Efficient Multisource Remote Sensing Image Matching Using Dominant Orientation of Gradient

Dongxing Liang, Jinshan Ding, *Member, IEEE*, and Yuhong Zhang, *Senior Member, IEEE*

***Abstract*—Image matching is the key step for image registration. Due to the existing nonlinear intensity differences between multisource images, their matching is still a challenging task. A fast matching approach based on dominant orientation of gradient (DOG) is proposed in this article, which is robust to nonlinear intensity variations. The DOG feature maps are constructed by extracting DOG feature of each pixel in the images in the first place. A template matching method is used to determine correspondences between images based on the feature representations. We define a similarity measurement, referred to as sum of cosine differences, which can be accelerated by fast Fourier transform. Subsequently, the subpixel accuracy can be achieved by fitting the similarity measurement using a quadratic polynomial modal. A new variable template matching (VTM) method has been developed to improve the matching performance. Experimental results confirm that the proposed matching approach is robust to nonlinear intensity differences and has time efficiency. The VTM method additionally improves the matching precision effectively.**

***Index Terms*—Dominant orientation of gradient (DOG), image matching, variable template matching (VTM).**

I. Introduction

With the development of geospatial information technology, quite a few types of remote sensing (RS) images become very accessible, such as visible image, infrared image, LiDAR image, and synthetic aperture radar (SAR) image. These multisource RS images provide complementary feature information for the observation scene, which can be utilized in many RS applications, including image fusion [1] and change detection [2]. Multisource image registration is the prerequisite of these applications. Multisource image registration aims to strictly align two or more images obtained by different sensors or at different viewing angles [3]. Although multisource RS image registration has been intensively studied for decades, there exists no high-precision automatic registration method that can be generally applied, especially for multisource RS image registration.

Manuscript received October 9, 2020; revised January 2, 2021; accepted January 14, 2021. Date of publication January 18, 2021; date of current version February 8, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFE0200400. *(Corresponding author: Jinshan Ding.)*

Dongxing Liang and Jinshan Ding are with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: ldxwwq@163.com; ding@xidian.edu.cn).

Yuhong Zhang is with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: yuhzhang@xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3052472

Fortunately, most modern RS images are usually attached with georeferencing information that can be used to preregister the reference and sensed images. The preregistration eliminates almost all global distortions including obvious rotation and scale differences and there are only several to dozens of pixel deviations between the preregistration images, which provides great convenience for further fine registration [4]–[6].

The crucial step of image registration is image matching which extracts and matches the correspondences or the control points (CPs) between reference and sensed images. The correspondences represent the distinctive and representative points of the observed scenes. However, the existing nonlinear intensity differences usually make multisource image matching in trouble. Hence, this article focuses on multisource RS image matching to cope with nonlinear intensity changes.

Generally, most image matching methods for multisource images can be divided into feature-based and area-based types [3]. For feature-based methods, the candidate local features with salient structure information are first extracted from both reference image and sensed image. The correspondences between images are determined based on similarity measures of their feature descriptors. The extracted features can be points [7], edges [8], and contours [9]. Recently, matching methods based on local invariant features, such as scale invariant feature transform (SIFT) [10], speeded up robust feature [11] and KAZE [12], have been widely utilized in the filed of RS image matching [13]–[15]. Among them, SIFT is recognized as the most robust method. However, the experimental results show that SIFT performs better in single-modal images, but not well in multisource images [16]. This is because it is vulnerable to nonlinear intensity differences. Many variants of SIFT are proposed to matching multisource RS images [17], [18]. They perform well in some cases, but not in others. A mixture model with multiple features combination is presented to improve the multiviewpoint RS image matching accuracy [19]. In general, the repeatability of feature detection is usually the main factor affecting the performance of these feature-based matching methods. Significant intensity, texture and potential resolution differences usually result in lower repeatability and lower matching performance [20]–[22].

Conversely, area-based methods, also known as template matching methods, realize the correspondences of CPs by evaluating the similarity of the corresponding window pairs or even the entire images, which have advantages in precision, distribution, stability, and other aspects [23]. The focus of area-based

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

methods is the similarity measurement that can be computed in either spatial or frequency domain.

There are some commonly used similarity measures operated in spatial domain [24], [25]. These similarity measures are calculated with the template patch sliding on the search region in the sensed image, which is usually time-consuming. Frequency-based image correlation is also a type of template matching technique, which can avoid the iterative search process by Fourier transform (FT) routine. Phase correlation (PC) is a well known frequency-based matching technique [26]. However, these conventional area-based matching methods usually utilize intensity information to matching images, which degrades the performance of multisource image matching due to nonlinear intensity differences and noise.

The intensity distortion between multisource images is very complicated, which can not be fitted with simple mapping rules, such as linearity, monotonicity, etc. Hence, some matching methods have poor matching performance because their similarity measures are based on intensity mapping rules [27]. The intensity inversion is a special case of intensity changes, which results in orientation reversal. The issue frequently degrades the performance of template matching methods based on gradient orientation [28]. Meanwhile, noise is usually inevitable in RS images, especially in LiDAR and SAR images. The gradient magnitude is more easily affected by noise, which may change the gradient orientation. This can cause issue to the similarity measure for these methods based on gradient information with magnitude and orientation of each individual pixel.

We propose a computational efficient and robust matching method. First, the dominant orientation of gradient (DOG) features are extracted as feature representations to reduce the influence of nonlinear intensity differences. Subsequently, the similarity measure, i.e., sum of cosine difference (SCD), is implemented to cope with orientation distortion caused by intensity changes and noise, which can be accelerated using FFT. And then, CPs with subpixel accuracy are determined by fitting the similarity measurement using a quadratic polynomial modal. Furthermore, a novel variable template matching (VTM) method is proposed to improve the matching performance. The VTM method is a general method. For template matching methods based on spatial similarity measurements that can be accelerated by FFT, the VTM method can be used in the measurement process of these methods and improve their matching performance without increasing computational complexity.

The rest of this article is organized as follows. Section II briefly reviews the related work for completeness. Section III details the framework for multisource RS image matching. Experimental results are presented in Section IV, and Section V concludes this article.

II. Brief Overview of Related Work

Nonlinear intensity difference is a major challenge for multisource RS image matching. Many area-based matching methods have been proposed to deal with this problem.

This matching method is utilized to optimize the coarse registration of RS images, and the adopted similarity measure is a decisive component of area-based methods [29]. Normalized cross correlation is a widely used similarity measure, which can deal with the linear intensity differences [30], while it is sensitive to nonlinear intensity differences. Further, a template matching method named matching by tone mapping (MTM) was proposed to handle nonmonotonic and nonlinear intensity mapping [27]. However, MTM relies on intensity mapping rule between two multisource images. Unfortunately, the intensity relationship between multisource RS images cannot be fitted by a simple function. Mutual information (MI) [25] can adapt to nonlinear intensity differences to some extent and has been extensively used in multisource image matching [31], [32]. Nevertheless, high computational load is a major limitation of MI-based methods.

Frequency-based image correlation is a specific type of area-based image matching technique, which realizes image matching with the translation property of FT or similarity model by means of the image information and operation in the frequency domain [23]. The frequency-based methods are usually time efficient due to the use of FFT. In the frequency domain, PC method can quickly estimate the translation based on Fourier theorem [26]. Nowadays, PC has been extended to deal with rotation and scale estimation using Fourier–Mellin transform [33], [34]. By means of FT and phase information, PC can realize superior performance in theoretical subpixel accuracy. The singular value decomposition and the unified random sample consensus were performed to achieve high subpixel accuracy by rank-one matrix approximation and 1-D fitting [35]. The rank-one matrix factorization with a mixture of Gaussian model on the PC matrix was utilized to consider more complicated noise [36]. Although these PC-based methods are more robust to noise and can get high subpixel matching accuracy, it is often at the expense of efficiency due to the use of iterative algorithm. The time-consuming effect is more serious when they are used to realize the correspondences of CPs, because the number of iterations increases with the number of CPs.

