reference image with some artificial distortions. Figure 2 shows an example of such simulation.

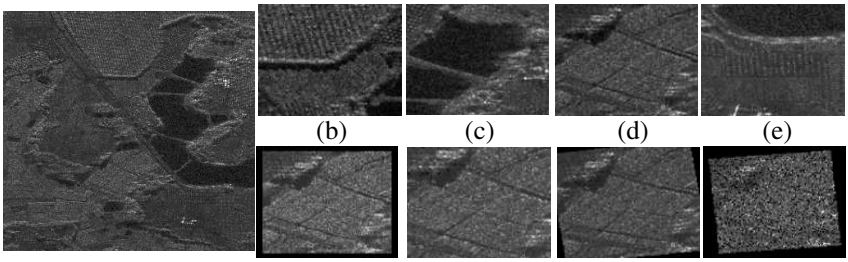



**Figure 2.** Example of created image pairs. (a) The reference image. (b)–(e) The sub-images intercepted from (a), from which the real-time images are created. (f)–(i) The simulated real-time images from (d), where (f) is the image zoomed out by 10%, (g) is the image zoomed in by 20%, (h) is the image rotated by 8°, (i) is the image zoom out by 10%, rotated by 5°, and added noise with variance 1.

An affine transformation is employed to form distorted images intercepted from reference image. The affine transformation from a reference image point $[x_i, y_i]$ to a real-time image point $[x_o, y_o]$ can be written as:

$$\begin{bmatrix} x_o \\ y_o \end{bmatrix} = S \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{5}$$

where the model translation is $[t_x, t_y]$ and the affine rotation, scale are represented by $[\theta, S]$.

Based on the matches detected by ISIFT, the transformation parameters can be solved. Let $m_1 = S\cos\theta$, $m_2 = S\sin\theta$, then the equation above can be rewritten to gather the unknowns into a column vector:

$$\begin{bmatrix} x_{o1} \\ y_{o1} \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} x_{i1} & -y_{i1} & 1 & 0 \\ y_{i1} & x_{i1} & 0 & 1 \\ & \cdots & & \\ & \cdots & & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ t_x \\ t_y \end{bmatrix} \tag{6}$$

We can rewrite this linear system as

$$\mathbf{A} = \mathbf{H} \cdot \mathbf{P} \tag{7}$$

where $\mathbf{A} = \begin{bmatrix} x_{o1} \\ y_{o1} \\ \vdots \end{bmatrix}$, $\mathbf{H} = \begin{bmatrix} x_{i1} & -y_{i1} & 1 & 0 \\ y_{i1} & x_{i1} & 0 & 1 \\ & \cdots & & \\ & \cdots & & \end{bmatrix}$, $\mathbf{P} = \begin{bmatrix} m_1 \\ m_2 \\ t_x \\ t_y \end{bmatrix}$, then the unknowns can be determined by

$$\mathbf{P} = \left[\mathbf{H}^T\mathbf{H}\right]^{-1}\mathbf{H}^T\mathbf{A} \tag{8}$$

The translation $[t_x,\ t_y]$ can be got directly, and the angle $\theta$ can be determined by solving the following equation

$$\theta = \mathrm{tg}^{-1}(m_2/m_1) \tag{9}$$

Figure 2(a) ($600 \times 600$ pixels) is a selected reference image. In order to simulate real-time images, blocks of sub-images with $128 \times 128$ pixels are intercepted casually from the reference image before thrown kinds of aberrations at random, as shown in Figure 2(b)–Figure 2(e). These aberrations include geometrical distortion, gray aberration and resolution changing. Figure 2(f)–Figure 2(i) show some examples of such simulative real-time images. The statistical results of the proposed method and other matching methods are shown in Table 2, where the matching with error less than 3 pixels is considered as a correct match; the angle accuracy is the mean angle error between the calculated value and the real one.

