



High-Precision Pixelwise SAR–Optical Image Registration via Flow Fusion Estimation Based on an Attention Mechanism

Qiuze Yu, Yuxuan Jiang , Wensen Zhao, and Tao Sun 

Abstract—Due to the severe speckle noise and complex local deformation in synthetic aperture radar (SAR) images, the problem of high-precision pixelwise registration (dense registration) between SAR and optical images remains far from resolved. In this article, an attention mechanism based optical flow fusion algorithm is proposed to achieve high-precision dense SAR–optical image registration. First, two descriptors, the scale-invariant feature transform (SIFT) and a descriptor based on phase congruency (PC), are used to describe SAR and optical images to eliminate their intensity differences. Then, a salient feature map is extracted as a query matrix to weight the optical flow energy function. When extracting the salient feature map, the Contour Robuste d’Ordre Non Entier detector and the ratio of exponentially weighted averages operator are used to eliminate additive and multiplicative noise in the optical and SAR images, respectively. Finally, the optical flow fields based on SIFT and the PC-based descriptor are fused to compensate for registration ambiguity. Experimental results show that our method is feasible, effective, and robust to noise, and it enables high-precision registration under local deformation.

Index Terms—Optical flow, pixelwise registration (dense registration), SAR and optical, synthetic aperture radar (SAR) image.

I. INTRODUCTION

WITH the development of aerospace technology and multisource observations, the amounts of multiplatform, multiband, multispectrum, and multiperiod image data available are increasing. Remote imaging sensors include optical sensors, synthetic aperture radar (SAR), and infrared radar, each of which has its own advantages. As an active imaging system, SAR is not affected by weather or time and provides high-resolution multiview, multiband, and multipolarization images. Because of these advantages, SAR is used in many remote sensing applications. An optical sensor is a passive sensor that can obtain rich grayscale and texture information, given good ground conditions. Similarly, each type of sensor also has shortcomings.

Manuscript received January 1, 2022; revised March 2, 2022, April 1, 2022, and April 19, 2022; accepted April 29, 2022. Date of publication May 5, 2022; date of current version May 26, 2022. This work was supported by the Open Project Program Foundation of the Key Laboratory of Opto-Electronics Information Processing, Chinese Academy of Sciences under Grant OEIP-O-202009. (Corresponding author: Yuxuan Jiang.)

The authors are with the School of Electronic Information, Wuhan University, Wuhan 430072, China (e-mail: yuhenry007@whu.edu.cn; 2016301200139@whu.edu.cn; 2017301200192@whu.edu.cn; suntao@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3172449

SAR images are complex, and the information they contain is not as rich as that in optical images. However, optical sensors can be blocked by clouds and cannot adapt to all-weather applications.

Because of the complementary advantages of different sensors, multisource information fusion has become an increasingly important image processing approach that is applied in many fields, such as change monitoring [1], image fusion [2], and target recognition [3]. One of the core tasks in multisource fusion applications is the high-precision registration of multisource images (such as SAR and optical images), especially pixelwise registration (dense registration).

Image registration [4] is the process of matching and overlapping two or more images acquired at different times, by different sensors or under different conditions [5] (weather, illumination, camera location, angle, etc.). This process is widely used in remote sensing data analysis, computer vision, image processing, and other fields [6]. In SAR image processing, image registration is generally included as a preprocessing step, and its accuracy has a great impact on subsequent applications. Image registration techniques can be divided into sparse registration and dense registration depending on whether all pixels are used for registration. Sparse registration is based on certain feature points that are extracted either manually or automatically and represented as a vector or tensor. Then, similarity calculations are performed to match these feature points between two images. The advantage of this approach is that it significantly reduces the computational complexity and the time required for registration. In contrast to sparse registration, all pixels are used in dense registration. The advantages of this approach are that local deformations can be modeled more precisely and it yields more accurate parameters for each pixel.

In recent years, learning-based methods have achieved success in a variety of scenarios [7], [8]. Ma *et al.* [9] utilized the visual geometry group network [10] to calculate approximate spatial relationships. Quan *et al.* [11] proposed a generative matching network to generate coupled optical and SAR images, thereby improving the registration quality. Merkle *et al.* [12] also achieved excellent results by utilizing conditional generative adversarial networks. However, an optical–SAR benchmark with dense ground-truth correspondences is still lacking, and this situation makes learning-based methods less feasible for the dense optical–SAR registration task. Therefore, this article focuses on traditional methods.

Generally, image registration is mainly performed between homogeneous images such as SAR images. For instance, SAR-SIFT [13] is an algorithm developed based on the scale-invariant feature transform (SIFT) [14] that improves the gradient definition in SIFT and has a good effect on SAR images. Moreover, Fan *et al.* [15] proposed an SAR image registration method based on SIFT, nonlinear diffusion [16], and phase congruency (PC) [17]. This algorithm uses nonlinear diffusion to eliminate the influence of noise in SAR images and a PC descriptor to screen the SIFT matching results, and it achieves good results.

In recent decades, numerous researchers have begun to focus on the registration of multisource images, such as SAR–optical registration [18]–[22], and convincing results have been reported. Li *et al.* [23] used the contours from an optical image as the initial conditions and applied an active contour model to obtain accurate contour locations in corresponding SAR images. Suri *et al.* [24] proposed a solely histogram-based method to achieve automatic registration. Huang *et al.* [25] used shape context to achieve registration. Ye *et al.* [26] applied the histogram of oriented phase congruency (HOPC) descriptor to improve the internal structure of SIFT. This approach uses the invariant structural features in multisource images in place of the descriptions in SIFT. Based on the concept of feedback, the improved iterative SIFT descriptor [27] is another extension of the SIFT framework.

In contrast, dense SAR–optical registration mainly relies on the optical flow field [28] to describe local deformation. The optical flow algorithm was first proposed in the field of video processing to estimate the motion field. Because of its ability to accurately estimate the local deformation field, it is also utilized for image registration. In the field of dense multisource image registration, Brigot *et al.* [30] proposed a local optical flow method based on the Lucas–Kanade (LK) algorithm [29] called GeFolki, which has a multiscale implementation and specific filters, including rank filtering, rolling guidance filtering, and local contrast inversion. In the SIFT flow algorithm proposed by Liu *et al.* [31], an entire image is described using the SIFT descriptor to eliminate the intensity differences between heterogeneous images. The description is optimized by means of a Markov random field with multiscale characteristics. Xiang *et al.* [32] proposed an improved SIFT flow algorithm called OS-flow. It includes both the Horn–Schunck implementation [33] and the LK implementation, whereas the SIFT flow utilizes only the latter. Additionally, it abandons the SIFT descriptor and uses the gradient location and orientation histogram (GLOH) descriptor [34] instead. The above methods utilize feature descriptors to eliminate the intensity differences between heterogeneous images and a multidimensional feature vector in place of the original gray value. For denoising, OS-flow uses a ratio operator to remove the multiplicative noise in SAR images, while GeFolki uses its own specific filters.

This article focuses on dense SAR–optical image registration with the ability to estimate the local deformation field, for which three challenges must be addressed. 1) Some salient edges may be distorted during optical flow estimation. 2) The multiplicative noise in SAR images needs to be considered. 3) There are obvious intensity differences between SAR and optical images.

To overcome these challenges, a novel algorithm for dense SAR–optical image registration based on an attention mechanism [35] and the optical flow field is proposed. The purpose of an attention mechanism is to highlight specific image regions. In this article, we propose a novel method of calculating and applying attention such that some salient edges can be maintained while reducing distortion. This method requires a corresponding query matrix that is assigned based on a proposed saliency map, which is extracted from structural information and is robust to noise. Moreover, SIFT and a proposed descriptor based on PC are applied to eliminate the intensity differences between SAR and optical images because these descriptors are both able to map the intensity to a modality-independent and discriminative vector form. SIFT, which considers the gradient histograms in local regions, is widely applied in image registration. PC can be used to model edges by means of a phase-based energy function, and this approach focuses more on single pixels and also achieves excellent results. To mitigate the potential detrimental impacts of the domain description property of SIFT, a fusion method is adopted.

The contributions of this article are summarized as follows.

- 1) A novel saliency map extraction method is proposed to obtain salient structural information.
- 2) For optical flow estimation, an attention calculation and application method is proposed with the aim of preserving salient edges.
- 3) A new descriptor based on PC is proposed, which is applied to mitigate the detrimental impacts of the domain description property of SIFT by fusing optical flow fields.

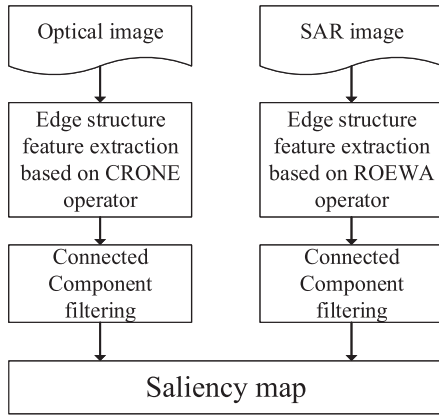
II. METHODOLOGY

In this section, feature maps that aim to eliminate the intensity differences between SAR and optical images are constructed via SIFT and a proposed PC-based descriptor. Then, a salient feature map is extracted by refining the edge structural information. For this calculation, a filter based on fuzzy theory is proposed and applied. Next, the optical field is estimated using an attention mechanism, and our proposed method of attention calculation and application is described. Finally, we fuse the optical flow fields to obtain a robust result. A flowchart of the proposed algorithm, which consists of all of the steps above, is shown in Fig. 1.