Moreover, some works have attempted to extract structural features as the replacement of image intensity and combined them with the conventional similarity measures to achieve image matching. Normalized gradient correlation (NGC) was proposed to realize image matching by directly using the gradient orientation of each individual pixel [28], which is more sensitive to noise and cannot handle the influence of nonlinear intensity differences. Phase congruency has been a representative feature representation to capture the structural information. In [37], the phase congruency is extended to build a novel feature descriptor named the histogram of orientated phase congruency, which can address the influence of nonlinear intensity differences. A robust matching method based on enhanced subpixel PC adopted phase congruency information as feature representations to reduce the influence of nonlinear intensity differences in multisource cases [38]. Phase congruency information was used as weights during the similarity calculation based on the adaptive multiscale structure orientation (AMSO) [39]. However, phase congruency is highly affected by noise especially in LiDAR and SAR which are usually more noisy, and the feature representation is sparse to realize image matching since most pixel values in the phase

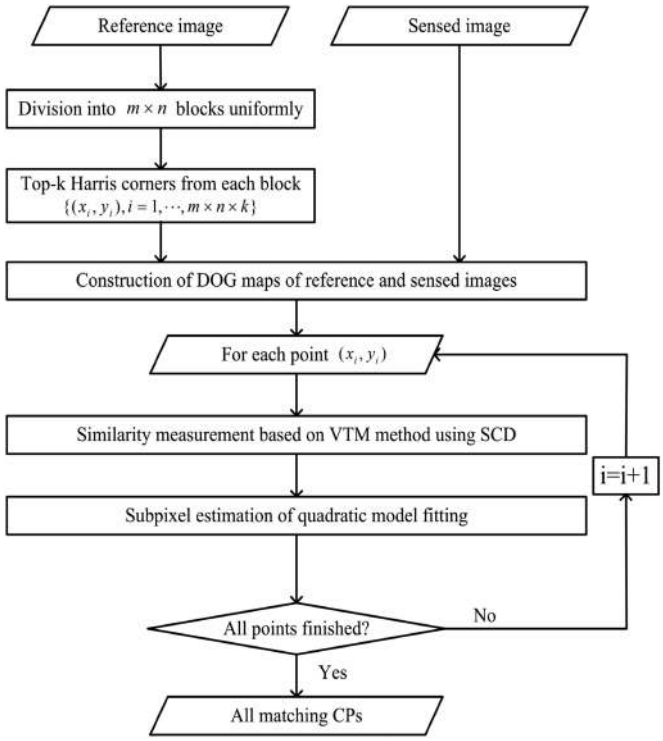

Fig. 1. Overall flowchart of the proposed matching method.

congruency map are close to zero [40], [41]. In [42], channel features of oriented gradients (CFOG) was proposed and measured in frequency domain using FFT. The novel feature is an extension of histogram of gradient (HOG) feature. Nevertheless, the gradient histogram consists of gradient magnitude which is more easily affected by noise.

III. Proposed Approach for Multisource Image Matching

A. Workflow of the Proposed Matching Method

The proposed matching method for multisource images estimates displacement between the corresponding window pairs in the frequency domain. The overall flowchart of the matching method is presented in Fig. 1, which mainly consists three steps as follows.

1) Feature points detection in reference image which is usually visible image. The block strategy is adopted to extract the Harris feature corners uniformly distributed over the reference image. In each block, the top k points with the largest response values are selected as the feature points.
2) Construction of DOG map of both reference and sensed images. In order to reduce the influence of complicated intensity differences and capture dense and useful structural information between multisource images, the DOG feature of each pixel in images is extracted as the replacement of the original image intensity.
3) Corresponding CPs matching. Considering the nonlinear intensity differences even intensity inversion between multisource images, SCD similarity measurement is evaluated between template window and search region, which can be accelerated with FFT and improved with VTM method. Subsequently, based on the similarity map, a quadratic model is utilized to obtain subpixel accuracy.

B. DOG Map Construction

The pixelwise orientation histogram is first established to obtain DOG feature map of image. For each image sample of $I(x, y)$, the gradient magnitude of $m(x, y)$ and orientation of $\theta(x, y)$ are precomputed using pixel differences as

$$m(x,y) = [(I(x+1,y) - I(x-1,y))^2 + ((I(x,y+1) - I(x,y-1))^2]^{\frac{1}{2}} \quad (1)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{I(x+1,y) - I(x-1,y)}{I(x,y+1) - I(x,y-1)}\right). \quad (2)$$

It should be noted that there may be orientation reversal at the matching points of the multisource images due to significant nonlinear intensity differences. We transfer the gradients from $[180^\circ, 360^\circ)$ to $[0^\circ, 180^\circ)$ to deal with the problem.

The pixelwise orientation histogram is formed with the gradient information of sample pixels in the surrounding region around the center point. In this article, the orientation histogram has nine bins covering the 180° range of orientations. Each sample pixel added to the histogram is weighted by its gradient magnitude and by a standard Gaussian-weighted window. The dense orientation histograms are obtained by performing the same operation for each pixel in the image. The orientation histogram can be constructed quickly using convolution with Gaussian kernel. The computation is defined as

$$h_i(x,y) = g_\sigma * h_i{}'(x,y) \quad (3)$$

where g_θ and $*$ operation mean standard Gaussian kernel and convolution operation, respectively. And $h_i{}'(x, y)$ denotes the initial histogram component quantized at the orientation of the ith bin, while $h_i(x, y)$ is the ith bin component of pixel-level orientation histogram.

The DOG feature of each pixel is the orientation corresponding to the bin in which the peak value of each orientation histogram locates. The DOG feature map is finally formed by extracting the corresponding DOG feature of each pixel. As described, the DOG feature map is a non-redundant feature representation with the same size as the original image.

The proposed method utilizes a Gaussian kernel to collect the local gradient information centered on each pixel quickly, which can suppress Gaussian noise effectively. And the unsigned DOG is extracted from orientation histogram. The unsigned orientation means the same angle for two opposite orientations, which can handle the orientation reversal caused by nonlinear intensity differences. Meanwhile, through this operation, the DOG feature discards gradient magnitude which is more easily affected by noise. The analysis of the sensitivity to noise will be shown in the latter experimental part.

In order to verify the robustness of DOG map to nonlinear radiation differences, we present the proposed feature maps, gradient features and phase congruency features of two images in Fig. 2. For two test images, there are significant nonlinear

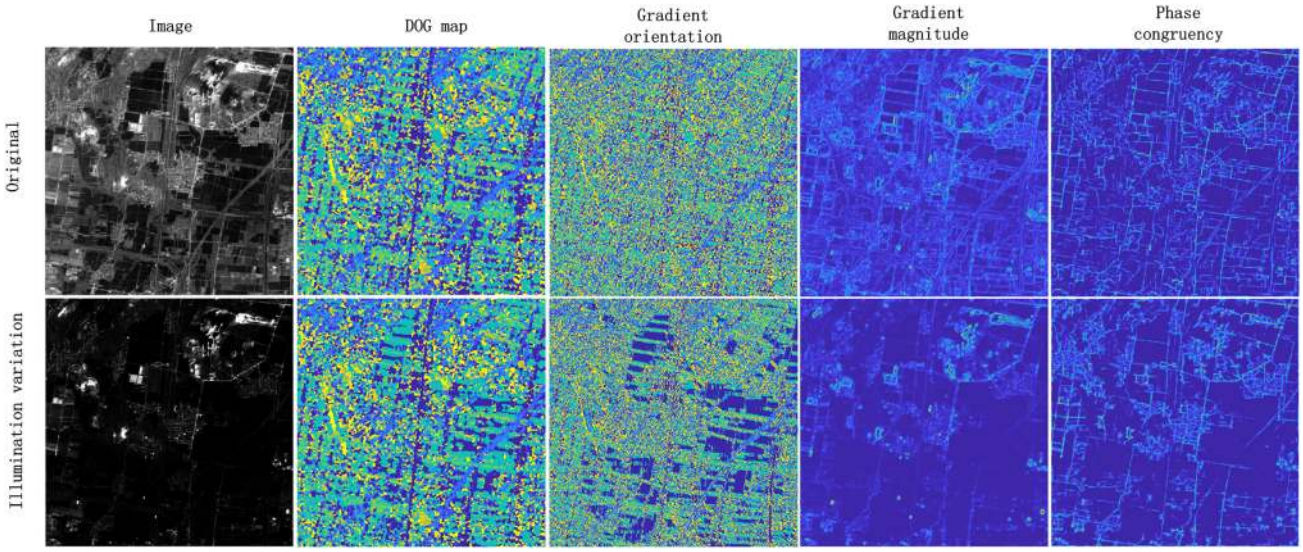

Fig. 2. Comparison of DOG map with gradient information and phase congruency.

intensity differences between them. One can see that phase congruency retains more structure information than gradient features including gradient magnitude and orientation. And the DOG feature map has less changes than phase congruency with illumination variation on the original image. This shows that DOG feature map of image is more robust to nonlinear intensity differences.

