

MOTIF: MULTI-ORIENTATION TENSOR INDEX FEATURE DESCRIPTOR FOR SAR-OPTICAL IMAGE REGISTRATION

Yongxiang Yao¹, Bin Zhang¹, Yi Wan¹, Yongjun Zhang^{1,2*}

¹ School of Remote Sensing and Information Engineering, Wuhan University, 430079 Wuhan Hubei Province, China – (yaoyongxiang, bin.zhang, yi.wan, zhangyj)@whu.edu.cn

² Hubei LuoJia Laboratory, 430079 Wuhan Hubei Province, China

Commission II, WG II /1

KEY WORDS: SAR image, Multi-orientation tensor, Multi-orientation tensor index feature, Bilateral matching, Registration

ABSTRACT:

The inherent speckle noise in synthetic aperture radar (SAR) images and the significant differences between SAR and optical images in nonlinear radiation give rise to the great difficulty in computing similarity between image features, improving detection accuracy of corresponding points and the efficiency of image matching, thus making the registration of SAR and optical images a long-standing challenging task. To address these issues, a new SAR-optical image registration method was proposed in this paper, namely, Multi-orientation Tensor Index Feature (MoTIF), which is characterized by a lightweight feature descriptor. Specifically, we firstly established a diffusion tensor model based on the information of image gradient orientation. Then, the model was parameterized using polar coordinates to help identify the MoTIF and get the array of indices of maximum value, with which we could draw a multi-orientation index map and thereupon construct the feature vector descriptors. To evaluate the proposed method, seven representative SAR-optical image pairs were tested along with a comparison with other four state-of-the-art methods. Results show that our MoTIF method outperforms the other methods in that it substantially de-speckles SAR images, overcomes nonlinear radiation distortions caused by the differences between SAR and optical images, and achieves high precision and efficiency in image registration. The average number of correct matches (NCM) of 151.0 and the root of mean-squared error (RMSE) of 1.66 pixels obtained by utilizing MoTIF with lower time consumption adds more evidence to its superior performance. The time consumption of the MoTIF method is better than that of the other four methods, and the calculation speed is 4 times faster than that of the LGHD method. Executable code and test data are published in the link <https://skyeart.org/publication/project/MoTIF/>

1. INTRODUCTION

The rapid development of computer science, remote sensing, artificial intelligence as well as other technologies brings about a great variety of multi-source images obtained by different kinds of sensors. Among those images, optical and SAR images and their joint use have been the heated topics receiving substantial scholarly attention, for their matching and registration can provide technical support for fields like 3D reconstruction, rescue and relief work, urban planning (Ye et al., 2017). Drawing on the time lag and spatial differences of two image sensors in obtaining information, we can enhance the temporal and spatial resolution of data sources and improve the reliability of data application and processing, thus giving full play to their respective advantages. However, SAR is usually pointed to the side for imaging, so it is not hard to figure out why it is strongly affected by the ground range and terrain relief. SAR-optical registration remains still difficultly formidable in view of the multiplicative noise caused by speckles distributed in the SAR image and the geometric and non-linear radiation distortions in its imagery.

In recent years, a lot of research have been carried out on SAR-optical image registration, and the methods used can be generally classified into three categories: area-based matching, deep learning-based matching, and feature-based matching. The area-based matching method features metrics like correlation coefficient, mutual information, etc. (Öfverstedt et al., 2019.

Viola et al., 1997). Commonly used for its accuracy and efficiency in identifying grayscale changes, the correlation coefficient cannot be used, however, to locate the nonlinear grayscale differences between SAR and optical images due to its high sensitivity to between-image discrepancies. In contrast, metrics of mutual information to non-linear gray-scale differences produce consistently sharper peaks in the surface of the similarity measurement, but it is not favored all the time because it often gets stuck at a locally optimal solution without giving a more global one. Deep learning-based method has been widely used in multi-model image matching and other materials, witnessed by the proposal of several robust deep learning algorithms, such as convolutional neural network matching (K. Yi et al., 2016), multi-source image feature extraction and description D2-Net network (Dusmanu et al., 2019), deep matching network based on co-attention (Wiles et al., 2021), VGG network feature extraction matching (Efe et al., 2021) and weak texture matching method LoFTR based on transformer network (J. Sun et al., 2021). At the same time, some deep learning matching methods based on SAR-optical have also achieved good results, such as conditional generative adversarial networks (Merkle et al., 2018), pseudo-siamese convolutional neural network (Hughes et al., 2018), convolutional neural network suitable for SAR-Optical match (Bürgmann et al., 2019), CorrASL network (Hughes et al., 2020). Nonetheless, deep learning-based method also has the problem of poor stability in matching brought by the complexity