A. Feature Map Construction Based on SIFT and a PC-Based Descriptor

SIFT is the most widely used feature extraction technique in image registration because it is invariant with respect to scale, rotation, and illumination changes. The process of SIFT can be divided into three steps: detection of feature points, description of feature vectors, and identification of matching points by comparing feature vectors. In this article, only the feature description step is addressed. For each pixel in an image, its neighborhood (e.g., 16×16) is divided into an array of 4×4 units, the possible directions are quantized into 8 units, and finally, a $4 \times 4 \times 8 = 128$ -D vector is obtained as the SIFT representation of the pixel. As a result of this calculation process, the structural information

A. Saliency map extraction



B. Fusion flow estimation using attention mechanism

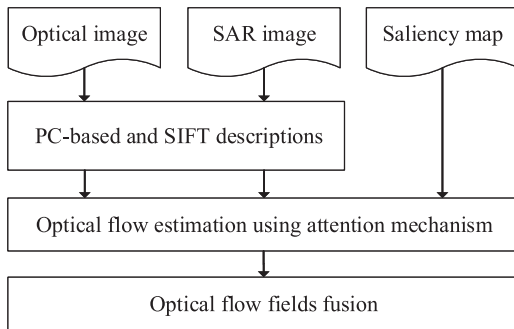


Fig. 1. Flowchart of the proposed algorithm.

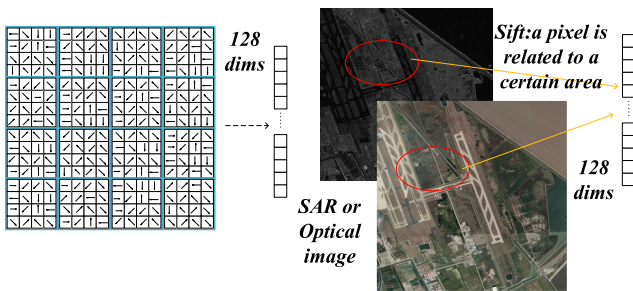


Fig. 2. Domain description property of SIFT, which introduces twist into the optical flow optimization algorithm.

contained in the resulting SIFT representation corresponds not only to a single pixel but also to the region around it. This is the domain description property of SIFT, which introduces twist into the optical flow optimization algorithm. Fig. 2 illustrates this.

PC [17] is a feature detection model based on the postulate that the feature points of an image occur in locations where the Fourier phase information is highly consistent. The expression for PC is as follows:

$$PC_{2D}(x) = \frac{\sum_j E_{\theta_j}(x)}{\sum_j \sum_n A_{n,\theta_j}(x) + \epsilon} \quad (1)$$

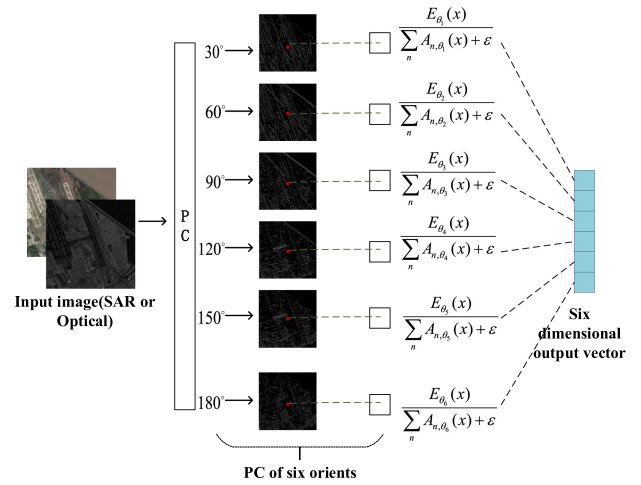


Fig. 3. Our proposed PC-based descriptor.

$$A_{n,\theta_j}(x) = \sqrt{e_{n,\theta_j}(x)^2 + o_{n,\theta_j}(x)^2} \quad (2)$$

$$E_{n,\theta_j}(x) = \sqrt{F_{n,\theta_j}(x)^2 + H_{n,\theta_j}(x)^2} \quad (3)$$

$$F_{n,\theta_j}(x) = \sum_n e_{n,\theta_j}(x), \quad H_{n,\theta_j}(x) = \sum_n o_{n,\theta_j}(x). \quad (4)$$

Here, $e_n(x)$ is the even-symmetric log Gabor wavelet filter response (real part), and $o_n(x)$ is the odd-symmetric filter response (imaginary part).

PC can be calculated in multiple directions. Based on this fact, we propose a descriptor that can better characterize a single pixel than the SIFT descriptor can. We take $\pi/6$ as the unit and compute PC in six directions, and the gray values of the image in each direction are taken as the elements of the feature vector. Finally, we obtain a 6-D vector. Fig. 3 illustrates this process.

Due to the underlying principle of PC, the proposed descriptor yields a stronger description of a single pixel and a weaker description of the surrounding region. Consequently, it plays a limited role in capturing relatively large-scale deformations but can preserve edges well. Kovessi [17] constructed a model of additive noise and removed the additive noise by subtracting the noise term from the energy function. However, in actual operations, we have found that although this step can eliminate noise, it also eliminates a large amount of useful information that is very important for optical flow estimation. Therefore, we retain the original result and instead perform denoising using a salient feature map.

B. Computation of the Salient Feature Map Based on Adaptive Edge Extraction

In the process of image registration using the SIFT flow algorithm, edge distortion often occurs, as shown in Fig. 4. This phenomenon is common in dense registration using optical flow optimization and descriptors with the domain description property, such as SIFT. Such an optimization algorithm treats blurred regions and salient regions equally. Moreover, the feature descriptions of adjacent pixels will be similar when a descriptor with the domain description property is used.

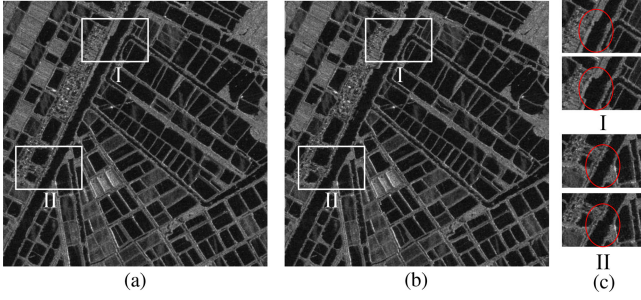


Fig. 4. Excessively twisted result caused by dense registration. (a) Original SAR image. (b) Result of SIFT flow. (c) Image patches. The upper patch is cut from (a), and the lower patch is cut from (b).

To preserve obvious edge information, a salient feature map that captures the edge structural features in both the SAR and optical images can be utilized. In this article, the Contour Robuste d'Ordre Non Entier (CRONE) detector [36] is used to extract edges from optical images, and the ratio of exponentially weighted averages (ROEWA) operator is used to extract edges from SAR images. The extracted edges are then superimposed and denoised to obtain the final salient feature map.

1) *Edge Structural Features of an Optical Image Based on the CRONE Detector*: The CRONE detector is a fractional calculus mask operator. Compared with the integer differential approach, the fractional differential approach can be used to extract edge information from an image while also retaining the weak edge information of smooth regions. In CRONE, a derivative order between 1 and 2 is favorable for detection selectivity. The horizontal and vertical components of the mask are calculated as follows:

$$X_{\text{mask}} = [-a_m, \dots, -a_2, -a_1, 0, +a_1, +a_2, \dots, +a_m] \quad (5)$$

$$Y_{\text{mask}} = [-a_m, \dots, -a_2, -a_1, 0, +a_1, +a_2, \dots, +a_m]^T. \quad (6)$$

Here, a_k is a real binomial coefficient, where n is the order of the derivative

$$a_k = (-1)^k \binom{n}{k} = (-1)^k \frac{n!}{k!(n-k)!}. \quad (7)$$

After the edges have been calculated using the CRONE operator, many messy lines can be observed. These lines are meaningless and need to be filtered out. However, merely performing another edge extraction step has little effect on the extracted information. Therefore, further filtering, such as filtering for connected components, needs to be performed based on the spatial information of the pixels themselves. Since the usefulness of a connected component is a fuzzy concept, we imitate the If–Then statements used in fuzzy theory. The rules are given as follows:

$$\text{If } size \leq \mu_1, \text{ Then } component = noise \quad (8)$$

$$\text{If } (size > \mu_1) \wedge (size \leq \mu_2),$$

$$\text{Then } component = weak\ edge \quad (9)$$

$$\text{If } size > \mu_2, \text{ Then } component = edge. \quad (10)$$

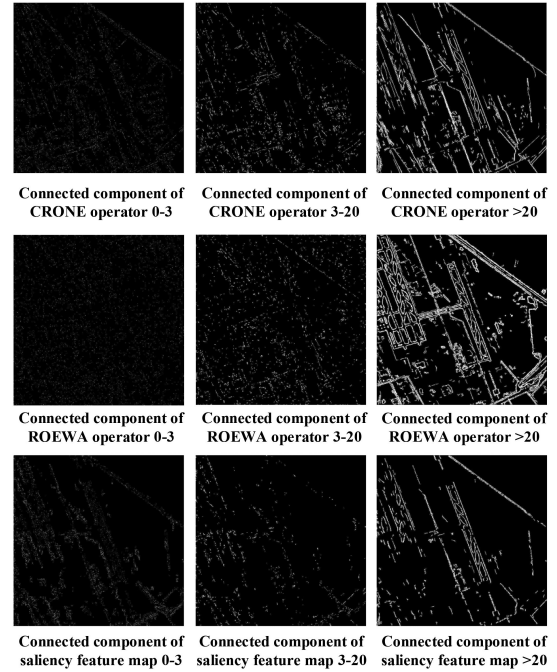


Fig. 5. Optical image (first row), SAR image (second row), and the saliency maps (third row) corresponding to different connected component sizes.