C. Fast Similarity Measurement

Based on the DOG feature maps of multisource images, template matching method is employed to detect the CPs between images. The sum of squared differences (SSD) is regarded as an obvious approach for similarity evaluation for image matching. The SSD between the two feature representations with the template window is defined as

$$S_i(u,v) = \sum_{x,y} |D_{1i}(x,y) - D_{2i}(x-u, y-v)|^2 W_i(x,y) \tag{4}$$

where D_{1i} and D_{2i} represent the sub feature maps picked from the DOG maps of reference image and sensed image around the ith keypoint, respectively. $W_i(x,y)$ means the template window function over reference subfeature map, where $W_i(x,y) = 1$ within the template window, otherwise $W_i(x,y) = 0$.

By minimizing the measurement $S_i(u,v)$, the matching function is defined as

$$u_i, v_i = \arg\min_{u,v} \{S_i(u,v)\} \tag{5}$$

where (u_i, v_i) represents the offset vector which matches D_{1i} with D_{2i}.

However, the experimental results show that the CPs detected by SSD measurement are not precise enough, which would be illustrated later. The main reason is the orientation distortion and reversal caused by local geometric and intensity distortions. Especially, when the distortions occur at the orientation periodic node (180° in this article), the similarity is minimal judged by SSD, which is opposite to the actual situation. For example, assuming the orientations of two corresponding pixels are 0° and 179°, respectively, that is caused by intensity inverse and slight local geometric distortion, the judgment of SSD is the least similarity, which is unreasonable.

Therefor, SSD is not suitable for the proposed DOG feature. The SCD method is proposed to measure similarity and cope with the interfere of orientation distortions. Considering the orientation range from 0° to 180°, we adjust the period of the cosine function to the same period as the orientation. The SCD is defined as

$$C = \cos(2 \cdot \Delta\phi) \tag{6}$$

where C and $\Delta\phi$ represent the similarity value and difference of two orientations, respectively. Fig. 3 shows the curve of the SCD measure as the absolute difference of two orientations. Because the function is a symmetric function, we use the absolute value of the orientation difference to analyze the proposed measurement here.

When the absolute difference is 0° or 180°, the proposed similarity measure obtain the maximum similarity value, while the similarity measure become the smallest value when the absolute orientation difference is 90°. Consequently, the proposed similarity measure can handle the influence of orientation reversal and orientation distortion.

In template matching, the SCD is presented as

$$C_i(u,v) = \sum_{x,y} \cos(2 \cdot (D_{1i}(x,y) - D_{2i}(x-u, y-v))) \cdot W_i \tag{7}$$

where $C_i(u,v)$ is the similarity measure map obtained by SCD. The rest of the variables have the same meaning as previously mentioned. Then the matching function is given as

$$u_i, v_i = \arg\max_{u,v} \{C_i(u,v)\} \tag{8}$$

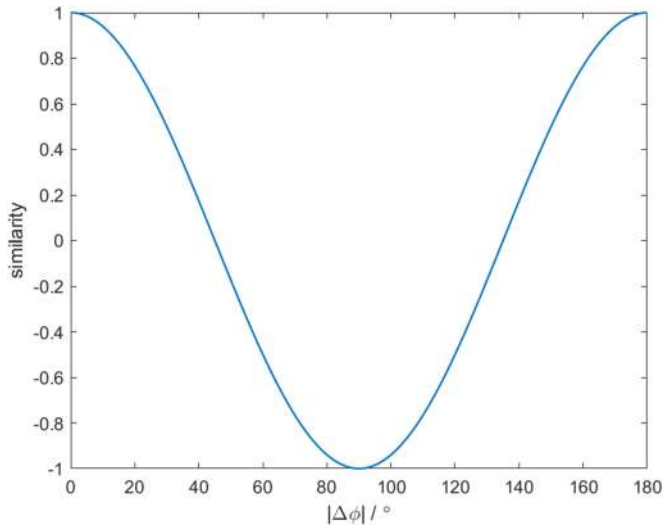

Fig. 3. Proposed similarity measurement versus the orientation difference.

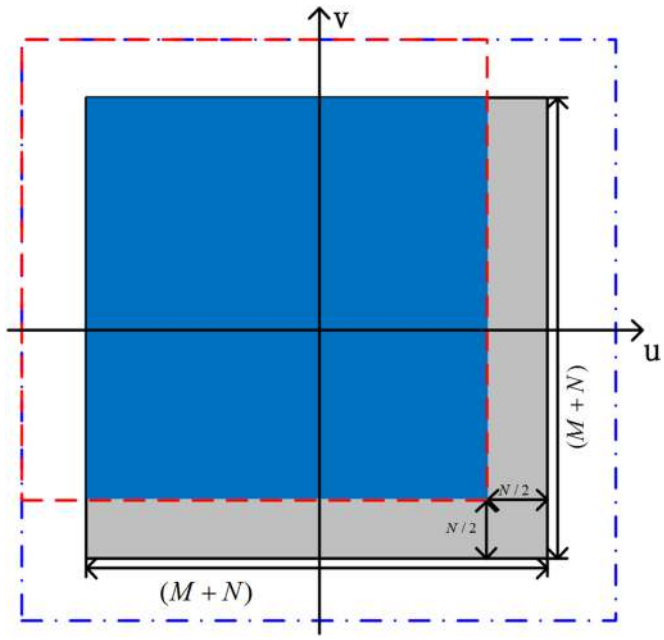

Fig. 4. Actual template with VTM method. The blue overlapping area is the actual matching template size at the corresponding matching point.

Although DOG feature map is a nonredundant feature representation, it is still time consuming for SCD to measure similarity in spatial domain, which can not meet some application requirements. To address that, FFT is considered to accelerate the measurement.

Considering that the cosine function can be expressed as the real part of Euler's formula, the SCD similarity function is rewritten as

$$C_i(u,v) = real\left(\sum_{x,y} e^{j2(D_{1i}(x,y)W_i - D_{2i}(x-u,y-v)W_i)}\right) = real\left(\sum_{x,y} e^{(j2D_{1i}(x,y)W_i)} \cdot e^{(-j2D_{2i}(x-u,y-v)W_i)}\right). \quad (9)$$

As stated above, SCD can be seen as a correlation operation of D_{1i} and D_{2i}, which can be speed up using FFT. The relationship of correlation operated in spatial domain and frequency domain is given as

$$corr(h(x,y), g(x,y)) \Leftrightarrow H^*(X,Y)G(X,Y) \quad (10)$$

where $corr(\cdot)$ is correlation operation. $H^*(X,Y)$ represents the complex conjugate of $H(X,Y)$, which is the forward FFT of $h(x,y)$, and $G(X,Y)$ denotes the forward FFT of $g(x,y)$.

Consequently, the SCD can be represented as

$$C_i(u,v) = real\{\mathrm{F}^{-1}[F^*(e^{j2D_{1i}W_i}) \cdot F(e^{-j2D_{2i}W_i})]\} \quad (11)$$

where F and F^{-1} denote the forward and inverse FFTs, respectively. And F^* is the complex conjugate of FFT. The matching result can be obtained using (8). Given template window of $M \times M$ pixels and search window of $N \times N$ pixels, the matching computing complexity reduces to $O((M+N)^2 log(M+N))$ from $O(M^2N^2)$ using (11). And with the larger template window or search window, the method will reduce more computation.