* Corresponding author

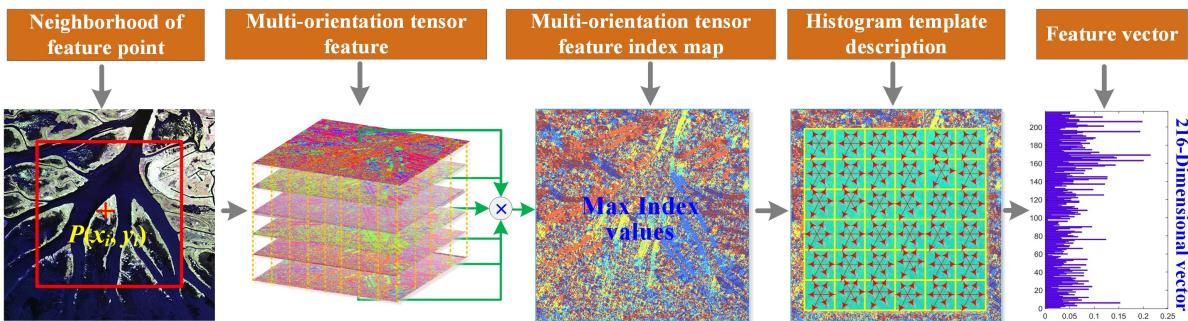


Figure 1. Construction flowchart of the MoTIF descriptor.

and unpredictability of scenarios where SAR-optical images are acquired.

Feature-based method such as SAR-scale-invariant feature transform (SIFT) (Dellinger et al., 2015), optical-to-SAR-SIFT (OS-SIFT) (Xiang et al., 2018), A new matching framework (Bas et al., 2021), Rotation-invariant self-Similarity descriptor (Mohammadi et al., 2022), have limitations in themselves as well due to their strong dependence on image gradients and high sensitivity to SAR-optical imaging matching. To address this, methods based on phase features have been put forward, including LGHD (Aguilera et al., 2015), channel features of orientated gradients (Ye et al., 2019), radiation-variation insensitive feature transform (RIFT) (Li et al., 2020), and histogram of absolute phase consistency gradients (HAPCG) (Yao et al., 2021). These methods make good use of the frequency-domain features of the image to make multi-source matching feasible. They did yield a great number of positive results, yet their deficiency of sensitivity to the multiplicative noise by registration of SAR and optical images, especially the registration taking place in non-urban, cannot be ignored. As is summarized above, area-based method enjoys good resistance to nonlinear radiation distortions, with obviously poor performance in the matching efficiency and the elimination of the coherent speckle of SAR images. Deep learning-based method has shown its great potential in image matching, but it is subject to the sample size and the complex scenarios where SAR-optical images get matched. Feature-based method is not quite useful in tackling problems such as multiplicative noise and nonlinear radiation distortions, though it makes feature matching faster. Clearly, if we want to further the research on SAR-optical image registration, we should firstly deal with how to achieve the accurate positioning of corresponding points and greater efficiency of matching both at once.

This paper presents a lightweight SAR-optical registration method based on a multi-orientation diffusion tensor index feature (MoTIF) description. It uses the diffusion coupled with the parametric expression of polar coordinates to construct rich multi-orientation tensor features, by which the maximum index value can be calculated and a corresponding descriptor would be generated as well. During this process, the registration of SAR-optical images could be advanced in a more accurate and efficient manner.

2. IMAGE REGISTRATION BASED ON MOTIF

The proposed MoTIF method is composed of four steps: (i). feature points extraction, (ii). MoTIF descriptor construction, (iii). bilateral matching and outliers removal, (iv). image fusion. The second step is the major focus of this paper which will be elaborated in the following section.

2.1 Feature point extraction

Feature point extraction is an important part in image matching. It is also a demanding task because of the significant nonlinear radiation differences between SAR and optical remote sensing images, and the multiplicative noise caused by the coherent speckles in the SAR images. Therefore, the anisotropic weighted moment image space is used to extract image features (Yao et al., 2021), and it is defined as (1):

$$W = \frac{1}{2} (M_{\max} + M_{\min} + \Delta\Phi * (M_{\max} - M_{\min})) \quad (1)$$

where W represents the final anisotropic weighted moment result; M_{\max} represents the maximum moment of phase consistency of the image; M_{\min} represents the minimum moment of phase consistency of the image; $\Delta\Phi$ represents the image weight coefficient (with the range of [-1...5]).

After the image's anisotropic weighted moments are established, the FAST (Rosten et al., 2006) operator is used to extract key points.

2.2 MoTIF descriptor construction

Although the previously established feature descriptors have achieved better results than traditional feature descriptors, they are still limited by different conditions, resulting in matching of SAR-optical images that does not satisfy actual production requirements. Hence the construction of a robust descriptor has been the major problem in SAR-optical image registration and yet to be solved. This part focuses on the construction of MoTIF descriptors to mainly address geometric deformations, non-linear radiation differences and speckle noise inherent in SAR images. The construction was conducted through the following three steps as shown in Figure 1: (i). multi-orientation tensor features construction, (ii). multi-orientation tensor index mapping, (iii). descriptor vector calculation.