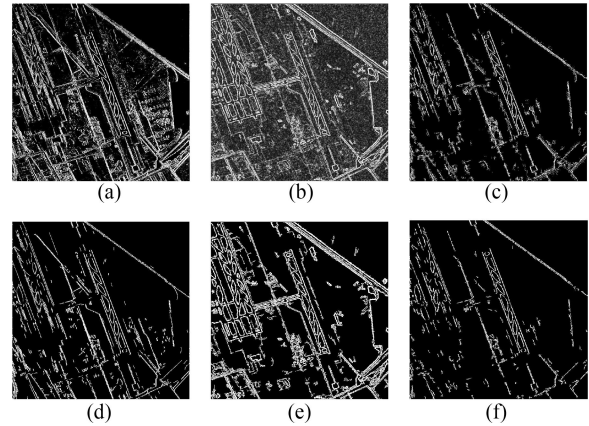


Fig. 6. Connected component filtering. (a–c) Before filtering, (a) CRONE, (b) ROEWA, and saliency map. (d–f) After filtering, (d) CRONE, (e) ROEWA, and (f) saliency map.

Here, *size* refers to the number of pixels that the component contains, and μ_1 and μ_2 are artificially specified thresholds. Based on our experience, μ_1 is set to 3, and μ_2 is set to 20. The performance of this connected component filtering method is illustrated in Figs. 5 and 6.

2) *Edge Structural Features of an SAR Image Based on the ROEWA Operator*: Due to the speckle nature of the noise in SAR images, a ratio operator such as ROEWA, which is an improved variant of the ratio of averages operator based on a multiedge model, is generally used for SAR image denoising

$$\overline{u_{x1}}(x, y) = f_1(x) * (f(y) \cdot Z(x, y)) \quad (11)$$

$$\overline{u_{x2}}(x, y) = f_2(x) * (f(y) \cdot Z(x, y)) \quad (12)$$

$$f(y) = \frac{1}{(1+b)}f_1(y) + \frac{b}{(1+b)}f_2(y+1) \quad (13)$$

$$f_{2-D}(x, y) = f(x)f(y). \quad (14)$$

Here, f_1 and f_2 are causal and anticausal filters, respectively. More details can be found in [37].

3) *Salient Feature Map*: After edge extraction, template matching [38], [39] is performed to obtain the global displacement, and overlapping parts are retained. In brief, the results need to be filtered using an algorithm described based on fuzzy theory. In this way, the salient edges in the optical and SAR images are identified, and the result is called the salient feature map, which can be regarded as a supporting tool for image registration.

C. Flow Fusion Estimation Using an Attention Mechanism

1) *Energy Function and Attention Mechanism*: Optical flow refers to the instantaneous velocity of a spatially moving object as represented by pixel motion in an image. In the optical flow method, the optical flow vector is calculated by using changes in the pixel scale and the corresponding relationships between the previous and current frames. There are two basic assumptions [40]–[44].

- 1) Constant brightness. When the same target moves between different frames, its brightness does not change. This assumption can be written as follows:

$$I(x, y, t) = I(x + u, y + v, t + 1). \quad (15)$$

Here, $I(x, y, t)$ is the intensity of pixel (x, y) , t is the time, and u and v represent the flows in the x -direction and the y -direction, respectively. Through first-order Taylor expansion, formula (16) can be simplified to formula (17)

$$I(x, y, t) = I(x, y, t) + u \frac{\partial I}{\partial x} + v \frac{\partial I}{\partial y} + 1 \frac{\partial I}{\partial t} \quad (16)$$

$$u \frac{\partial I}{\partial x} + v \frac{\partial I}{\partial y} + 1 \frac{\partial I}{\partial t} = 0. \quad (17)$$

- 2) Small displacement. The displacement between adjacent frames is assumed to be relatively small.

The energy function of the optical flow method is generally divided into a data term E_{term} and a space term E_{space} , which satisfy assumptions (1) and (2), respectively

$$E(u, v) = E_{\text{term}} + E_{\text{space}} \quad (18)$$

$$E_{\text{term}} = \sum_p \min(\|s_1(p) - s_2(p + w(p))\|_1, t) \quad (19)$$

$$E_{\text{space}} = \sum_p \eta(|u(p)| + |v(p)|) + \sum_{(p,q) \in \epsilon} [\min(\alpha|u(p) - u(q)|, d) + \min(\alpha|v(p) - v(q)|, d)]. \quad (20)$$

Here, p and q refer to the grid coordinates of the previous and current frames. The optical flow method was first used in video processing. For image registration, the real image is treated as the previous frame, while the reference image is treated as the current frame. Instead of the gray value, s_1 and s_2 are used,

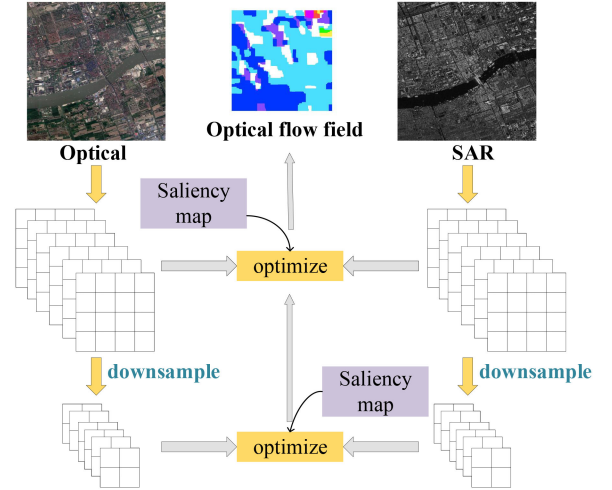


Fig. 7. Optical flow estimation using an attention mechanism. The saliency map is used as the query matrix.

referring to the feature spaces of the real and reference images, respectively, to address the problem of intensity differences.

All pixels are considered in the computation. However, it is clearly unreasonable to treat salient regions and uninteresting regions equally; the former contains more useful information than the latter and should, therefore, receive more attention. Accordingly, we use the salient feature map and the attention mechanism to accomplish this

$$\text{Attention}(Q, K, V) = W(Q, K) \cdot V \quad (21)$$

$$W(Q, K) = \lambda \cdot \text{step}(Q \wedge K). \quad (22)$$

An attention function can be described as mapping a query Q and a set of key–value pairs K – V to an output. In this article, Q is the salient feature map, K is the matrix of grid coordinates, V is the E_{term} part of the energy function, and λ is a ratio. Thus, the energy function is modified as follows:

$$E(u, v) = \sum_p \min(W(p) \cdot \|s_1(p) - s_2(p + w(p))\|_1, t) + E_{\text{space}}. \quad (23)$$

In this way, edge integrity can be maintained during optimization. Fig. 7 illustrates the main steps of optical flow estimation using the attention mechanism. The saliency map is used as the query matrix, and the optimization method is dual-layer loop belief propagation based on a Markov field.

2) *Flow Fusion Estimation*: The optical flow as described by SIFT can appropriately capture relatively large displacements because of its domain description property, while the PC-based descriptor focuses more on single pixels. The domain description property of the former will tend to cause distortion, and the latter has a weak ability to capture relatively large displacements. Therefore, we fuse these complementary approaches to obtain better results [45].

The image is first divided into several regions. Then, a region in the optical flow field of the PC-based descriptor ($field_p$) is copied into the corresponding region in the optical flow field of SIFT ($field_s$), and $E(u, v)$ is calculated again. The change is

Algorithm 1: Fusion of the Optical Flow Fields of SIFT and the PC-Based Descriptor.

Input: the optical flow fields of SIFT and the PC-based descriptor

Output: the fused field

```

1: Generate a region set  $X$  related to  $field_p$ 
2: Initialize  $x^{cur}$  as a random region from  $X$ 
3: while traversing  $X$  do
4:  $field_{fusion} = x^{cur} \rightarrow field_s$ 
5: if  $E_{fusion} < E_{origin}$  then
6:  $field_s = field_{fusion}$ 
7: end if
8:  $x^{cur} = another\ region$ 
9: end while
10:  $field_{fusion} = field_s$ 
11:
12: return  $field_{fusion}$ 

```

TABLE I
DATASET DESCRIPTION

Image ID	1	2	3	4
Image type	Suburb	Airport	Port	Field
Image size	1200 × 1200	1200 × 1200	1200 × 1200	1000 × 1000
Image ID	5	6	7	8
Image type	Suburb	City with river	City	City
Image size	1000 × 1000	1200 × 1200	1200 × 1200	1400 × 1400
Image ID	9	10		
Image type	City	City with river		
Image size	1400 × 1400	1400 × 1400		

maintained if $E(u, v)$ decreases relative to the original value. This process is iterated until all regions have been traversed. Additional details are illustrated in the pseudocode given in Algorithm 1.

III. EXPERIMENTAL RESULTS

In this section, we present an experimental analysis of our work. The discussion is separated into four topics: the dataset used, the parameter settings applied, the evaluation metrics calculated, and the results obtained. The algorithms considered for comparison are the most advanced or most well-established methods in dense SAR–optical image registration from recent years: GeFolki, SIFT flow, and OS-flow.

A. Dataset Description

Ten groups of image data, including suburban, field, airport, port and city categories, were prepared to verify the algorithm performance, as shown in Table I. These data come from Shanghai. Each data group consists of a SAR–optical image pair. The SAR image data were obtained from GF-3, while the optical image data were taken from the Google Earth dataset. Fig. 8 shows the first and sixth image pairs.

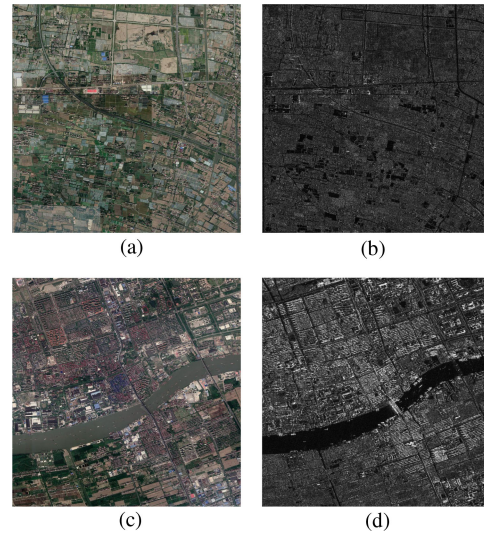


Fig. 8. Dataset description. (a) Optical image in the first pair. (b) SAR image in the first pair. (c) Optical image in the sixth pair. (d) SAR image in the sixth pair.