D. VTM Method

During the conventional template matching procedure, the size of the features participating in the similarity measurement each time remains constant when a template window sliders pixel-by-pixel over a search region. Just as mentioned above, with the computing complexity of $O((M+N)^2 log(M+N))$, only the reference sub feature map of $M \times M$ is used to be measured with the corresponding sensed subfeature map. And generally speaking, larger template results higher matching precision in a local area. Accordingly, we propose VTM method to use the feature information in the sub feature maps as much as possible. The VTM method is defined as

$$C'_i(u,v) = real\{\mathrm{F}^{-1}[F^*(\exp(j2D_{1i})) \cdot F(\exp(-j2D_{2i}))]\} \quad (12)$$

where $C'_i(u,v)$ is similarity map obtained by VTM method. In the case, without of limitation of template window, the actual template size varies spatially. Fig. 4 shows the size of actual template with VTM method. All of feature representations which can be utilized are $(M+N) \times (M+N)$ pixels for both D_{1i} and D_{2i}. The actual matching template size varies with the 2-D offset when D_{1i} sliders over D_{2i}, just like the overlapping region as shown in Fig. 4. That means that all candidate matching points in a search region correspond to different actual template sizes.

Hence, the relationship between the actual template size and 2-D offset is expressed as

$$Mask(u,v) = (M+N-|u|) \times (M+N-|v|) \quad (13)$$

where $Mask(u,v)$ is the actual template size corresponding 2-D offset (u,v). Since the template size is not constant, it is not suitable to measure similarity by SCD no longer. Accordingly, based on $C'_i(u,v)$, mean of cosine difference (MCD) measurement is

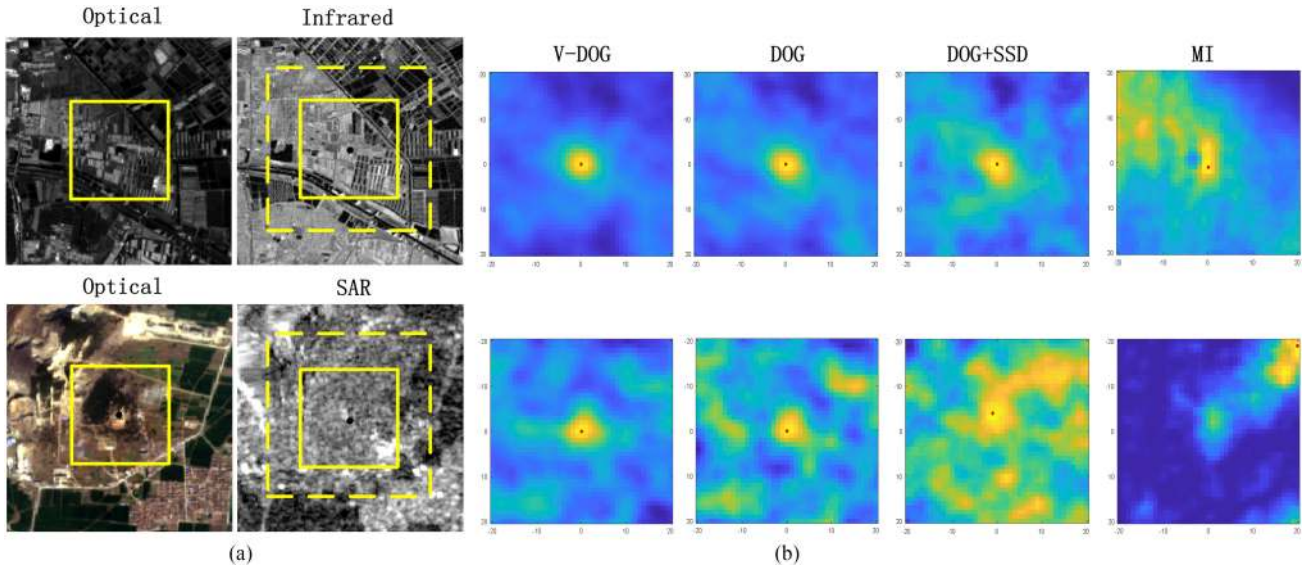

Fig. 5. Similarity maps of V-DOG, DOG, DOG+SSD, and MI, where the template size is 50×50 pixels and the search window is 40×40 pixels. (a) Test images. (b) Similarity maps where the red points mean the positions with the maximum similarities.

defined as

$$C_i(u,v) = \frac{C_i{}'(u,v)}{Mask(u,v)} \tag{14}$$

where $C_i(u,v)$ is the similarity map obtained with VTM and MCD. The matching result can be obtained in (8). Compared with the conventional template matching method, the template gain with VTM method is represented as $(1+(N-|u|)/M) \times (1+(N-|v|)/M)$. Given the search window of $N \times N$ pixels, the maximum offset is $N/2$, which means that using VTM method, the maximum template gain is $(1+N/M)^2$ and the minimum gain is $(1+N/2M)^2$. The gain effect is related to template to search window ratio, which is presented as α.

The similarity maps of the proposed method are compared with SSD and MI to verify the effectiveness of the proposed matching methods. Two pairs of multisource images are used in the test, which are the optical-to-infrared and optical-to-SAR image pairs with different resolutions, respectively.

Fig. 5 shows comparison results of these measures, where V-DOG means test results obtained with VTM method and DOG feature map. All measures perform well for optical-to-infrared images. However, the DOG features with SSD measure (DOG+SSD) and MI occur location errors for optical-to-SAR images and the similarity maps look more noisy. In contrast, both methods proposed in this article find the correct match relationships in all cases. Moreover, the similarity maps of V-DOG are smoother with sharp peaks. The test results confirm that the proposed methods are resistant to nonlinear intensity differences between images and can cope with orientation distortions.

The VTM method can not only be applied to the matching method proposed, but also in some methods of template matching through FFT, which can improve the matching performance without increasing the computational complexity. More analysis is given in Section IV.

E. Subpixel Calculation in the Spatial Domain

Subpixel calculation aims to determine the precise subpixel location of the similarity measurement peak. In this article, a quadratic model, which is commonly used fitting model, is applied as an approximation to estimate the subpixel shifts. Assuming a paraboloid function is denoted as $P(x,y) = a_0x^2 + a_1y^2 + a_2xy + a_3x + a_4y + a_5$, the similarity peak location $(\Delta x, \Delta y)$ can be acquired through coefficients $a_i (i = 0, \ldots, 5)$ as

$$\Delta x = \frac{2a_1a_3 - a_2a_4}{a_2^2 - 4a_0a_1} \tag{15}$$

$$\Delta y = \frac{2a_0a_4 - a_2a_3}{a_2^2 - 4a_0a_1} \tag{16}$$

where the coefficients can be calculated by least square fitting given some neighbors around the integer-valued peak location.

IV. Processing Results

The datasets, experimental results and their evaluation are presented. The results are compared with some classic and state-of-the-art similarity measures, such as MI, NGC, FHOG [42], CFOG, and AMSO.