2.2.1 Multi-orientation tensor features construction.

Step 1: Tensor feature calculation. The image gradients cannot be directly relied on in image registration owing to their high sensitivity to image distortions, especially to speckle noise in SAR image. To this end, some scholars have employed Sobel and Laplacian operators (Ma et al., 2016) for gradient optimization and filtering, making its use more promising in cross-modal image matching. In this section, the second-order gradient was firstly calculated, followed by the computing of the second-order gradient amplitude in the horizontal and vertical directions via using the Sobel template [-1,0,1; -2,0,2; -1,0,1] (see equation (2)).

$$\mathbf{G}(x, y)_\sigma^1 = \sqrt{(\mathbf{L}(x, y) \cdot \sigma \cdot \nabla S_x)^2 + (\mathbf{L}(x, y) \cdot \sigma \cdot \nabla S_y)^2} \quad (2)$$

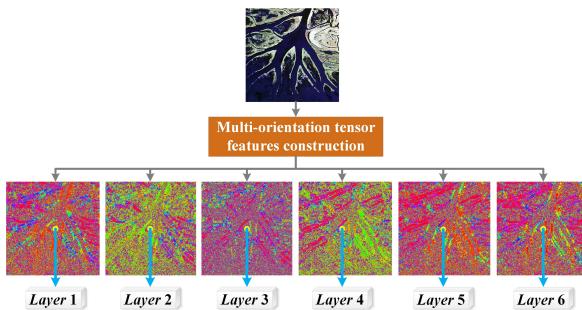


Figure 2. Construction process of multi-orientation tensor feature sets.

where $\mathbf{G}(x, y)_\sigma^1$ denotes the first-order gradient, $\mathbf{L}(x, y)$ represents the grayscale of the image and σ is the standard deviation of Gaussian distribution. ∇S_x and ∇S_y denote the Sobel template in the horizontal and vertical directions respectively. The more detailed calculation is presented in the equation (3):

$$\begin{cases} \mathbf{L}(x, y)_{xx} = [\mathbf{G}(x, y)_\sigma^1 \otimes \nabla S_x]^2 \\ \mathbf{L}(x, y)_{yy} = [\mathbf{G}(x, y)_\sigma^1 \otimes \nabla S_y]^2 \\ \mathbf{L}(x, y)_{xy} = \mathbf{G}(x, y)_\sigma^1 \otimes \nabla S_x \cdot \mathbf{G}(x, y)_\sigma^1 \otimes \nabla S_y \end{cases} \quad (3)$$

where $\mathbf{L}(x, y)_{xx}$ and $\mathbf{L}(x, y)_{yy}$ represent the sum of squares of second order gradients in the x and y directions respectively. $\mathbf{L}(x, y)_{xy}$ denotes trace of the second-order gradient and \otimes denotes the convolution operator.

The tensor provides the edge information in terms of its shape and direction. Despite its shape changes with the contrast and illumination, edge direction always remains unaltered. Consequently, the tensor model is frequently used in extracting the structural features of the image (Köthe et al., 2003). The definitive structure tensor expression is given as equation (4):

$$\mathbf{T}(x, y) = \begin{bmatrix} G_\sigma \otimes \mathbf{L}(x, y)_{xx} & G_\sigma \otimes \mathbf{L}(x, y)_{xy} \\ G_\sigma \otimes \mathbf{L}(x, y)_{yx} & G_\sigma \otimes \mathbf{L}(x, y)_{yy} \end{bmatrix} \quad (4)$$

Where, G_σ is the Gaussian kernel function with standard deviation σ ; The results of $\mathbf{L}(x, y)_{yx}$ gets the identical result with $\mathbf{L}(x, y)_{xy}$. We also calculated the image tensor, the parallel and orthogonal eigenvectors of the tensor $\mathbf{T}(x, y)$ and denoted the latter two as $\mathbf{V}_p(x, y)$ and $\mathbf{V}_o(x, y)$.

Step 2: The coherent speckle noise of the SAR image still existed even when we finished the calculation of the image tensor features. It will undermine the robustness of the registration of the SAR-optical image we introduced the coherence-enhancing diffusion function (Weickert et al., 1999) to retain image edges features by reducing the multiplicative speckle noise in SAR images, particularly in their uniform area. This function defined as equation (5) and (6).

$$\mathbf{D} = [\mathbf{V}_p(x, y) \quad \mathbf{V}_o(x, y)] \begin{bmatrix} G_{\lambda_1}(\lambda_1, \lambda_2) & 0 \\ 0 & G_{\lambda_2}(\lambda_1, \lambda_2) \end{bmatrix} [\mathbf{V}_p(x, y)^T \quad \mathbf{V}_o(x, y)^T] \quad (5)$$

$$\begin{cases} G_{\lambda_1}(\lambda_1, \lambda_2) = \alpha \\ G_{\lambda_2}(\lambda_1, \lambda_2) = \alpha + (1-\alpha) e^{-\frac{C_m}{k^{(\lambda_1-\lambda_2)^m}}} \end{cases} \quad (6)$$

Where, \mathbf{D} is the tensor matrix after coherence enhancement diffusion. α permits a small diffusivity (usually $\alpha=0.05$) even when no preferential direction exists and k acts as a threshold to

$(\lambda_1 - \lambda_2)^2$ value ($k=125$). The m is 2. The C_m is 2.33666.