B. Parameter Settings

Many parameters are involved in our algorithm. For CRONE, the derivative order n was set to 1.5, and the template length $2m + 1$ was 7. For our filter based on fuzzy theory, u_1 was set to 3, and u_2 was set to 20. For the attention mechanism, λ was set to 0.1. For optical flow field fusion, the optical flow fields were each divided into 256 blocks, and the scale-space level in the bpf flow optimization algorithm was set to 2.

Because of the difficulty of dense SAR–optical image registration, the number of algorithms available for comparison is limited. SIFT flow is a classical algorithm, while OS-flow is an improved version of SIFT flow. We set the relevant parameters of these algorithms based on the default code. GeFolki can also achieve good results in dense SAR–optical image registration. We set the relevant parameters of this algorithm in accordance with the literature. During the experiment, the same images were used; so the variables could be strictly controlled.

In addition, the reference and real images need to be specified. Due to their different imaging mechanisms, optical images have smaller and more controllable local deformations, whereas the local deformations in SAR images are seriously affected by the terrain and other factors. Therefore, the optical image in each pair was used as the reference image, and the SAR image was used as the real image.

C. Evaluation Metrics

Due to the differences between SAR and optical images, the traditional evaluation metrics used for the optical flow method are not suitable. Moreover, because the real values of a remote sensing image cannot be obtained, it is impossible to obtain the real residual image. The global parameters can be described by means of scattering and mapping transformations, but expressing the local deformations using the same functional form is difficult. Therefore, based on previous SAR–optical image registration studies, multiple pairs of truth points were

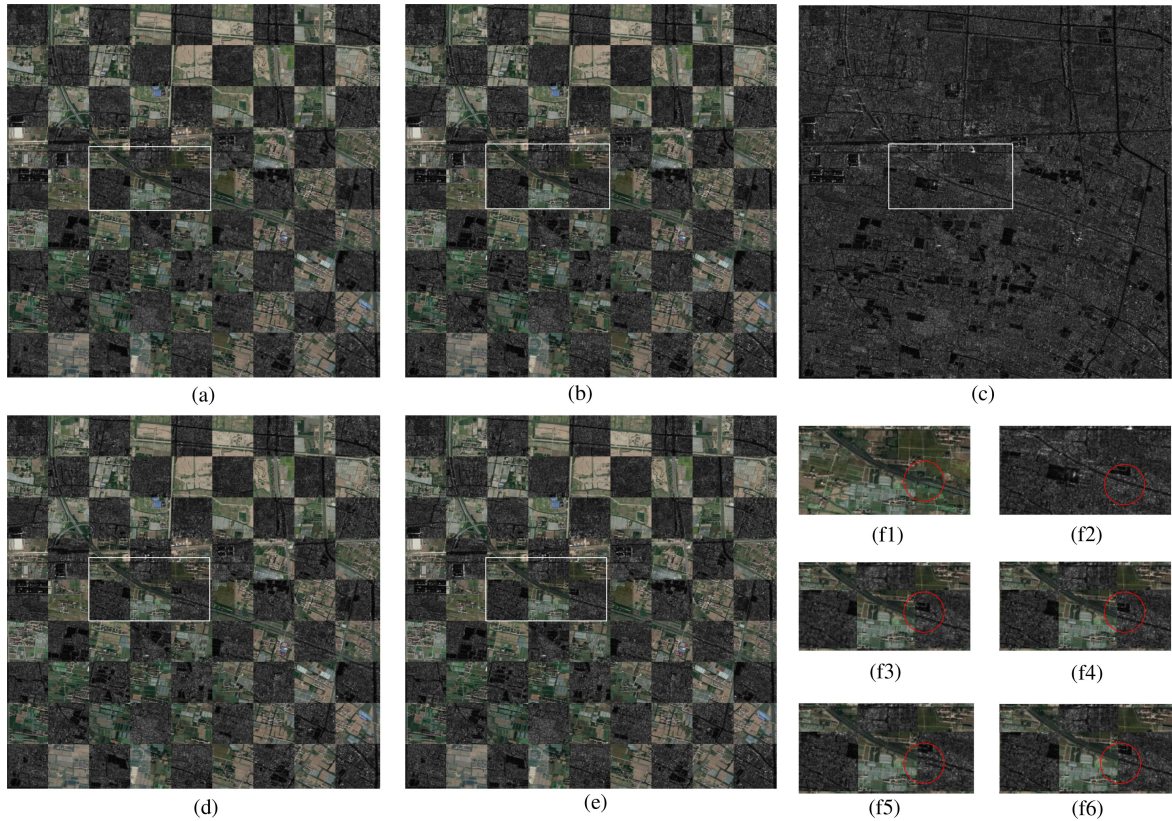


Fig. 9. Suburban image. (a) GeFolki. (b) OS-flow. (c) SAR image. (d) SIFT flow. (e) Our method. (f) White-framed region. (f1) Optical image. (f2) SAR image. (f3) GeFolki. (f4) OS-flow. (f5) SIFT flow. (f6) Our method.

simulated to evaluate the error, and the root mean square error (RMSE) and the error statistics of the matched inliers (ESMI) [46] were taken as the evaluation metrics.

The RMSE is usually expressed by the following formula:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]}. \quad (24)$$

The ESMI is the variance of the matching error of the matched inliers, which is defined as follows:

$$E[\Delta x_i] = \frac{1}{m} \sum_{i=1}^m (x_i - \hat{x}_i) \quad (25)$$

$$E[\Delta y_i] = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i) \quad (26)$$

$$D[\Delta x_i] = \frac{1}{m} \sum_{i=1}^m [(x_i - \hat{x}_i) - E[x_i - \hat{x}_i]]^2 \quad (27)$$

$$D[\Delta y_i] = \frac{1}{m} \sum_{i=1}^m [(y_i - \hat{y}_i) - E[y_i - \hat{y}_i]]^2. \quad (28)$$

D. Results

In this section, we present the experimental performance of our algorithm, which solves the problem introduced at the beginning of this article and reduces the influence of excessive twist to improve the registration quality. Six image pairs (IDs 1, 2, 3, 4, 6, and 7), which represent different types of topography, are selected for specific analysis. Finally, the experimental results are given. We evaluate the registration performance by comparing our proposed algorithm with three existing algorithms, namely, GeFolki, SIFT flow, and OS-flow, which are all state-of-the-art methods in the field of dense SAR–optical image registration. All experiments were conducted on a computer with an Intel Core i3-6100 CPU.

Fig. 9 shows the suburban image pair (image ID 1). Cities and villages are mixed in the depicted region. In addition to the main roads, there are many winding field trails, which can easily interfere with the registration process. Some houses on farmland cannot be observed in the SAR image, and the amount of detailed information in the optical image is obviously greater than that in the SAR image. The color differences between the SAR image and the optical image are also significant. Farmland and houses are dark in the SAR image, but they can be easily observed in the optical image. In panel (f) of Fig. 9, the road is well registered by GeFolki, OS-flow, and our algorithm, but there is a large twist in the SIFT flow result due to ambiguous pixel correspondences. In contrast, OS-flow uses GLOH descriptions instead of SIFT. Depending on the intensity ordering within a local

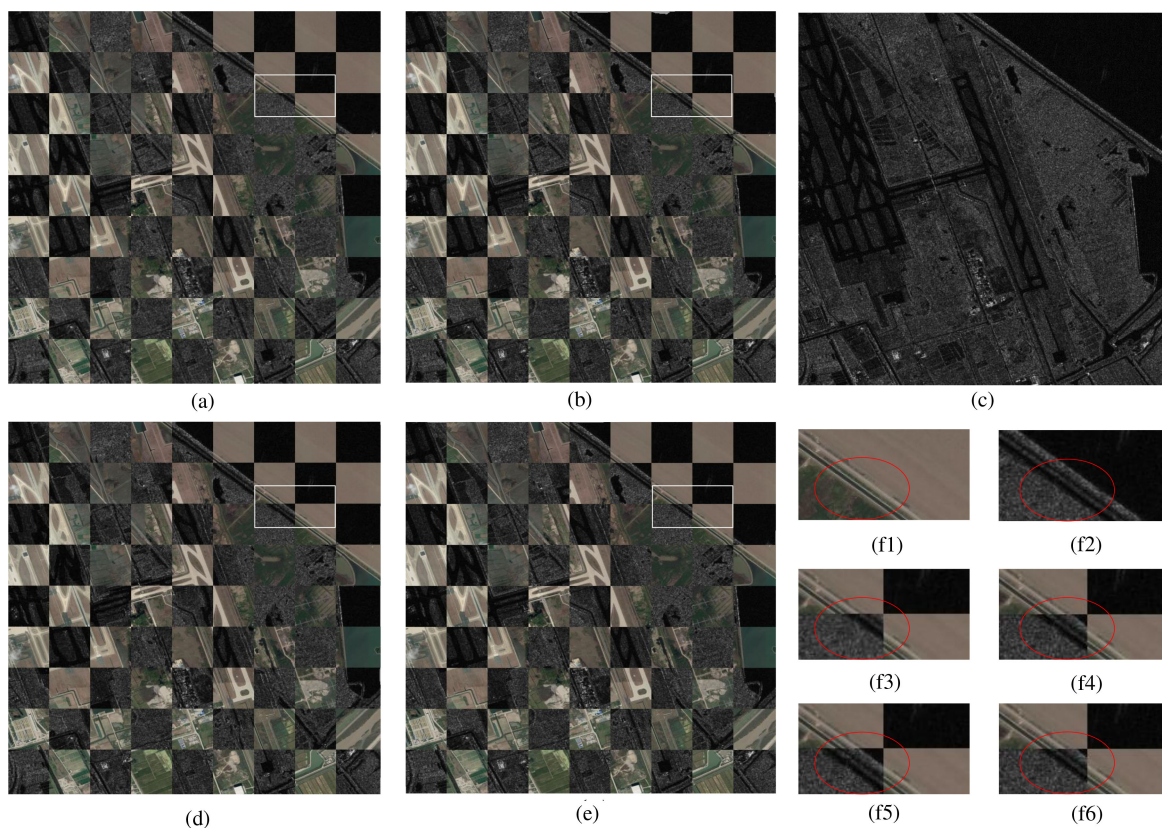


Fig. 10. Airport image. (a) GeFolki. (b) OS-flow. (c) SAR image. (d) SIFT flow. (e) Our method. (f) White-framed region. (f1) Optical image. (f2) SAR image. (f3) GeFolki. (f4) OS-flow. (f5) SIFT flow. (f6) Our method.

neighborhood, the rank filter in GeFolki can highly compress the signal dynamics, leading to performance enhancements. Meanwhile, our algorithm can easily extract the trunk road.