A. Datasets

Eight pairs of multisource RS images are used to analyze the matching performance of the proposed methods. The detailed information of these test images are shown in Table I. The test images consist of optical-to-infrared (No. 1, No. 2, No. 3), optical-to-LiDAR (No. 4) and optical-to-SAR (No. 5, No. 6, No. 7, No. 8). These images cover a variety of terrain, such as urban, suburb, farmland, river, and island. Fig. 6 shows the test image pairs, which have been prerectified using their physical sensor models. The rectification is employed to remove nearly all

TABLE I
DETAIL INFORMATION OF THE TEST IMAGE PAIRS

Category	Image No.	Source	GSD(m)	Size	Date	Location
Optical-to-Infrared	No.1	Daedalus optical	0.5	512×512	04/2000	Images cover urban area with buildings
		Daedalus infrared	0.5		04/2000	
	No.2	Landsat8	30	512×512	08/2013	Images cover suburban area with farmland
		Landsat8	30		08/2013	
	No.3	Daedalus optical	0.5	512×512	04/2000	Images cover a factory
		Daedalus infrared	0.5		04/2000	
Optical-to-LiDAR	No.4	Airborne optical	2.5	524×524	06/2012	Images cover urban area with high buildings
		Lidar depth	2.5		06/2012	
Optical-to-SAR	No.5	TM band3	3	500×500	12/2014	Images cover urban area with buildings
		TerraSAR-X	3		12/2013	
	No.6	Sentinel-1	10	512×512	07/2017	Images cover rivers and islands
		Sentinel-2	10		07/2017	
	No.7	Landsat	30	600×600	12/2014	Images cover suburban area with farmland
		TerraSAR-X	30		12/2013	
	No.8	Sentinel-1	10	512×512	07/2017	Images cover urban area with buildings
		Sentinel-2	10		07/2017	

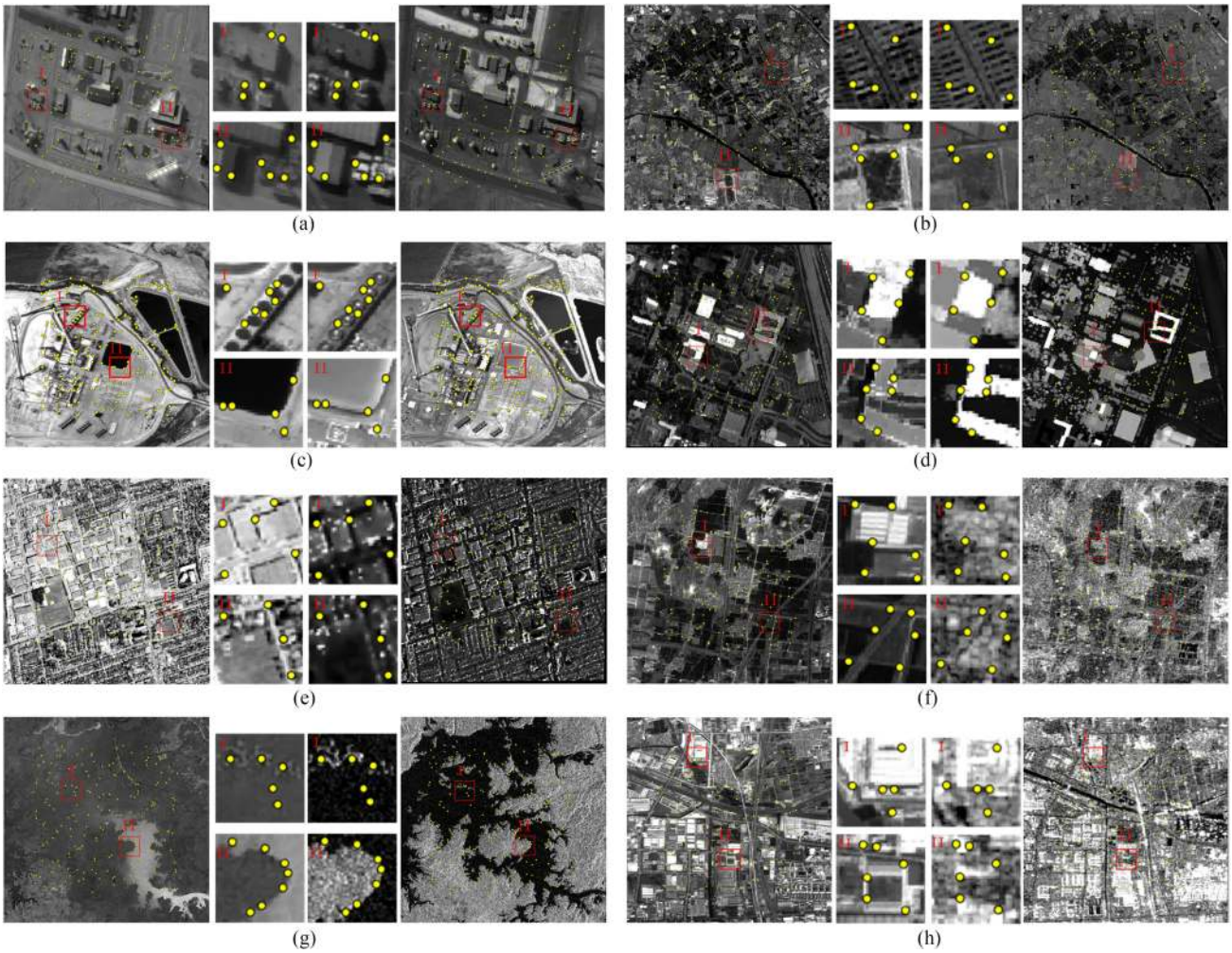

Fig. 6. Matching results of all test image pairs. (a) No. 1. (b) No. 2. (c) No. 3. (d) No. 4. (e) No. 5. (f) No. 6. (g) No. 7. (h) No. 8.

the global geometric distortions from the unregistered images, including obvious rotation and scale differences. In addition, the images of each pair are resampled to the same resolution or ground sample distance (GSD) to facilitate the subsequent matching process.

B. Implementation and Evaluation

In the experiments, block strategy is first employed to obtain 200 uniformly distributed Harris keypoints in the reference images [43], [44]. Then, the CPs can be determined within search window in the sensed image. Moreover, the subpixel accuracy of each CP is achieved by local fitting technique based on quadratic polynomial.

The matching performance is evaluated in terms of correct matching ratio (CMR), root-mean-square errors (rmse), and time consumption. In order to determine the correct matching points, the projective mapping model for each image pair is estimated using 30 manually selected CPs. The projective model is used to calculate the location errors of the matching points obtained by template matching method. The matching points within positioning errors of 1.5 pixels are defined as the correct matching points. The CMR is defined as the ratio between the number of correct matching points and the number of total matching points. The rmse value is denoted as

$$RMSE = \sqrt{\frac{1}{N_c}\sum_{i=1}^{N_c}(x_i - m_i)^2 + (y_i - n_i)^2} \qquad (17)$$

where N_C is the number of correct matching points. (x_i, y_i) and (m_i, n_i) are actual matching point transformed by mapping model and matching point obtained by template matching method for each keypoint, respectively.

C. Effect Analysis of VTM

As analyzed in Section III, the template gain will raise with the increase of α when using VTM method, which is expected to improve the performance of CPs matching. To verify the effect of VTM method, the test images are matched in VTM and conventional template matching method, respectively. For better reflecting the influence of α on the performance of VTM method, the template size is fixed as 50×50 pixels in the test, which is expected to remain stable in the performance of conventional template matching method. In this case, α varies with the size of search window. It is worth noting that in the experiment, each actual matching point is added with a uniform random offset away from the center of search region. Because the matching performance would be better if the actual matching points always locate in the center of the search region where the matching points have larger actual template size. This is just to achieve more representative and convincing statistical results of experimental performance.

Figs. 7 and 8 show the average CMR and rmse values versus the variable α, respectively. It can be seen from Fig. 7 that the CMR index of fixed template matching method is not stable as expected. When α is small, there is no more selection for matching point, so as to achieve a higher CMR value. In contrast, when α become larger, there is more interference to the CPs matching, which results in the decline of CMR value. However, the CMR values achieved by VTM method raise with the increase of α. They are much higher than the ones obtained by conventional template matching method.

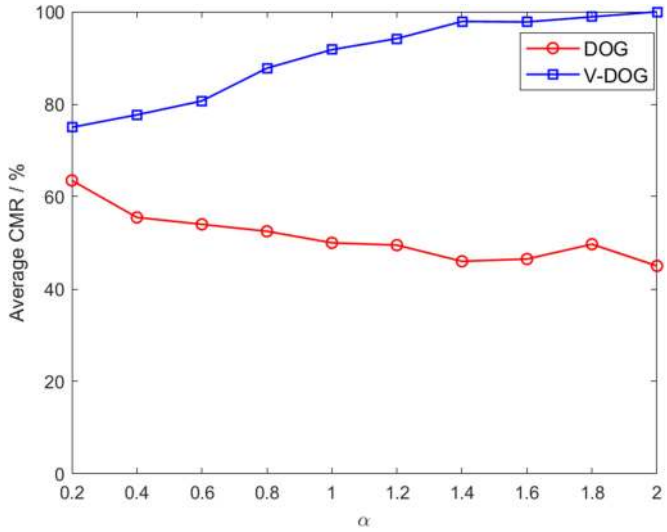

Fig. 7. Average CMR values with different values of α, the ratio between the template size, and search window size.