$\mathbf{V}_p(x, y)$ represents the parallel feature vector of the image tensor. $\mathbf{V}_o(x, y)$ is the orthogonal feature vector of the image tensor. T is the transpose operator of the matrix. Moreover, the positive constant is introduced to correct the bias in the original Perona-Malik diffusivity function (Weickert et al., 1999).

Step 3: The enhanced diffusion tensor features of SAR-optical images can be obtained computationally by Step 2. This operation weakens the effect of multiplicative noise, leaving problems such as nonlinear radiometric distortion and geometric distortion between SAR-optical unsolved. On that account, Ye et al. (2019) have attempted at enriching the image structure features by generating multi-directional gradient features. Based on their findings, this paper adopts a parametric representation of polar coordinates to generate create the map sets centered on the MoTIF.

Firstly, the \mathbf{D} -matrices obtained from Step 2 are decomposed along the x -directions and y -directions, which are denoted as

\mathbf{CoT}_x and \mathbf{CoT}_y . Then, the feature images are rotated in the orientation range $[0 \sim \pi]$ with the rotation interval angle (π/o) . After the rotation is completed, the fast fourier transform is executed to further filter the fine noise, and the final equation is shown in (7).

$$\mathbf{F}_o(x, y) = \left\| FFT \left[(\cos(\frac{\pi}{o}) \otimes \mathbf{CoT}_x) + (\sin(\frac{\pi}{o}) \otimes \mathbf{CoT}_y) \right] \right\| \quad (7)$$

Where, $\mathbf{F}_o(x, y)$ is the eigenvalue of the diffusion tensor of the o -th layer. o is the number of layers of the multi-orientation tensor feature (The value of o is taken as 6 in this paper). $\cos(\cdot)$ and $\sin(\cdot)$ are signs of trigonometric functions. FFT is a Fast Fourier Transform function. $\|\cdot\|$ is the sign for absolute value. The construction of multi-orientation tensor feature sets has been completed, as shown in Figure 2.

2.2.2 Multi-orientation tensor index map. Although values of coherence-enhancing diffusion tensor feature under different layers were obtained, it is still difficult to directly describe the between-layer feature similarity of SAR-optical images, which are easily affected by multiplicative noise and nonlinear radiation distortions. To enhance the robustness of the descriptor, using index features between different coherence-enhancing diffusion tensor feature maps is necessary and effective. The values of corresponding coherence-enhancing diffusion tensor features of each pixel p , and the maximum value of each pixel to construct a feature map were all calculated. This paper is divided into 6-layers into multi-orientation tensor feature layers.

Firstly, a multi-channel feature map was formed with reference to the features of the $\mathbf{F}_o(x, y)$ after the calculation of coherence-enhancing diffusion tensor features in Step 2. The mathematical formula is defined as equation (8):

$$\mathbf{MoTF}(x, y) = \{\mathbf{F}_i(x, y)\}, \quad i=[1, \dots, o] \quad (8)$$

Then, the maximum value of each pixel of different images is counted and its location is marked. Finally, the values of each pixel location in the image, the corresponding pixel value is obtained, and an o -dimensional ordered array were obtained. The channel index value $\mathbf{MoTF}(x, y)$ which incorporated the maximum value of the o -dimensional array was calculated. The mathematical expression is defined as equation (9):

$$\mathbf{MoTF}(x|_{\max}, y|_{\max}) = \text{argmax}(\mathbf{MoTF}_o(x, y)) \quad (9)$$

Where, $\text{MoTIF}(x|_{\max}, y|_{\max})$ denotes the multi-orientation tensor index map result. The superscript $o = [1, 2, 3, \dots, 6]$ indicates different channels. A multi-orientation tensor index map was finally made after the completion of the above operations on all pixels.

2.2.3 Feature vector statistics. The proposed MoTIF descriptor (a 216-unit vector) is generated through calculating the histogram of oriented gradient (Dalal et al., 2005) on the multi-orientation tensor index map. For each feature point, their location serves as the central point for determining the neighborhood window range. The Gaussian function is used to assign weight to each pixel to make the feature description vectors consistent when they are confronted with the changes of the window position. Since the value range of the multi-orientation tensor index map is set from 1 to o , the partial image block is divided into $o \times o$ sub-regions. Then a histogram vector of $n=6$ is fixed in each subregion, and all histogram vectors are sequentially connected to form $o \times o \times n$ -unit feature vector and normalize the feature vector. Finally, a 216-dimensional descriptor vector was generated as shown in Figure 1.

2.3 Bilateral Matching

After the MoTIF descriptor calculation is completed, feature matching ensues. In this paper, we used the similarity measure of Euclidean distance and the matching strategy of the bilateral matching method making sure the one-to-one correspondence of obtained matching points. Besides, outliers are unavoidable after the bilateral matching. For their removal we used the fast sample consensus (FSC) algorithm (Wu et al., 2014) to cast out wrong matchings. The FSC algorithm can steadily extract the correct matching point pairs from mismatches with fewer iterations.