Fig. 10 shows the airport image pair (image ID 2). The sea surface lies in the upper right corner, which is blank in both the SAR and optical images. This part of the imaged region has no structural features. There is slight distortion inside the airport in the SAR image, which is caused by the SAR imaging mechanism. Our goal is to correct this local distortion, which is a key advantage of dense registration compared to sparse registration. In fact, all four algorithms generally achieve this goal well, but a problem appears in Fig. 10(f). The roads along the coastline are not successfully registered by GeFolki, OS-flow, or SIFT flow, although our algorithm performs well in this situation. This is a common weakness of optical flow algorithms in dense registration: the optical flow optimization function pays more attention to local information than global information. Consequently, the large blank area in the upper right corner disturbs the registration process and makes it difficult for edges near the blank area to be properly aligned. Our algorithm overcomes this difficulty through the introduction of the attention mechanism.

Fig. 11 shows the port image pair (image ID 3). Similar to the airport image pair, the sea occupies a large area in the upper right corner, which is blank in both the optical and SAR images. However, in this case, there are ships on the sea surface. Due to the different image acquisition times, the positions and styles of the ships that appear in the SAR and optical images

are different. Hence, these are obvious noise points for the registration process. In addition, there is considerable noise in the port container area, which appears as obvious blurriness. Such noisy areas usually lead to very poor results because the optical flow energy function treats all points equally. If the antinoise ability is not good, such areas will twist the result. As seen in Fig. 11(f), our attention-based optical flow fusion method performs well, and the joint filter combination in GeFolki also produces acceptable results. In contrast, the results of OS-flow and SIFT flow are excessively twisted.

Fig. 12 shows the field image pair (image ID 4). The edges of the farmland and the farmhouses are seriously obscured in the SAR image, and the outlines cannot be fundamentally distinguished. Only the roads between patches of farmland are clear. However, in the optical image, the details of these areas are easy to distinguish. Similar to Fig. 10, the edges near the blank areas are difficult to align. However, our algorithm performs well because of the introduction of the attention mechanism, while the effects of SIFT flow, OS-flow, and GeFolki are not satisfactory.

Fig. 13 shows the image pair depicting a city region with a river (image ID 6), and Fig. 14 shows the city image pair without a river (image ID 7). City areas contain both considerable structural information and many noise points. Consequently, it is easy to achieve good results at the global scale, but the results at the local scale can easily be twisted. Moreover, the color intensity trends are usually opposite between SAR and optical

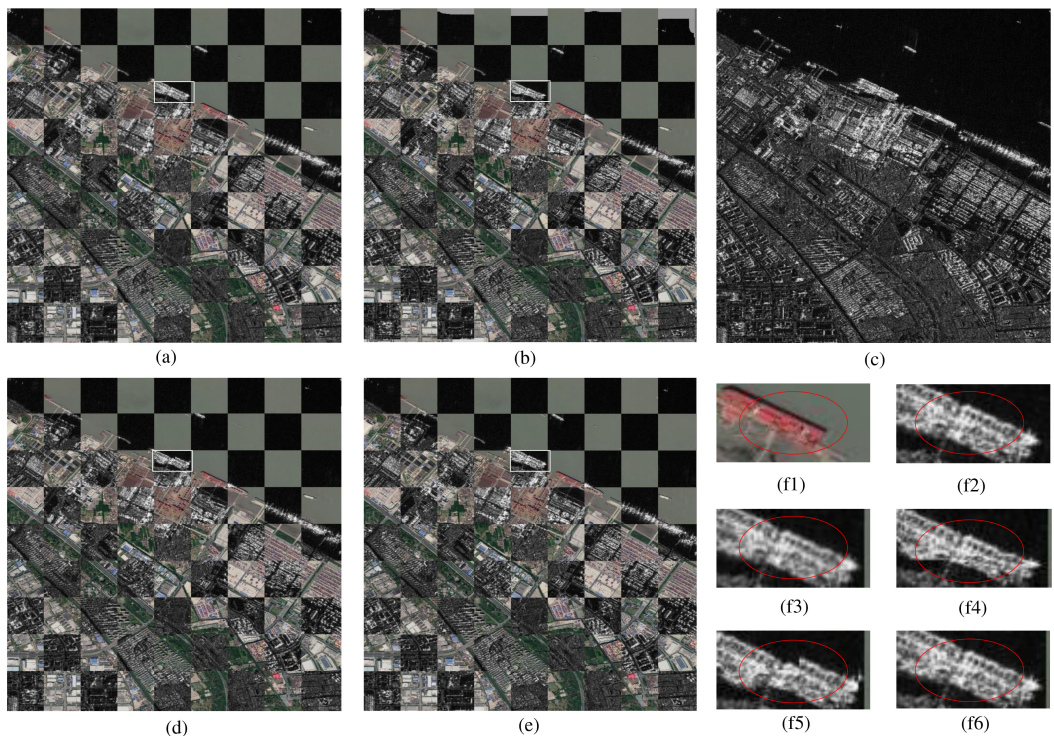


Fig. 11. Port image. (a) GeFolki. (b) OS-flow. (c) SAR image. (d) SIFT flow. (e) Our method. (f) White-framed region. (f1) Optical image. (f2) SAR image. (f3) GeFolki. (f4) OS-flow. (f5) SIFT flow. (f6) Our method.

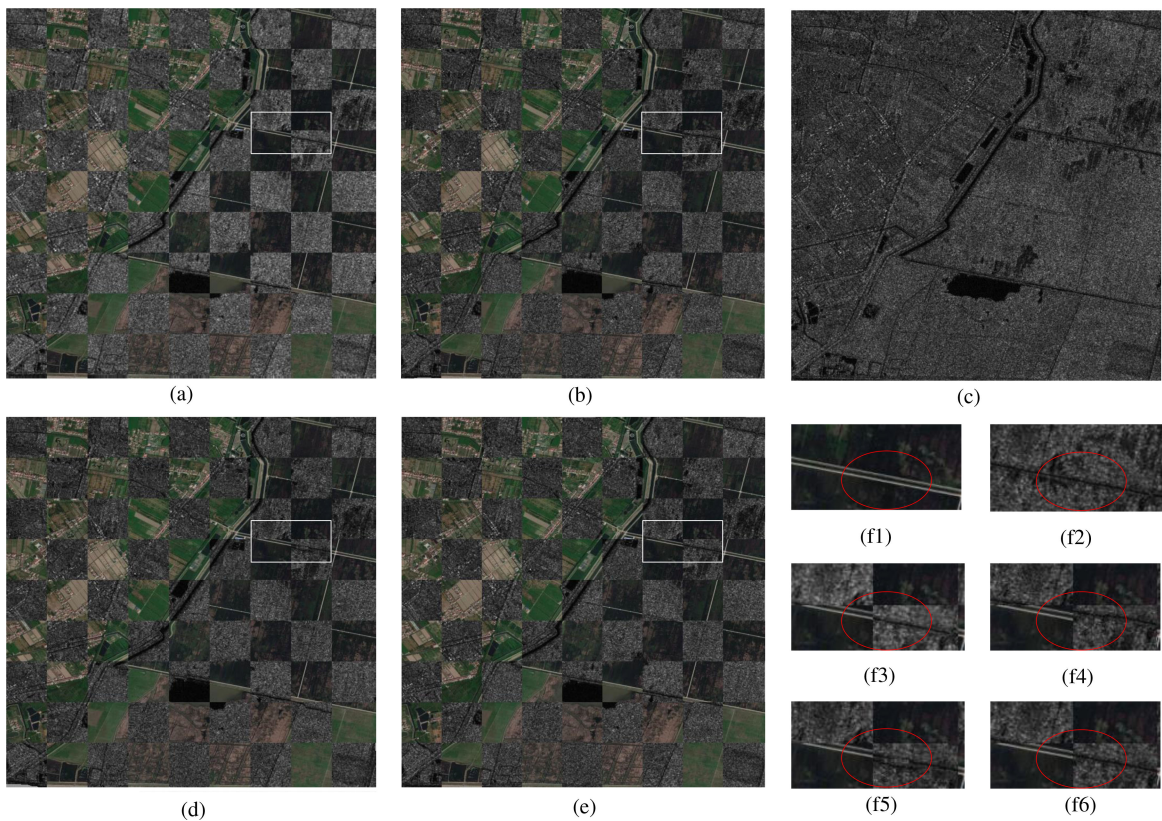


Fig. 12. Field image. (a) GeFolki. (b) OS-flow. (c) SAR image. (d) SIFT flow. (e) Our method. (f) White-framed region. (f1) Optical image. (f2) SAR image. (f3) GeFolki. (f4) OS-flow. (f5) SIFT flow. (f6) Our method.