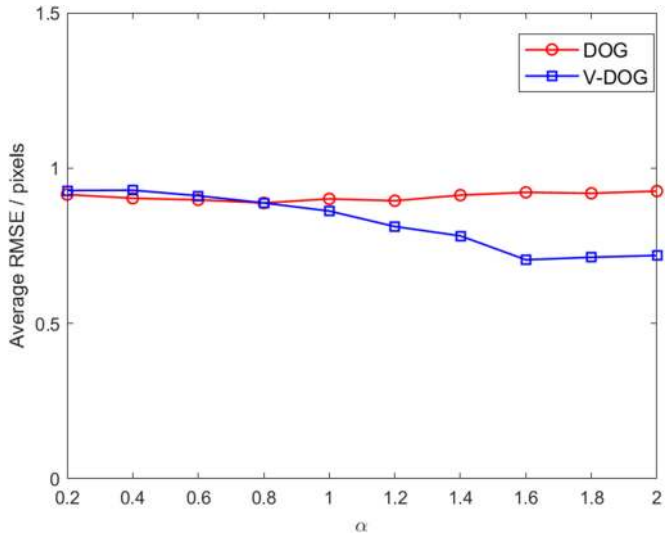

Fig. 8. Average rmse values with different values of α, the ratio between the template size, and search window size.

It can be noticed in Fig. 8 that the rmse values of the correct CPs are stable with the change of α in conventional template matching method, while the values achieved by the VTM method decrease slowly with the increase of α, which indicates that the VTM method can effectively improve matching performance.

D. Analysis of Noise Sensitivity

In this section, some images with noise are used to assess all template matching methods mentioned above. Because it is impossible to use a simple mathematical model to perfectly fit

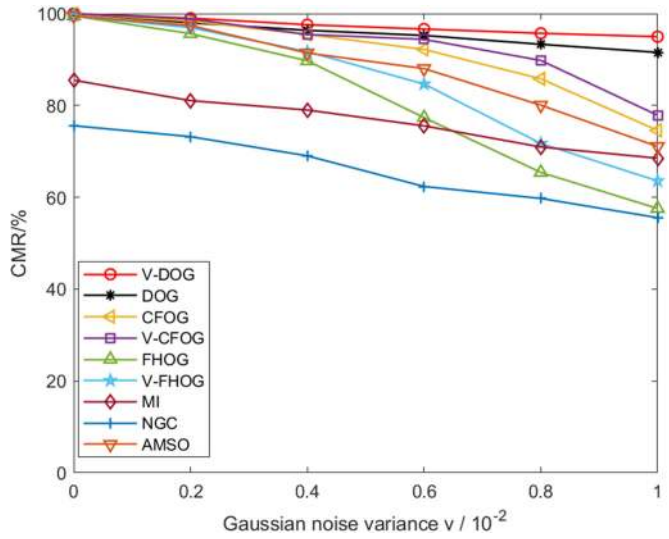

Fig. 9. Average correct matching rates of similarity measures versus various Gaussian noise.

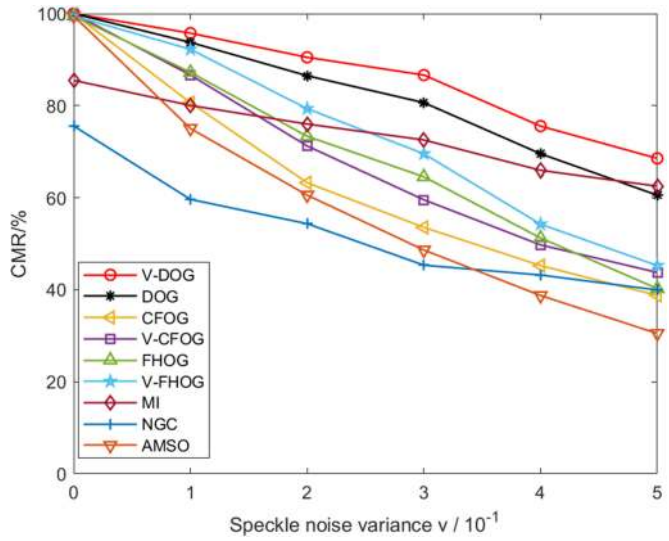

Fig. 10. Average correct matching rates of similarity measures versus various speckle noise.

the nonlinear intensity relationship between multisource images. Meanwhile, LiDAR and SAR images are usually more noisy than infrared images, which makes it hard to measure the effect of noise on image matching. Consequently, we use three pairs of real optical-to-infrared images instead of synthetic images to test the noise sensitivities of these methods. Here, Gaussian and speckle noise are considered to be added to the sensed images. The different levels of Gaussian noise with mean 0 and variances in the range of [0, 0.01] with interval of 0.002, and the different levels of speckle noise with variances in the range of [0, 0.5] with interval of 0.1 are added to the sensed images, respectively. All the similarity measures are performed with template window of 100×100 pixels. The average CMRs are presented in Figs. 9 and 10.

It can be observed that the proposed matching method achieves superior results under increased noise. The results of CFOG and FHOG decrease significantly as noise increased. This is because the two methods are based on the gradient orientation histogram which is more closely related to gradient amplitude, while DOG feature extract the dominant component from the gradient orientation histogram, which discards the amplitude information and reduce the influence of noise. Although MI usually presents stable results, its CMR values are lower than the best performance. NGC method always gets the lowest precision with both Gaussian and speckle noise, which indicates that compared with gradient orientation of each individual pixel, the DOG feature is more robust to noise. In addition, AMSO also presents a higher sensitivity to noise compared with V-DOG.

E. Performance Analysis

We compare the performance of the proposed method with some classic and state-of-the-art similarity measures, such as NGC, MI, FHOG, CFOG, and AMSO. The VTM method can be utilized in the CFOG and FHOG to obtain the corresponding VTM+ methods, which are termed as V-FHOG and V-CFOG, respectively. And like mentioned in the experiment of analyzing the effect of VTM, we do a random translation for each feature point to be matched, but it still remains in the search region. And the random shift against the ground truth is recorded for analyzing the matching performance.

Fig. 11 shows the CMR values of all similarity measures on the test image pairs. In most cases, NGC presents the worst results, and it can hardly work in some cases, which indicates that NGC cannot handle the multisource image matching. This is because there is no any process to handle the nonlinear intensity differences in both feature representation and similarity measure, and at the same time it loses the ability to cope with multisource image matching. The matching performance of AMSO is obviously lower than the DOG-based, CFOG-based, and FHOG-based methods, especially when matching optical-to-SAR images. This may be because AMSO uses phase congruency as weights during the similarity calculation, which is easily affected by noise. The performance of MI is also not satisfactory, which shows that MI can not tackle multisource RS image matching effectively. The proposed V-DOG measure always achieves the best results in all cases. Moreover, all VTM+ methods get better CMR values than those obtained by original methods, which verifies the effectiveness of the VTM method. CFOG-like and FHOG-like methods are all based on orientation histogram. The amplitudes of histogram are composed of gradient amplitudes of neighborhood pixels, which are easily affected by pixel intensity changes. The proposed DOG feature map is constructed without amplitude information, which is most likely to avoid the influence of nonlinear intensity differences and noise. Hence, although DOG feature map is a non-redundant structure representation, which may not be as informative as CFOG or FHOG features, it still achieves similar matching results. V-DOG even achieves the better results than those by V-CFOG and V-HOG.