2.4 Image fusion

When the correct correspondence points are obtained, the transformation matrix between images needs to be calculated to achieve the image fusion. Here Affine model the chessboard grid fusion (Li et al., 2017) method were employed for the former was used to calculate the homography matrix between images, while the latter for the registration of SAR-optical images.

3. EXPERIMENTAL RESULT

Four state-of-the-art methods, i.e., OS-SIFT (Xiang et al., 2018), LGHD (Aguilera et al., 2015), RIFT (Li et al., 2020) and HAPCG (Yao et al., 2021) methods, were used for comparison. During the tests, the feature point extraction threshold was set to 0.4 the image scale difference was set to 1.6 with six image multi-orientation tensor feature maps and 72 pixels concerning the neighborhood window. The parameters of the compared methods were adjusted to the optimal stage accordingly. The proposed MoTIF method, OS-SIFT, LGHD, RIFT and HAPCG were implemented in Matlab-R2018a. When those methods, the number of matched key-points were kept under 3500. The experiments were performed on a Dell-G3 laptop with an Intel(R) Core(TM) i7-9750H CPU, 16GB-RAM, and Windows 10 x64 operating system.

Image-space affine transformation was used to model the geometric relationships of image pairs. For each pair, over 15

well-distributed corresponding points were manually collected to calculate the affine transformation as the ground truth, which is used to measure the location accuracy of the automatically matched points. Three indices, i.e. the number of correct matches (NCM), the root of mean-squared error (RMSE) of the correct matches, and matching time (MT) are used to quantitatively evaluate the performance of several methods.

3.1 Image Datasets

The test data consists of seven SAR-optical image pairs acquired respectively by SAR and optical sensors. The SAR images are made up of the data from German Terra SAR-X satellite and Chinese Gaofen-3 satellite, while the optical data are from Google Earth data and Chinese Gaofen-2. These images, with three types of low-resolution, medium-resolution, and high-resolution, have different spectral characteristics covering different scenes such as urban areas, suburban areas and mountainous areas. Notably, differences in the imaging mechanisms between SAR and optical images bring about the significant nonlinear radiation variances amid each pair. We hence categorized the data into two groups in line with their geometric differences. The first group contains four pairs, namely, no. 1, no. 2, no. 6, and no. 7, whose sizes from 520 to 1000 pixels. These images, though substantially similar in rotation and scale, cannot be registered immediately for SAR images are strongly affected by coherent speckles. The registration of no. 6 and 7 image pairs with unclear structural features seems more challenging for the information they offer is all about mountains with no buildings captured. The second group has three pairs, i.e. no. 3, 4 and 5, whose sizes varying from 450 to 923 pixels. As shown in Figure 3, images of no. 3 pair have a temporal difference of more than one year, images of no. 4 pair have a scale difference, and images of pair 5 has a rotational difference.

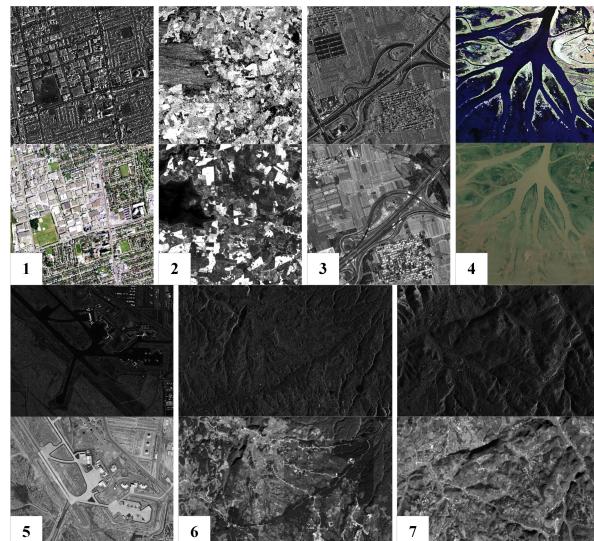
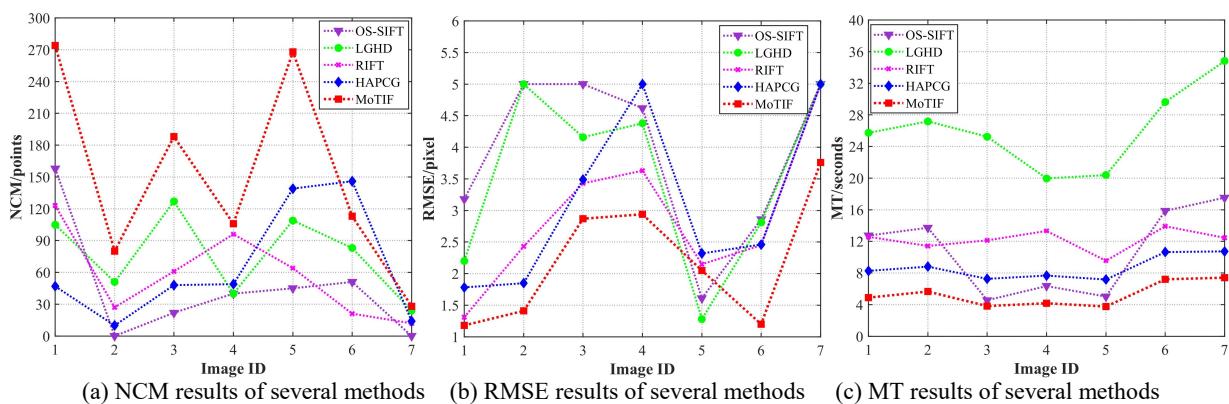


Figure 3. SAR-optical datasets.