All the algorithms are run in MATLAB on an Intel Pentium 4/2.4 GHz processor, 2 GB physical memory with Windows OS. In Table 2, the matching probabilities of the three methods are all "1", which means all of them can give correct matching. However, the matching position and angle will be transformed to the position errors and the yawing errors of SAR platform, which are fused with the

**Table 2.** The statistical results.

| Performance \ Method | | SIFT | SR-SIFT | ISIFT |
|---|---|---|---|---|
| Matching probability | | 1 | 1 | 1 |
| Position accuracy (pixel) | Row | 2.184 | 0.895 | 0.486 |
| | Column | 2.406 | 1.106 | 0.354 |
| Angle accuracy (°) | | 0.152 | 0.102 | 0.081 |
| Time (s) | | 0.245 | 1.256 | 0.261 |

current navigation estimates to correct the navigation error [13]. The more accurate matching position and angle are, the more improved navigation precision will be achieved. Compared with SIFT and SR-SIFT, the ISIFT can obtain the position accuracy at sub-pixel level, and improve the angle accuracy to 0.01°. Though the performance of SR-SIFT is also at a high level, its computation time is too long compared with SIFT and ISIFT. Therefore, the ISIFT method can get the best navigation precision.

## 5. CONCLUSION

In this paper, an improved SIFT-based matching method is presented as an enhancement of the scene matching method for navigation. Some problems when SIFT is used in SAR image are analyzed, then a preprocessing method and a scale, rotation restriction criteria is proposed as a solution. Plentiful SAR images of multi-frequency-band and different distortions are created to complete simulative experiments. Compared with the results of different matching methods, the improved SIFT-based matching method can not only increase the number of correct matches, but also improve the matching accuracy. Furthermore, there is almost no increase in computation time.

## REFERENCES

1. James, E. B. and A. M. Charles, "Precision aided inertial navigation using SAR and digital map data," *IEEE Position Location and Navigation Symposium*, 490–496, Las Vegas, NV, USA, 1990.
2. Sheng, Y. W. and D. E. Alsdorf, "Automated georeferencing and orthorectification fo Amazon basin-wide SAR mosaics using SRTM DEM data," *IEEE Trans. Geosci. Remote Sens.*, Vol. 43, No. 8, 1929–1940, 2005.
3. Mikolajczyk, K. and C. Schmid, "An affine invariant interest point detector," *ECCV 2002*, 128–142, Berlin Heidelberg, Germany, 2002.
4. Mikolajczyk, K. and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 27, No. 10, 1615–1630, 2005.
5. Lowe, D., "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, Vol. 60, 91–110, 2004.

6. Grabner, M., H. Grabner, and H. Bischof, "Fast approximated SIFT," *Asian Conferonce on Comput. Vis.*, 918–927, Hyderabad, India, 2006.
7. Ke, Y. and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *Proceedings of IEEE Comput. Soci. Conference on Comput. Vis. Pattern. Recog.*, 506–513, Washington D.C., USA, 2004.
8. Lu, J., B. Wang, H. M. Gao, and Z. Q. Zhou, "SAR images matching based on local shape descriptors," *IET International Radar Conference*, 235–238, Guilin, China, 2009.
9. Yi, Z., Z. G. Cao, and X. Yang, "Multi-spectral remote image registration based on SIFT," *Electron. Lett.*, Vol. 44, No. 2, 107–108, 2008.
10. Bastanlar, Y., A. Temizel, and Y. Yardimci, "Improved SIFT matching for image pairs with scale difference," *Electron. Lett.*, Vol. 46, No. 5, 346–348, 2010.
11. Xie, H., L. Pierce, and F. Ulaby, "Statistical properties of logarithmically transformed speckle," *IEEE Trans. Geosci. Remote Sens.*, Vol. 40, No. 3, 721–727, 2002.
12. Xie, H., L. Pierce, and F. Ulaby, "SAR speckle reduction using wavelet denoising and Markov random filed modeling," *IEEE Trans. Geosci. Remote Sens.*, Vol. 40, No. 10, 2196–2212, 2002.
13. James, E. B. and A. M. Charles, "Precision aided inertial navigation using SAR and digital map data," *IEEE Position Location and Navigation Symposium*, 490–496, Las Vegas, NV, USA, 1990.