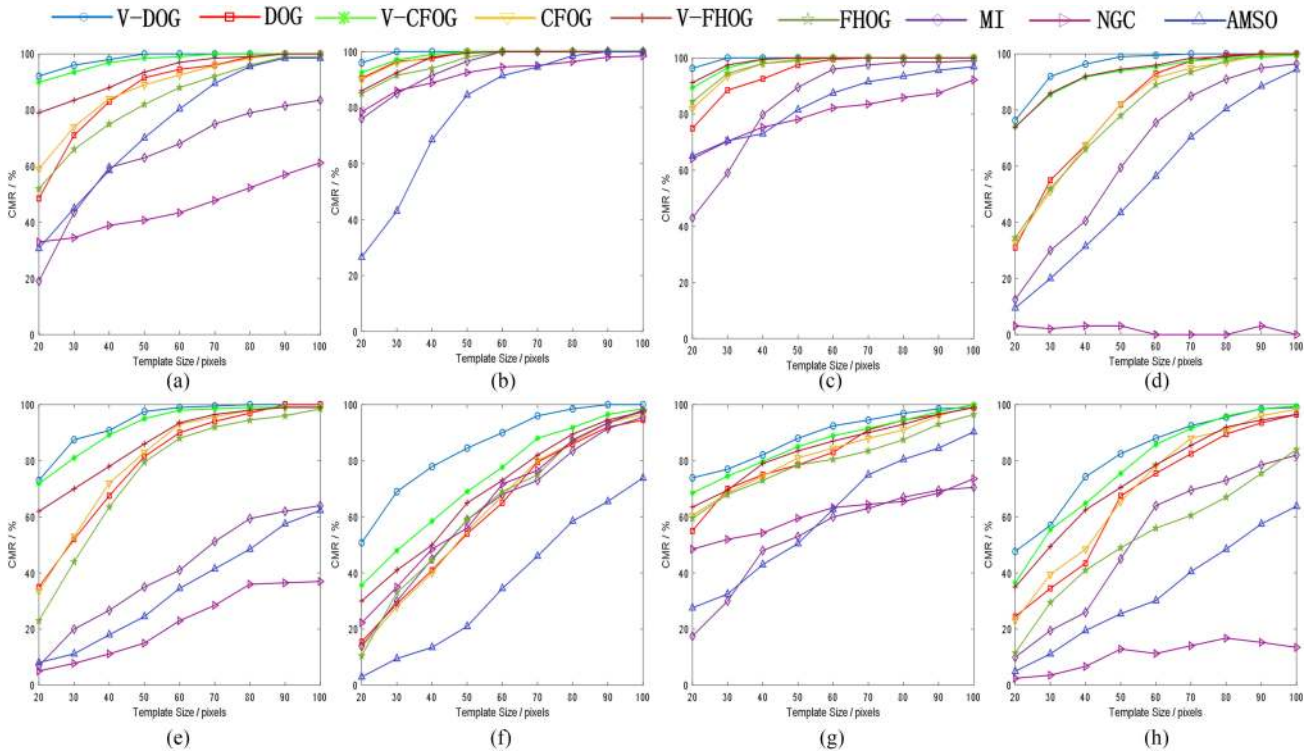

Fig. 11. CMR values of all similarity measures versus the template size. (a) No. 1. (b) No. 2. (c) No. 3. (d) No. 4. (e) No. 5. (f) No. 6. (g) No. 7. (h) No. 8.

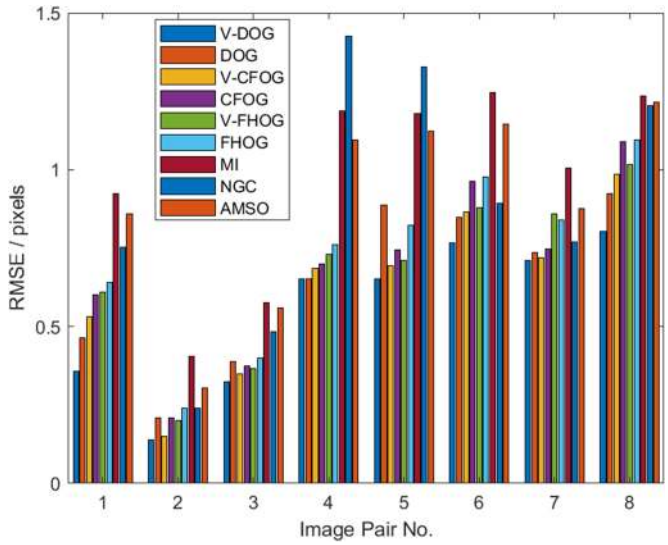

Fig. 12. RMSE values of all similarity measures for each image pair.

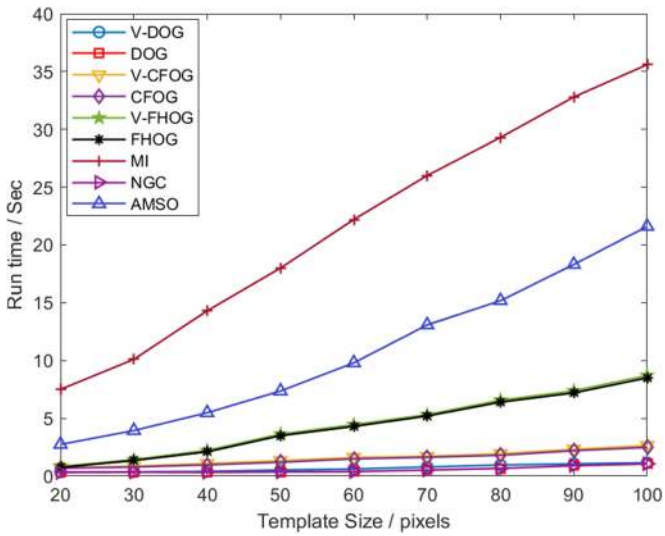

Fig. 13. Time consumption of all similarity measures with different template sizes.

Fig. 12 shows the rmse values of correct matches of all similarity measures mentioned above with template window of 100×100 pixels. As we can see, the proposed V-DOG achieves the smallest rmse values. And all VTM+ methods improve the performance of the original methods. Furthermore, the matching results obtained by V-DOG are given in Fig. 6. It can be seen from the partial enlarged drawing that the correspondences between images are detected precisely.

F. Computational Load

Both nonredundancy of DOG feature and acceleration in frequency domain are for fast matching. Therefore, run time of all similarity measures are compared under different template sizes in Fig. 13. The run time is tested on a platform with an Intel Xeon Gold 6128 CPU. As MI needs to compute statistical information about pixel values, it costs most time among these

measures. AMSO is the second most time-consuming measurement method because it measures similarity with iterative process in spatial domain. Except for MI and AMSO, the rest measurements are all accelerated using FFT, which are time efficient. Due to the mean operation, the run time of VTM+ methods are slightly more than the corresponding original methods. Moreover, DOG and NGC measures cost the shortest time than CFOG and FHOG due to nonredundant feature representation instead of histogram information.

All these results have confirmed that the proposed method is time efficient and robust against the nonlinear intensity differences between images. Furthermore, the VTM method can improve the performance of image matching without increasing the computational complexity.

V. Conclusion

Multisource image matching has always been challenging, which often requires human intervention to achieve a better matching result, and is time-consuming and laborious. In this article, an efficient template matching method is used to realize fast matching for multisource RS images. We propose a nonredundant structural representation termed as DOG feature, which is robust to significant nonlinear intensity differences. Correspondingly, SCD measure is proposed to effectively reduce the influence of orientation distortions caused by noise, intensity deformations, and local geometric distortions. Meanwhile, the similarity measure is accelerated using FFT. In addition, a relatively general method, the VTM method, for template matching is proposed to improve the performance of multisource RS images matching. The VTM method can effectively improve the matching performance without increasing the computational complexity, especially for those template matching methods based on spatial similarity measurements that can be accelerated by FFT.

The proposed approach is based on preregistration images and cannot deal with rotation and scale changes, which limits the application scope at this stage. Future work will focus on improving its robustness to rotation and scale differences.

References

[1] Y. Zhang, "Understanding image fusion," *Photogram. Eng. Remote Sens.*, vol. 70, no. 6, pp. 657–661, 2004.