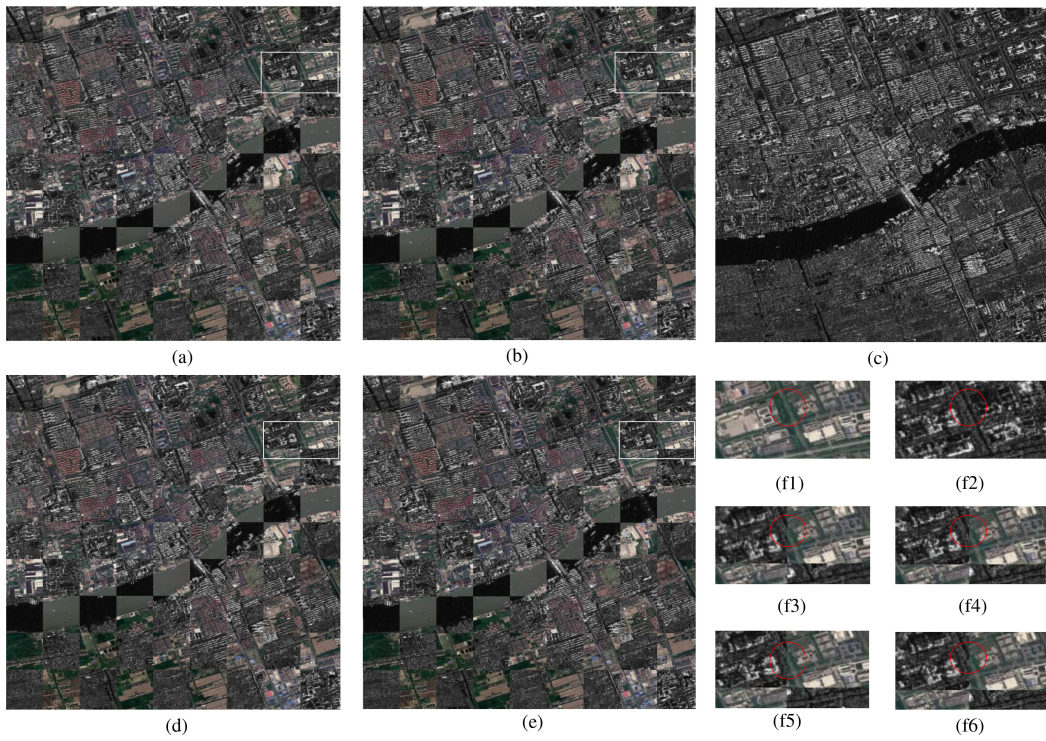


Fig. 13. City image with river. (a) GeFolki. (b) OS-flow. (c) SAR image. (d) SIFT flow. (e) Our method. (f) White-framed region. (f1) Optical image. (f2) SAR image. (f3) GeFolki. (f4) OS-flow. (f5) SIFT flow. (f6) Our method.

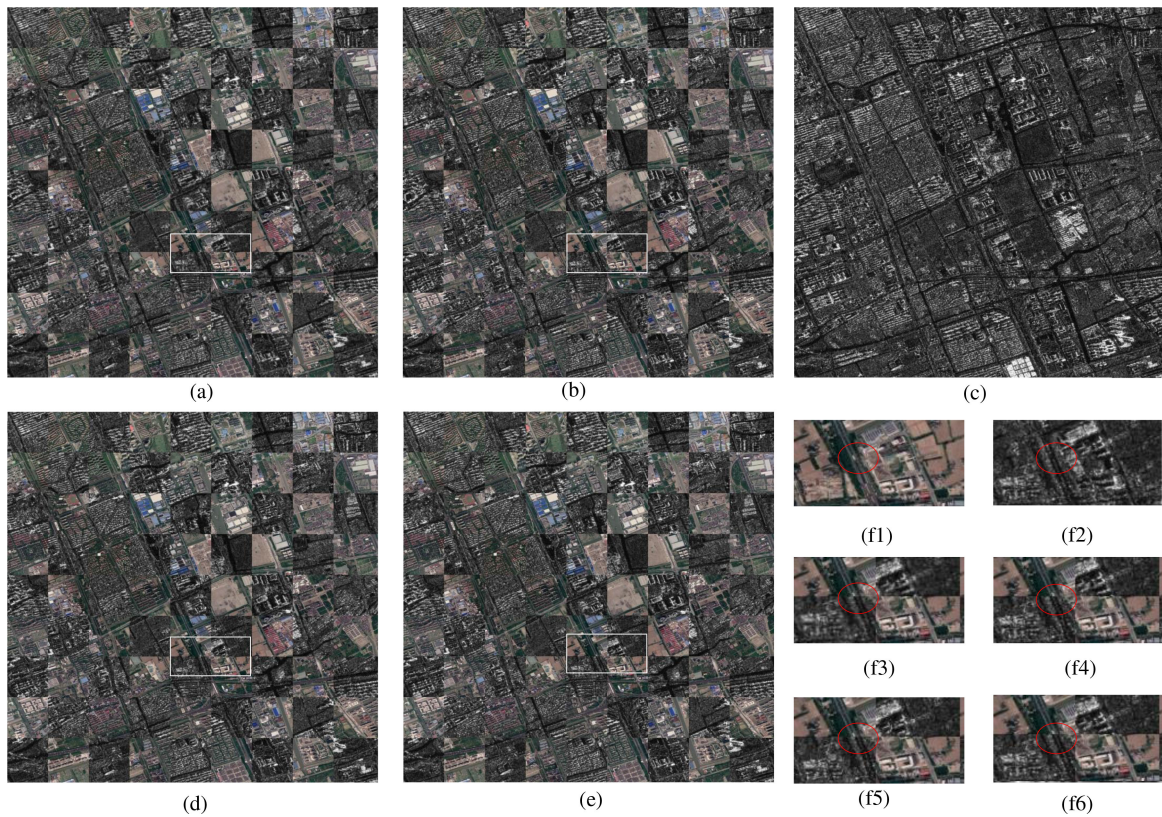


Fig. 14. City image. (a) GeFolki. (b) OS-flow. (c) SAR image. (d) SIFT flow. (e) Our method. (f) White-framed region. (f1) Optical image. (f2) SAR image. (f3) GeFolki. (f4) OS-flow. (f5) SIFT flow. (f6) Our method.

TABLE II
QUANTITATIVE COMPARISONS OF DIFFERENT METHODS ON TEN IMAGE PAIRS

Image ID	Description	Metric	GeFolki	SIFT flow	OS-flow	Our method
1	Suburb	RMSE	4.3286	5.3127	1.7022	0.8634
		ESMI (Varx/Vary)	1.6439/1.0933	2.2981/1.7364	1.5476/1.3630	0.3340/0.4001
		Time	16.35 s	66.78 s	55.52 s	154.74 s
2	Airport	RMSE	7.4972	6.9308	2.2476	0.8374
		ESMI (Varx/Vary)	1.9204/0.3548	5.6913/36.807	2.6213/1.8772	0.4300/0.1566
		Time	15.30 s	64.88 s	55.64 s	148.27 s
3	Port	RMSE	4.1616	6.5465	1.6557	0.9830
		ESMI (Varx/Vary)	1.5710/1.5280	5.5028/9.5474	1.4610/0.8225	0.8719/0.1292
		Time	15.15 s	64.12 s	55.48 s	147.80 s
4	Field	RMSE	2.1119	3.175	2.6246	0.8720
		ESMI (Varx/Vary)	0.8319/3.2196	5.0016/4.5469	2.6619/3.8945	0.3203/0.2993
		Time	10.18 s	44.29 s	38.34 s	98.70 s
5	Suburb	RMSE	2.573	3.5426	1.6609	0.9436
		ESMI (Varx/Vary)	0.2468/0.2192	2.1806/1.7345	0.5839/0.4998	0.1336/0.1369
		Time	10.45 s	43.83 s	38.23 s	101.29 s
6	City with river	RMSE	4.5054	7.9613	1.7718	0.9764
		ESMI (Varx/Vary)	2.3681/0.9471	5.7507/0.4524	1.2776/1.8870	0.2773/0.3917
		Time	17.14 s	64.24 s	55.45 s	148.14 s
7	City	RMSE	1.8489	2.4428	2.2121	1.2366
		ESMI (Varx/Vary)	0.9918/0.8337	2.8512/0.8892	2.2057/1.5130	0.4028/0.0919
		Time	15.35 s	63.93 s	55.68 s	149.59 s
8	City	RMSE	3.5801	4.164	2.5487	1.0591
		ESMI (Varx/Vary)	1.9767/1.3842	3.2461/1.1791	3.8101/1.8597	0.7085/0.3702
		Time	24.06 s	87.64 s	76.35 s	210.79 s
9	City	RMSE	4.4503	7.4003	1.9036	0.9511
		ESMI (Varx/Vary)	0.7208/0.3030	1.6406/0.7029	1.5414/0.6726	0.1883/0.3176
		Time	21.41 s	88.49 s	76.81 s	211.84 s
10	City with river	RMSE	5.0401	7.6246	1.716	1.0803
		ESMI (Varx/Vary)	1.0431/0.3666	3.6434/2.1812	2.1167/0.6588	0.1685/0.2643
		Time	21.50 s	88.57 s	76.30 s	207.85 s

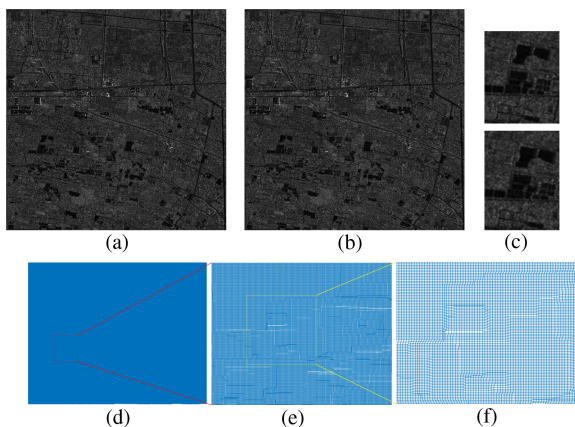


Fig. 15. Schematic diagram of the deformation field. (a) Image before deformation. (b) Image after deformation. (c) Local amplification. (d) Deformation field. (e) Primary amplification of the deformation field. (f) Secondary amplification of the deformation field.

images of such regions. For example, in Figs. 13(f) and 14(f), the roads show relatively low intensities in the SAR images but high reflectivity in the optical images. Although GeFolki and OS-flow each have mechanisms to address such intensity differences, they do not achieve a very good effect in these examples. In contrast, our algorithm extracts these roads by means of its attention mechanism and achieves good results.