3.2 Quantitative Results

Figure 4 illustrates the matching results of three indices used to compare the proposed MoTIF method with other four state-of-the-art methods. The unit of NCM (see Figure 4 (a)), RMSE (see Figure 4 (b)) and MT (see Figure 4 (c)) is the number of points, pixel and seconds.



(a) NCM results of several methods (b) RMSE results of several methods (c) MT results of several methods

Figure 4. Quantitative results of several methods.

As shown in Figure 4 that OS-SIFT is superior to LGHD, RIFT and HAPCG on MT while the latter three work better on NCM and RMSE. However, MoTIF outperforms these four methods on all three indices and it is remarkably less time-consuming. At the same time, the average scores of the seven groups of image pairs in the three indices were further calculated with the aim of comparing several methods more comprehensively.

as its NCM is increased by approximately 1 to 2.3 times, RMSE accuracy increased by 0.25 to 0.9 times, and the average MT increased by 0.64 to 4 times. It can thus be concluded that MoTIF, a lightweight feature registration method, makes possible the robust registration of SAR-optical images in a faster fashion.

Drawing on, the correct correspondence points obtained, we then calculated the homography matrix between images which was used to promote image fusion (see Figure 5). Figure 5 (a) is the distribution of ground truth points, and Figure 5 (b) is the matching result of the SAR-optical image.

Figure 5 demonstrates that MoTIF works well on robust matching and fusion of images with scale and rotation differences. It boosts the accurate fusion of SAR-optical images with no mismatch or artifacts, as is proofed by maps (c) and (d).

4. PARAMETRIC ANALYSIS

To fully evaluate the performance of the proposed method, we analyzed its two core parameters, the size of the neighborhood windows and the number of calculations required by MoTIF, which would be estimated by two indicators, RMSE and NCM. Settings of the parameters are shown in Table 2.

Methods	OS-SIFT	LGHD	RIFT	HAPCG	MoTIF
Pair 1	✓	✓	✓	✓	✓
Pair 2	✗	✗	✓	✓	✓
Pair 3	✗	✓	✓	✓	✓
Pair 4	✓	✓	✓	✓	✓
Pair 5	✓	✓	✓	✓	✓
Pair 6	✓	✓	✓	✓	✓
Pair 7	✗	✗	✗	✗	✓
Success rate	4/7	5/7	6/7	6/7	7/7
NCM	45.14	77.00	57.71	64.71	151.00
RMSE	3.18	2.83	2.20	2.41	1.66
MT	10.83	26.14	12.20	8.66	5.28

Table 1. Average quantitation matching scores of seven groups SAR-optical images of our MoTIF and four competitors.

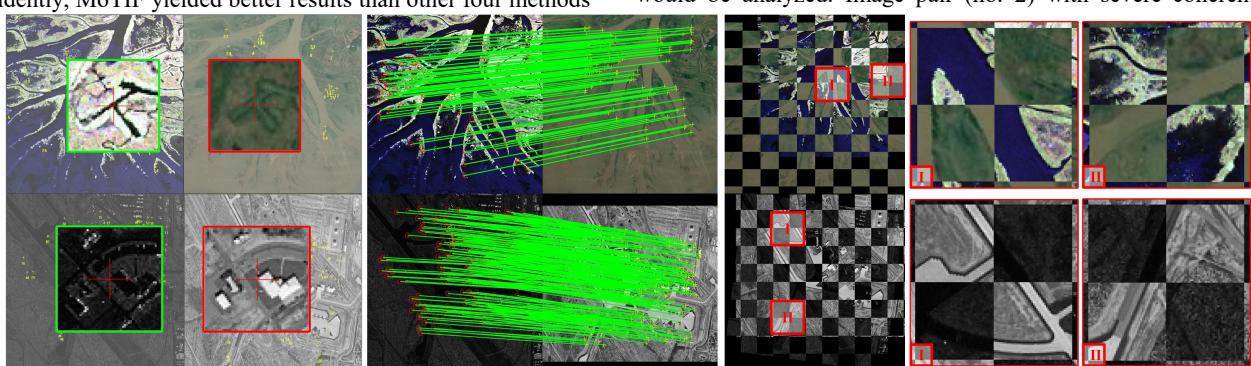
Note that, "✗" failed pair matching. The value of NCM is positively associated with the performance, the higher the number, the better the performance. As of RMSE, the RMSE results higher than 5 are set to 5 pixels.