[2] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proc. IEEE.*, vol. 101, no. 3, pp. 609–630, Mar. 2013.

[3] B. Zitov and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[4] H. Gonalves, J. A. Gonalves, L. Corte-Real, and A. C. Teodoro, "CHAIR: Automatic image registration based on correlation and hough transform," *Int. J. Remote Sens.*, vol. 33, no. 24, pp. 7936–7968, 2012.

[5] P. Bunting, F. Labrosse, and R. Lucas, "A multi-resolution area-based technique for automatic multi-modal image registration," *Image Vis. Comput.*, vol. 28, no. 8, pp. 1203–1219, Aug. 2010.

[6] M. L. Uss, B. Vozel, V. V. Lukin, and K. Chehdi, "Multimodal remote sensing image registration with accuracy estimation at local and global scales," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6587–6605, Nov. 2016.

[7] L. Yu, D. Zhang, and E.-J. Holden, "A fast and fully automatic registration approach based on point features for multi-source remotesensing images," *Comput. Geosci.*, vol. 34, no. 7, pp. 838–848, Jul. 2008.

[8] M. Chen and Z. Shao, "Robust affine-invariant line matching for high resolution remote sensing images," *Photogrammetric Eng. Remote Sens.*, vol. 79, no. 8, pp. 753–760, 2013.

[9] 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.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[11] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 346–359.

[12] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "Kaze features," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 214–227.

[13] A. Sedaghat and H. Ebadi, "Remote sensing image matching based on adaptive binning SIFT descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5283–5293, Oct. 2015.

[14] A. Sedaghat and N. Mohammadi, "Uniform competency-based local feature extraction for remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 135, pp. 142–157, Jan. 2018.

[15] L. Huang and Z. Li, "Feature-based image registration using the shape context," *Int. J. Remote Sens.*, vol. 31, no. 8, pp. 2169–2177, 2010.

[16] P. Schwind, S. Suri, P. Reinartz, and A. Siebert, "Applicability of the SIFT operator to geometric SAR image registration," *Int. J. Remote Sens.*, vol. 31, no. 8, pp. 1959–1980, Mar. 2010.

[17] B. Fan, C. Huo, C. Pan, and Q. Kong, "Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT," *IEEE Geosci. Remote Sens. Lett*., vol. 10, no. 4, pp. 657–661, Jul. 2013.

[18] W. Ma *et al.*, "Remote sensing image registration with modified SIFT and enhanced feature matching," *IEEE Geosci. Remote Sens. Lett*., vol. 14, no. 1, pp. 3–7, Jan. 2017.

[19] K. Yang, A. Pan, Y. Yang, S. Zhang, S. H. Ong, and H. Tang, "Remote sensing image registration using multiple image features," *Remote Sens.*, vol. 9, no. 6, 2017, Art. no. 581.

[20] M. Gesto-Diaz, F. Tombari, D. Gonzalez-Aguilera, L. Lopez-Fernandez, and P. Rodriguez-Gonzalvez, "Feature matching evaluation for multimodal correspondence," *ISPRS J. Photogrammetry Remote Sens.*, vol. 129, pp. 179–188, Jul. 2017.

[21] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale-invariant feature matching for optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4516–4527, Nov. 2011.

[22] Y. Ye, J. Shan, S. Hao, L. Bruzzone, and Y. Qin, "A local phase based invariant feature for remote sensing image matching," *ISPRS J. Photogrammetry Remote Sens.*, vol. 142, pp. 205–221, Aug. 2018.

[23] X. Tong *et al.*, "Image registration with fourier-based image correlation: A comprehensive review of developments and applications," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 4062–4081, Oct. 2019.

[24] J. P. Lewis, "Fast template matching," in *Proc. Vis. Interface*, 1995, pp. 120–123.

[25] A. A. Cole-Rhodes, K. L. Johnson, J. LeMoigne, and I. Zavorin, "Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1495–1511, Dec. 2003.

[26] R. N. Bracewell and R. N. Bracewell, *Fourier Transform and Its Applications*. New York, NY, USA: McGraw-Hill, 1965.

[27] Y. Hel-Or, H. Hel-Or, and E. David, "Matching by tone mapping: Photometric invariant template matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 317–330, Feb. 2014.

[28] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou, and T. Stathaki, "Robust FFT-based scale-invariant image registration with image gradients," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1899–1906, Oct. 2010.

[29] R. Feng, Q. Du, X. Li, and H. Shen, "Robust registration for remote sensing images by combining and localizing feature- and area-based methods," *ISPRS J. Photogrammetry Remote Sens.*, vol. 151, pp. 15–26, 2019.

[30] J. Ma, J. C. W. Chan, and F. Canters, "Fully automatic subpixel image registration of multiangle CHRIS/Proba data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2829–2839, Jul. 2010.

[31] H.-M. Chen, M. K. Arora, and P. K. Varshney, "Mutual informationbased image registration for remote sensing data," *Int. J. Remote Sens.*, vol. 24, no. 18, pp. 3701–3706, Jan. 2003.

[32] 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.

[33] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.
[34] X. Wan, J. Liu, H. Yan, and G. L. Morgan, "Illumination-invariant image matching for autonomous UAV localisation based on optical sensing," *ISPRS J. Photogrammetry Remote Sens.*, vol. 119, pp. 198–213, Sep. 2016.
[35] X. Tong *et al.*, "A novel subpixel phase correlation method using singular value decomposition and unified random sample consensus," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4143–4156, Aug. 2015.
[36] Y. Dong, T. Long, W. Jiao, G. He, and Z. Zhang, "A novel image registration method based on phase correlation using low-rank matrix factorization with mixture of Gaussian," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 446–460, Jan. 2018.
[37] 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, Mar. 2017.
[38] Z. Ye *et al.*, "Robust fine registration of multisensor remote sensing images based on enhanced subpixel phase correlation," *Sensors*, vol. 20, no. 15, Aug. 2020, Art. no. 4338.
[39] J. Lu, F. Sun, and J. Dong, "A novel multi-sensor image matching algorithm based on adaptive multiscale structure orientation," *IEEE Access*, vol. 7, pp. 177474–177483, 2019.
[40] P. Kovesi, "Phase congruency detects corners and edges," in *Proc. Australian Pattern Recognit. Soc. Conf., DICTA*, vol. 2003, 2003, pp. 309–318.
[41] Y. Xiang, F. Wang, L. Wan, and H. You, "SAR-PC: Edge detection in Sar images via an advanced phase congruency model," *Remote Sens.*, vol. 9, no. 3, 2017, Art. no. 209.
[42] 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.
[43] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vis. Conf.*, 1988, pp. 1–5.
[44] Y. Ye and J. Shan, "A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences," *ISPRS J. Photogrammetry Remote Sens.*, vol. 90, pp. 83–95, Apr. 2014.

Dongxing Liang received the B.Eng. degree in electrical engineering in 2018 from Xidian University, Xi'an, China, where he is currently working toward the Ph.D. degree in radar image processing.

Jinshan Ding (Member, IEEE) is currently a Professor with the School of Electronic Engineering, Xidian University, Xi'an, China. He founded the millimeter-wave and THz research group in Xidian University in 2014. His research interests include millimeter-wave and THz radar, video SAR, and machine learning in radar.

Yuhong Zhang (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Xidian University, Xi'an, China, in 1988.

From 1988 to 1993, he was with the Institute of Electronic Engineering, Xidian University, where he has served as an Associate Professor and the Deputy Director. From 1994 to 1998, he was with Syracuse University, Syracuse, NY, USA, as a Visiting Associate Professor. From 1998 to 2014, he was a Senior Scientist with Stiefvater Consultants, Rome, NY, USA, and Research Associates for Defense Conversion Inc., Rome, where he worked on-site with the Air Force Research Laboratory from 1998 to 2010. He is currently a Professor with the School of Electronic Engineering, Xidian University. His research interests include array signal processing, signal modeling and simulation, synthetic aperture radar imaging, and waveform diversity.