Table II summarizes the experimental results for all image data. Our proposed algorithm performs better than SIFT flow, OS-flow, and GeFolki, with significant improvements in both the RMSE and ESMI results, thus showing that our algorithm is feasible and effective. Due to the introduction of the attention mechanism, several problems related to roads and edges are solved. By combining the optical flow fields of different

descriptors, the domain description property of SIFT is compensated. These changes allow our algorithm to achieve advantages in accuracy. However, these changes also increase the computational cost. Fig. 16 presents a visual representation of Table II. Fig. 15 shows a schematic diagram of the obtained deformation field.

To demonstrate the robustness of the proposed method against speckle noise, we used a multiplicative noise model to generate simulated SAR images. We added speckle noise with different variances to the same SAR image, and the results obtained under different noise levels are shown in Fig. 16(j) and (k). Because the values of the SIFT flow and GeFolki algorithms are too large, we show only the comparison between OS-flow and our algorithm. All experiments indicate that the proposed method performs much better than the other methods.

IV. DISCUSSION

A. Salient Feature Map

The salient feature map plays an important role in our algorithm as the Q matrix for introducing preprocessing information into the optical flow optimization function. In this article, this information is represented in the form of edge structural features. Other types of image features include color features and textural features. Because of the differences between SAR and optical images, color features are not considered. Between textural and edge structural features, the latter are more convenient to use in subsequent experiments. Nevertheless, we believe that investigating how to make use of textural features will be an interesting topic for future work. In addition, some recent methods related to speckle noise, such as nonlocal filters [47], sparse representation [48], and nonlinear partial differential equations [49], will be considered.

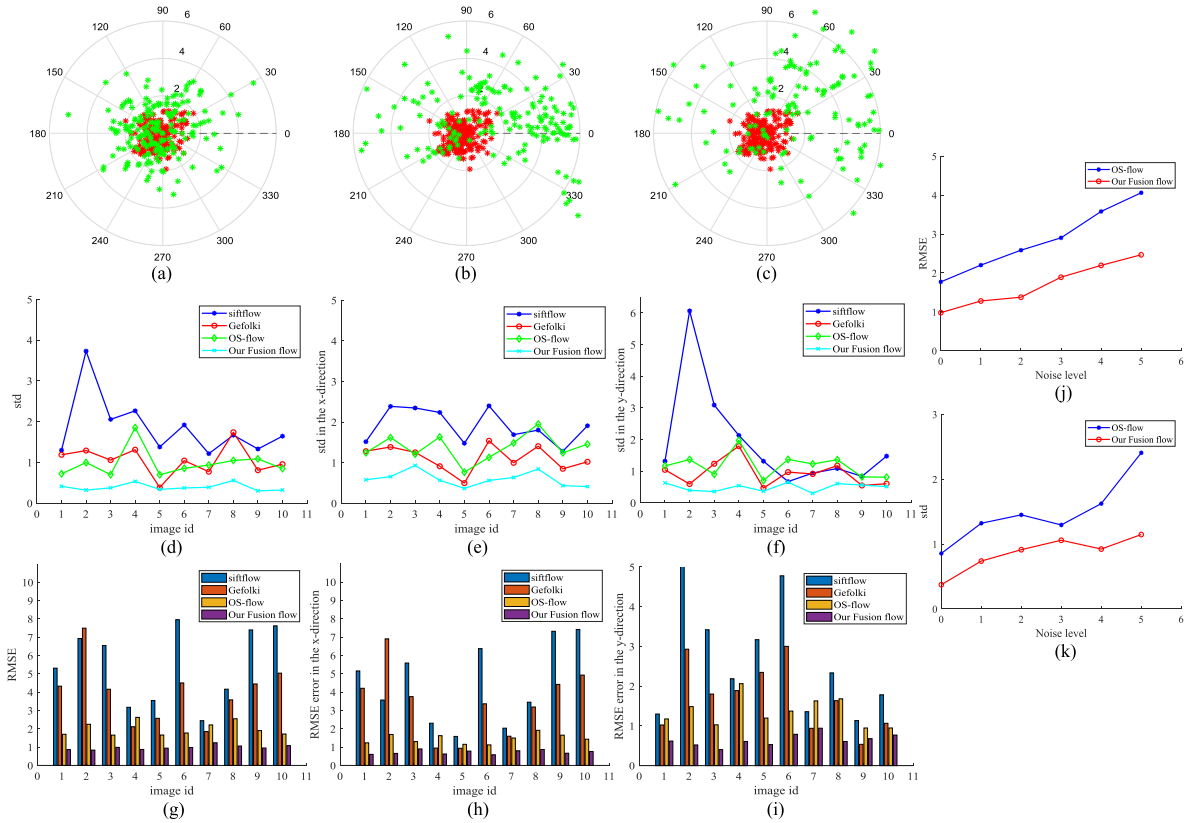


Fig. 16. (a–i) Quantitative comparisons of different methods on ten image pairs. (a–c) Comparison of coordinate errors (Δx , Δy) in the Cartesian coordinate system between our method (red) and OS-flow (green) (a), our method (red) and GeFolki (green) (b), and our method (red) and SIFT flow (green) (c). (d) Standard deviation of the matching error. (e) Standard deviation of the matching error in the x -direction. (f) Standard deviation of the matching error in the y -direction. (g) RMSE. (h) Mean of the matching error in the x -direction. (i) Mean of the matching error in the y -direction. (j–k) Quantitative comparisons at different noise levels. (j) RMSE. (k) Standard deviation.

B. Attention Mechanism

The core question addressed in this article is how to make use of an attention mechanism in dense registration tasks. We apply our attention mechanism to identify and highlight salient feature regions, which contain rich information and significantly contribute to the registration task. On the other hand, noisy areas often make a negative contribution. Accordingly, how to apply an attention mechanism to noisy areas to reduce their influence is expected to be a question worth studying.

C. Fusion Method

The purpose of fusion is to combine the advantages of two methods while neutralizing their disadvantages to enhance the robustness of the results. In this article, each image is divided into square blocks, and the final optical flow field is improved through direct block-by-block replacement of the SIFT flow field with the PC-based flow field. Although fusion of the two flow fields can be achieved in this way, this approach also has obvious disadvantages. First, the use of square blocks will result in stepwise changes at the block edges, which may lead to discontinuities and increase the error. Second, the direct replacement method is rather crude.

To address the above shortcomings, two possible improvements could be to adopt a Gaussian kernel instead of division

into square regions and to perform weighted fusion instead of direct replacement. However, the feasibility and effectiveness of these modifications require further investigation.

D. Sparse Registration and Dense Registration

The approaches to image registration can be divided into sparse registration [26], [50]–[52] and dense registration depending on whether all pixels are used. Sparse registration produces only a set of global parameters as output, which cannot correct local distortions but offers real-time performance. On the other hand, dense registration aims to solve the problem of local distortions and thus produces output in the form of the optical flow field, but its real-time performance is poor.

In this section, our algorithm is compared with the HOPC descriptor [26] and the rotation-invariant feature transform (RIFT) [53], which are state-of-the-art algorithms in the field of multisource sparse registration. The specific results are shown in Table III. Although our algorithm yields improved RMSE results, it takes significantly longer than the others as a consequence of the algorithm design itself.

Our algorithm is fundamentally different from HOPC and RIFT. They perform sparse registration, whereas our algorithm performs dense registration. Both approaches have their own

TABLE III
QUANTITATIVE COMPARISONS WITH SPARSE REGISTRATION METHODS

Image ID	HOPC		RIFT		Our method	
	Time	RMSE	Time	RMSE	Time	RMSE
1	31.71 s	1.0855	22.82 s	1.2689	154.74 s	0.8634
2	33.10 s	1.2315	20.92 s	1.0321	148.27 s	0.8374
3	30.72 s	1.1722	21.40 s	1.7075	147.80 s	0.9830
4	28.75 s	1.7858	17.76 s	3.9195	98.70 s	0.8720
5	27.48 s	1.0786	17.74 s	2.0814	101.29 s	0.9436
6	30.66 s	2.1226	21.16 s	2.0092	148.14 s	0.9764
7	30.87 s	1.2651	21.04 s	1.5952	149.59 s	1.2366
8	35.57 s	1.1846	24.60 s	1.5441	210.79 s	1.0591
9	34.74 s	1.2787	24.10 s	1.2063	211.84 s	0.9511
10	34.78 s	1.1710	24.81 s	1.1533	207.85 s	1.0803

advantages and disadvantages, and which approach should be chosen depends on the specific task.

E. Critical Discussion

Compared with the previous dense registration algorithms, our algorithm has an advantage in accuracy, but its computational cost is relatively high. The attention mechanism enhances the importance of salient features and causes them to have a greater impact on the energy function, thereby ensuring that salient edges can be maintained in the calculation of the optical flow optimization function to avoid excessively twisted results. This is also the main reason why our method is superior to previous methods. In addition, a new descriptor based on PC is proposed. By combining the results of this descriptor with those of SIFT, the error caused by the domain description property of SIFT is compensated, also contributing to the enhanced accuracy. However, these additional steps not only improve the accuracy but also increase the computational cost, making our algorithm markedly slower than previous algorithms. In our future work, we will focus on reducing the computational cost.

Moreover, our method relies too much on the saliency map. If the edge structural information is not extracted well, then the performance will be affected; in particular, the gradient reversal phenomenon may occur. This is a limitation of our algorithm. In the future, we will investigate the adoption of advanced descriptors to address this problem.