From Table 1, seven pairs are perfectly matched through the use of MoTIF method, evidenced by the average NCM, average RMSE and average MT of 151.00, 1.66 pixels and 5.28 seconds. Evidently, MoTIF yielded better results than other four methods

Parameter Name	Symbol	Variable Value	Fixed Parameters
Neighborhood window	N_w	[32,40,48, 56,64,72, 80,88,96,104, 120,130,140]	$N_L = 6$
Layers number of MoTIF	N_L	[3,4,5,6,7,8,9,10,11,12]	$N_w = 72$

Table 2. Matching parameters configuration.

The performance of MoTIF under different parametric settings would be analyzed. Image pair (no. 2) with severe coherent



(a) Ground truth points (b) Matching results of MoTIF method (c) Images fusion (d) Local subplots of the fusion results
 Figure 5. Matching and registration results of SAR-optical image.

speckle noise is used as an example. Figure 6 shows that as window N_w increases from 32 to 64, NCM results also increase with a steady upward trend. When the window N_w reaches 64 and beyond, NCM goes higher than 70. When the window N_w equals 88, NCM arrives at its maximum. But when N_w exceeds 130, NCM begins to decrease. As for RMSE, $N_w = 72$ is the most favorable in that RMSE on that point gets to its optimal value, 1.41 pixels. Therefore 72 is recommended when setting N_w .

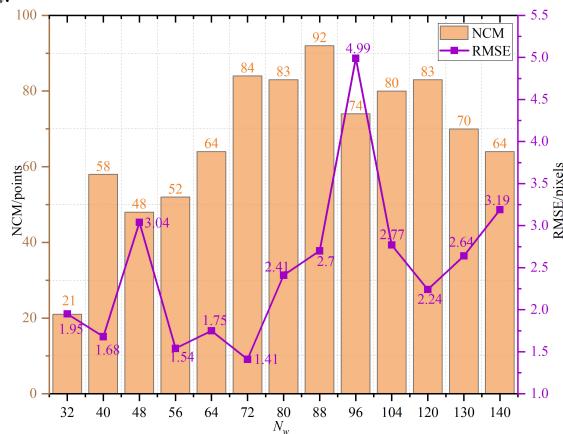


Figure 6. Results of different neighborhood windows.

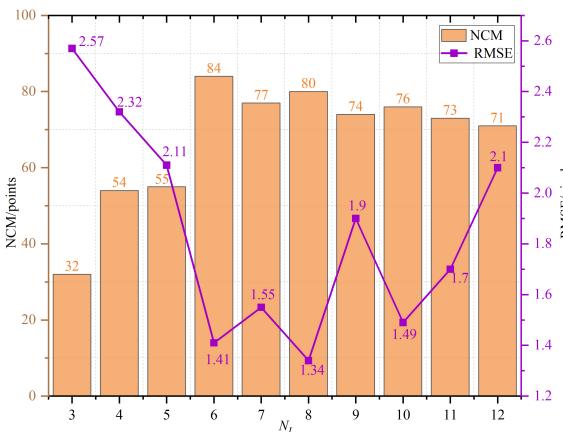


Figure 7. Results of the MoTIF layers number.

Figure 7 demonstrates that RMSE won't go lower than 2 pixels when the value of N_L is set minimal or maximal. When N_L ranges from 6 to 8, it fluctuates between 1.3 and 1.6 but NCM goes higher than 75 which is surely unsatisfactory. Only when N_L equals 6, can we get good values of RMSE and NCM both at once. As the N_L increases, more corresponding points will be brought while the matching growth rate is declining and the increase in the number of matching points can no longer compensate for the deterioration of the matching accuracy itself. Therefore, $N_L=6$ is recommended.

5. CONCLUSION

This paper proposed a novel lightweight SAR-optical image registration method mainly to resolve the problems of significant nonlinear radiation distortions and multiplicative noise caused by speckles inherent in SAR images. Firstly, the tensor features of the image to collect information on its salient structure were extracted. Then we used the coherence enhancement diffusion model to cope with the strong effect from the coherent speckle. Finally, enriched features of the images were obtained by performing MoTIF extraction through

using polar coordinates. The maximum index value was also calculated so the descriptor vectors can be established. Experimental results show that MoTIF boasts better accuracy and efficiency in SAR-optical image registration than OS-SIFT, LGHD, RIFT, and HAPCG, etc. With this method, NCM has increased approximately by 1 to 2.3 times, the accuracy of RMSE improved by 0.25 to 0.9 times and the average MT increased significantly by 0.64 to 4 times.

However, the MoTIF method also has its limitations. It is not so useful when applied in the registration of images whose scale- and rotation- distortions cannot be corrected or eliminated merely based on their locations or attitudes. Future research could address this issue of the scale and rotation invariance to help further MoTIF and its application in SAR-optical and other multi-modal images registration.

ACKNOWLEDGEMENTS

This work was supported by the Key Program of the National Natural Science Foundation of China under project number 42030102, the National Natural Science Foundation of China under project number 42192583 and 42001406, Basic Research Strengthening Program of China (173 Program) 2020-JCJQ-ZD-015-00-04, the Fund for Innovative Research Groups of the Hubei Natural Science Foundation under project number 2020CFA003, and China Postdoctoral Science Foundation under Grant No. 2020 M672416.