V. CONCLUSION

In this article, we have proposed a novel method for dense SAR–optical image registration via optical flow fusion estimation based on an attention mechanism. The main idea of the proposed method is to introduce the use of an attention mechanism into dense registration, which distinguishes it from previous methods. The architecture of the method is composed of two branches: one aims to extract a salient feature map to be used as the Q matrix to introduce preprocessing information into the optical flow optimization function, and the other is optical flow fusion estimation. The input consists of a SAR–optical image pair, and the output is an optical flow field. The proposed method outperforms previous methods by overcoming edge distortions. The method has been tested on image data representing ten different scenarios. The results illustrate that the proposed

method exhibits good performance in dense SAR–optical image registration.

REFERENCES

- [1] H. Nagendra *et al.*, “Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats,” *Ecol. Indicators*, vol. 33, pp. 45–59, 2013.
- [2] Q. Li *et al.*, “Coupled GAN with relativistic discriminators for infrared and visible images fusion,” *IEEE Sensors J.*, vol. 21, no. 6, pp. 7458–7467, Mar. 2021.
- [3] J. Xu, L. Li, and M. Ji, “Ensemble learning based multi-source information fusion,” in *Proc. Int. Conf. Image Video Process. Artif. Intell.*, 2019, Art. no. 1132123.
- [4] B. Zitova and J. Flusser, “Image registration methods: A survey,” *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.
- [5] X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, “A review of multimodal image matching: Methods and applications,” *Inf. Fusion*, vol. 73, pp. 22–71, 2021.
- [6] L. G. Brown, “A survey of image registration techniques,” *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.
- [7] F. Sharifzadeh, G. Akbarzadeh, and Y. Seifi Kavian, “Ship classification in SAR images using a new hybrid CNN-MLP classifier,” *J. Indian Soc. Remote Sens.*, vol. 47, no. 4, pp. 551–562, 2019.
- [8] N. Davari, G. Akbarzadeh, and E. Mashhour, “Intelligent diagnosis of incipient fault in power distribution lines based on corona detection in UV-visible videos,” *IEEE Trans. Power Del.*, vol. 36, no. 6, pp. 3640–3648, Dec. 2021.
- [9] W. Ma, J. Zhang, Y. Wu, L. Jiao, H. Zhu, and W. Zhao, “A novel two-step registration method for remote sensing images based on deep and local features,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4834–4843, Jul. 2019.
- [10] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14.
- [11] D. Quan *et al.*, “Deep generative matching network for optical and SAR image registration,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 6215–6218.
- [12] N. Merkle, S. Auer, R. Müller, and P. Reinartz, “Exploring the potential of conditional adversarial networks for optical and SAR image matching,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 6, pp. 1811–1820, Jun. 2018.
- [13] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, “SAR-SIFT: A SIFT-like algorithm for SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2015.
- [14] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [15] J. Fan, Y. Wu, F. Wang, Q. Zhang, G. Liao, and M. Li, “SAR image registration using phase congruency and nonlinear diffusion-based SIFT,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 562–566, Mar. 2015.
- [16] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [17] P. Kovesi *et al.*, “Image features from phase congruency,” *Videre: J. Comput. Vis. Res.*, vol. 1, no. 3, pp. 1–26, 1999.
- [18] Z. Hongsheng, L. Hui, and Z. Yuanzhi, “A practical semi-automatic approach for optical and SAR data co-registration,” in *Proc. 3rd Int. Workshop Earth Observ. Remote Sens. Appl.*, 2014, pp. 1–4.
- [19] H. Wang *et al.*, “Optical-to-SAR image registration based on Gaussian mixture model,” *ISPRS-Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 39, pp. 179–183, 2012.
- [20] Y. Wang and Z. Chen, “Automatic registration of SAR and optical images based on linear features and neural network,” in *Proc. 3rd Int. Symp. Multispectral Image Process. Pattern Recognit.*, 2003, pp. 553–557.
- [21] Q. Yu, D. Ni, Y. Jiang, Y. Yan, J. An, and T. Sun, “Universal SAR and optical image registration via a novel SIFT framework based on nonlinear diffusion and a polar spatial-frequency descriptor,” *ISPRS J. Photogramm. Remote Sens.*, vol. 171, pp. 1–17, 2021.
- [22] S. Hoffmann, C.-A. Brust, M. Shadaydeh, and J. Denzler, “Registration of high resolution SAR and optical satellite imagery using fully convolutional networks,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5152–5155.
- [23] H. Li, B. S. Manjunath, and S. K. Mitra, “A contour-based approach to multisensor image registration,” *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 320–334, Mar. 1995.

[24] S. Suri and P. Reinartz, "Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 939–949, Feb. 2010.

[25] L. Huang, Z. Li, and R. Zhang, "SAR and optical images registration using shape context," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 1007–1010.

[26] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2941–2958, May 2017.

[27] S. Chen, S. Zhong, B. Xue, X. Li, L. Zhao, and C.-I. Chang, "Iterative scale-invariant feature transform for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3244–3265, Apr. 2021.

[28] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, 1994.

[29] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, 1981, pp. 674–679.

[30] G. Brigot, E. Colin-Koeniguer, A. Plyer, and F. Janez, "Adaptation and evaluation of an optical flow method applied to coregistration of forest remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 2923–2939, Jul. 2016.

[31] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.

[32] Y. Xiang, F. Wang, L. Wan, N. Jiao, and H. You, "OS-flow: A robust algorithm for dense optical and SAR image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6335–6354, Sep. 2019.

[33] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1–3, pp. 185–203, 1981.

[34] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[35] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[36] B. Mathieu, P. Melchior, A. Oustaloup, and C. Ceyral, "Fractional differentiation for edge detection," *Signal Process.*, vol. 83, no. 11, pp. 2421–2432, 2003.

[37] R. Fjortoft, A. Lopes, P. Marthon, and E. Cubero-Castan, "An optimal multiedge detector for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 793–802, May 1998.

[38] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, 2016.

[39] A. Gruen, "Development and status of image matching in photogrammetry," *Photogrammetric Rec.*, vol. 27, no. 137, pp. 36–57, 2012.

[40] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.

[41] T. Brox, A. Bruhn, N. Papenber, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 25–36.

[42] Y. Mileva, A. Bruhn, and J. Weickert, "Illumination-robust variational optical flow with photometric invariants," in *Proc. Joint Pattern Recognit. Symp.*, 2007, pp. 152–162.

[43] H. Zimmer *et al.*, "Complementary optic flow," in *Proc. Int. Workshop Energy Minimization Methods Comput. Vis. Pattern Recognit.*, 2009, pp. 207–220.

[44] N. Papenber, A. Bruhn, T. Brox, S. Didas, and J. Weickert, "Highly accurate optic flow computation with theoretically justified warping," *Int. J. Comput. Vis.*, vol. 67, no. 2, pp. 141–158, 2006.

[45] S. Roth, V. Lempitsky, and C. Rother, "Discrete-continuous optimization for optical flow estimation," in *Statistical and Geometrical Approaches to Visual Motion Analysis*, Berlin, Germany: Springer, 2009, pp. 1–22.

[46] Q. Yu, S. Zhou, Y. Jiang, P. Wu, and Y. Xu, "High-performance SAR image matching using improved SIFT framework based on rolling guidance filter and ROEWA-powered feature," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 920–933, Mar. 2019.

[47] E. Ferrentino, A. Buono, F. Nunziata, A. Marino, and M. Migliaccio, "On the use of multipolarization satellite SAR data for coastline extraction in harsh coastal environments: The case of solway firth," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 249–257, 2021.

[48] J. Fan, Y. Wu, F. Wang, P. Zhang, and M. Li, "New point matching algorithm using sparse representation of image patch feature for SAR image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1498–1510, Mar. 2017.

[49] M. Pourfard *et al.*, "KAZE-SAR: SAR image registration using KAZE detector and modified surf descriptor for tackling speckle noise," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5207612.

[50] Y. Ye and J. Shan, "A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences," *ISPRS J. Photogramm. Remote Sens.*, vol. 90, pp. 83–95, 2014.

[51] Y. Ye, L. Bruzzone, J. Shan, F. Bovolo, and Q. Zhu, "Fast and robust matching for multimodal remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9059–9070, Nov. 2019.

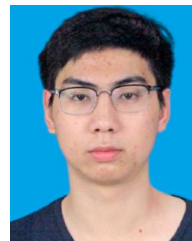
[52] Y. Ye, L. Shen, M. Hao, J. Wang, and Z. Xu, "Robust optical-to-SAR image matching based on shape properties," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 564–568, Apr. 2017.

[53] J. Li, Q. Hu, and M. Ai, "RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform," *IEEE Trans. Image Process.*, vol. 29, pp. 3296–3310, 2020.



Qiuzhe Yu received the Ph.D. degree in electronic engineering from the Institute of Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, China, in 2004.

From 2008 to 2012, he served as a Postdoctoral Fellow and an Associate Professor with Shanghai Jiao Tong University, Shanghai, China. In 2013, he worked as a Visiting Scholar with the California Institute of Technology (Caltech), Pasadena, CA, USA. He is currently a Principle Investigator with the Laboratory for Multisensor Signal Processing and UAV Navigation, Wuhan University, Wuhan, China. He has authored more than 40 papers in international conference proceedings and journals. His research interests include multisensor (radar-optical) signal and image processing, multisensor automatic target recognition, deep learning, and unmanned aerial vehicle navigation.



Yuxuan Jiang received the bachelor's degree in communication engineering in 2020 from Wuhan University, Wuhan, China, where he is currently working toward the master's degree with the School of Electronic Information.

His research interests include SAR image registration and navigation.



Wensen Zhao received the bachelor's degree in communication engineering in 2021 from Wuhan University, Wuhan, China, where he is currently working toward the master's degree with the School of Electronic Information.

His research interests include SAR image registration and navigation.



Tao Sun received the Ph.D. degree in computer science from Wuhan University, China, in 2003.

He did his postdoctorate work at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, from 2004 to 2006, where he is currently an Associate Professor with the Electronic Information School. He majored in photogrammetry and remote sensing. His research interests include remote sensing image processing, environmental modeling, and computer vision.