REFERENCES

- Ye, Y., Shen, L., Hao, M., 2017. Robust Optical-to-SAR Image Matching Based on Shape Properties. *IEEE Geosci. Remote Sens. Lett.*, 14(4): 564-568.
- Öfverstedt, J., Lindblad, J., Sladoje, N., 2019. Fast and robust symmetric image registration based on distances combining intensity and spatial information. *IEEE Transactions on Image Processing*, 28(7): 3584-3597.
- Viola, Paul., Wells, III., William, M., 1997. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2): 137-154.
- Yi, K., M., Trulls, E., Lepetit, V., 2016. Lift: Learned invariant feature transform. *European conference on computer vision*. Springer, Cham, 467-483.
- Dusmanu, M., Rocco, I., Pajdla, T., 2019. D2-net: A trainable cnn for joint description and detection of local features. *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*. pp.8092-8101. (16 June 2019).
- Wiles, O., Ehrhardt, S., Zisserman, A., 2021. Co-Attention for Conditioned Image Matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 15920-15929. (19 June 2021).

- Efe, U., Ince, K., G., Alatan, A., 2021. DFM: A Performance Baseline for Deep Feature Matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 4284-4293. (19 June 2021).
- Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X. 2021. LoFTR: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 8922-8931. (19 June 2021).
- Merkle, N., Auer, S., Müller, R., & Reinartz, P. 2018. Exploring the potential of conditional adversarial networks for optical and SAR image matching. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(6), 1811-1820.
- Hughes, L. H., Schmitt, M., Mou, L., Wang, Y., & Zhu, X. X. 2018. Identifying corresponding patches in SAR and optical images with a pseudo-siamese CNN. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 784-788.
- 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.
- Hughes, L. H., Marcos, D., Lobry, S., Tuia, D., & Schmitt, M. 2020. A deep learning framework for matching of SAR and optical imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 166-179.
- Dellinger, F., Delon, J., Gousseau Y., 2015. SAR-SIFT: A SIFT-like algorithm for SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1): 453-466.
- Xiang, Y., Wang, F., You, H., 2018. 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*, 56(6), 3078-3090.
- Bas, S., & Ok, A. O. 2021. A new productive framework for point-based matching of oblique aircraft and UAV-based images. *The Photogrammetric Record*, 36(175), 252-284.
- Mohammadi, N., Sedaghat, A., & Jodeiri Rad, M. 2022. Rotation-invariant self-Similarity descriptor for multi-temporal remote sensing image registration. *The Photogrammetric Record*, 37(177), 6-34.
- Aguilera, C. A., Sappa, A. D., Toledo, R., 2015. LGHD: A feature descriptor for matching across non-linear intensity variations. *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp.178-181. Quebec, QC, Canada. (10 December 2015).
- 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.
- J, Li., Q, Hu., M, Ai., 2020. RIFT: Multi-Modal Image Matching Based on Radiation-Variation Insensitive Feature Transform. *IEEE Transactions on Image Processing*, 29: 3296-3310.
- Yao, Y, X., Zhang, Y, J., Wan, Y., 2021. Heterologous Images Matching Considering Anisotropic Weighted Moment and Absolute Phase Orientation. *Geomatics and Information Science of Wuhan University*, 46(11):1727-1736.
- Rosten, E., Drummond, T., 2006. Machine learning for high-speed corner detection. In *European conference on computer vision*. pp. 430-443. Springer, Berlin, Heidelberg. (May 2006).
- W, Ma, Z., Wen, Y., Wu, L., Jiao, M., Gong, Y, Zheng., L, Liu., Remote sensing image registration with modified SIFT and enhanced feature matching, *IEEE Geoscience and Remote Sensing Letters*., vol.14, no.1, pp.3-7.
- Köthe, U., Edge and junction detection with an improved structure tensor. In *Joint Pattern Recognition Symposium*. pp. 25-32. Springer, Berlin, Heidelberg. (September 2003).
- Weickert, J., 1999. Coherence-enhancing diffusion filtering. *International journal of computer vision*, 31(2), 111-127.
- Wu, Y., Ma, W., Gong, M., Su, L., Jiao, L., 2014. A novel point-matching algorithm based on fast sample consensus for image registration. *IEEE Geoscience and Remote Sensing Letters*, 12(1), 43-47.
- N, Dalal., B, Triggs., 2005. Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, 1(12): 886-893.
- Wu, Y., Ma, W., Gong, M., Su, L., Jiao, L., 2014. A novel point-matching algorithm based on fast sample consensus for image registration. *IEEE Geoscience and Remote Sensing Letters*, 12(1), 43-47.
- Li, J., Hu, Q., Ai, M., Zhong, R., 2017. Robust feature matching via support-line voting and affine-invariant ratios. *ISPRS Journal of Photogrammetry and Remote Sensing*, 132, 61-76.