Automatic Matching of Multispectral Images Based on Nonlinear Diffusion of Image Structures

Ruixiang Li, Haitao Zhao, Xiaoye Zhang, Xiaosan Ge, Zhanliang Yuan, and Qin Zou, *Senior Member, IEEE*

***Abstract*—Imaging with different spectrums often leads to significant nonlinear differences in the intensity, which brings a great challenge to the automatic matching of multispectral images. Considering that there are often limitations to only using intensity information, this article studies multispectral image matching based on the structure consistency. First, an extended log-Gabor filter is constructed to build phase congruency maps, which encodes the structure information and provides rich and robust features. Then, a nonlinear diffusion-based algorithm is developed to detect the salient feature points on the phase congruency map, which are expected to be illumination and contrast invariant. Finally, the structure descriptors are built according to the orientation of the maximum log-Gabor filter responses, and the image matching is achieved by computing the correspondence. Extensive experiments are carried out on various multispectral image datasets. The results show that the proposed method holds a stable performance over the nonlinear intensity variance across spectrums, and outperforms the comparison methods in terms of the number of correct matches and the matching precision.**

***Index Terms*—Multispectral image matching, nonlinear diffusion, phase congruency (PC), structural descriptor.**

I. Introduction

MULTISPECTRAL imaging has attracted intensive research interests in recent years as it captures richer scene information than common visible-light imaging [1]. The fusion of multispectral images can utilize complementary spectral information and increase the accuracy of image analysis. In image fusion, a critical process is the image matching, which automatically establishes the correspondences between two images. Image matching is also a prerequisite step to integrate their spectral information for the subsequent multispectral image processing and analysis tasks, e.g., object detection [2], environmental surveillance [3], image registration [4], and image fusion and classification [5], [6], etc. However, the spectral inconsistency often leads to nonlinear intensity differences and local detail discrepancies between multispectral images [7], [8], e.g., the visible and infrared imageries, which makes the multispectral image matching a challenging problem.

Generally, the multispectral image matching methods can be divided into two categories: the intensity-based and the feature based. The intensity-based methods use similarity metrics in intensity to determine the correspondence between a pair of images, and obtain the best result by maximizing a similarity metric or minimizing an objective function[9]. Common similarity metrics include the sum of squared differences, the normalized cross-correlation, the mutual information (MI), etc. Among these metrics, MI is considered to be robust to nonlinear intensity differences, and has been successfully applied in multispectral or multisensor image matching [10]. In addition, some intensity-based methods try to adapt to the nonlinear intensity differences by improving the similarity measurement [11], [12]. These methods avoid the step of feature detection and generate good results, but they are sensitive to geometric differences and computationally expensive. Different from intensity-based methods, feature-based methods first extract salient features, e.g., points, edges, lines, contours, and region, etc., between images [13]–[17], construct descriptors to depict the features properties, and then use the similarity measure of the feature descriptors to establish accurate correspondences. The feature-based methods are widely used in image matching because of their robustness to geometric and illumination variance, and achieve remarkable performance. However, image intensities of different spectral bands are in a nonlinear relationship, increasing the difficulty of detection and matching identical feature points in the two images [8], which makes it difficult to establish a stable correspondence for the matching. Though numerous feature-based methods have been proposed for multispectral image matching in the past few decades [4], [7], [8], [18]–[23], it is still a very challenging task.

Current feature-based multispectral image matching methods encounter the following problems.

1) Due to the difference in imaging mechanism, there are obvious intensity differences in the same scene content of different spectrums [8], [18], which makes it difficult to extract reliable common features from them, and seriously affects the result of image matching. When the spectral

Manuscript received August 20, 2020; revised November 6, 2020; accepted December 4, 2020. Date of publication December 8, 2020; date of current version January 6, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 41572341 and in part by the Youth Innovation Promotion Association, Chinese Academy of Sciences under Grant Y8YR2700QM. *(Corresponding author: Haitao Zhao.)*

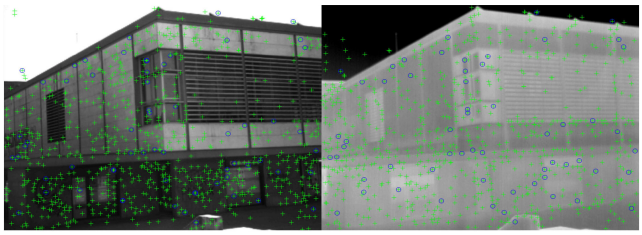
Ruixiang Li, Xiaosan Ge, and Zhanliang Yuan are with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China (e-mail: lirxhpu@163.com; gexiaosan@163.com; yuan6400@hpu.edu.cn).

Haitao Zhao is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China (e-mail: zhaoht@aircas.ac.cn).

Xiaoye Zhang is with the Guangdong Diankeyuan Energy Technology Company Ltd., Guangzhou 510080, China, and also with the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: xiaoyz@whu.edu.cn).

Qin Zou is with the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: qzou@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3043379



(a) (b)

Fig. 1. Detection results produced by SIFT detector. (a) Visible image. (b) Infrared image. The detected features are shown as "+" and the repeated ones as "o."

difference of multispectral images increases, the common information between multispectral images will decrease. The feature points appear in one image may not present in another image. Fig. 1 shows the detection results obtained by the scale-invariant feature transform (SIFT) detector. We can see there are few common features detected between the visible and infrared images.

2) The nonlinear grayscale difference between two spectrums can severely reduce the performance of image-matching algorithms that are based on intensity and gradient. In addition, the infrared images lacking contrast and detail, as compared to visible images [20], would make the situation even worse. These factors undermine the robustness of feature descriptors and lead to unstable image matching.

The gray distortion caused by the nonlinear intensity difference is a great challenge to the multispectral image matching. Through research on image matching, some researchers found that the geometric structure information between multispectral images is more stable than gradients or intensities information under nonlinear intensity variation in multispectral images [20], [24], [25]. Based on this observation, feature points can be detected based on the structure consistency of image, which can be evaluated by calculating similarity metrics on structure or shape descriptors. Phase congruency (PC) model has been demonstrated to capture the structural information consistent with human visual and is robust to illumination and contrast differences [26]. Therefore, some algorithms based on image PC are applied to multispectral or multimodal image matching [11], [25]–[28]. However, PC has limits on itself as it holds the following[26], [29].

1) PC is highly affected by noise since it mainly contains edges structural.
2) The aliasing effect greatly affects the quality of images and the detection accuracy of PC algorithms.
3) PC is not enough for feature description since most pixel values in the PC maps are close to zero.

Motivated by the observation mentioned earlier, we propose a multispectral image matching method based on structure consistency (MMSC). First, the PC maps of the original images are calculated. PC maps reflect the main structure consistency information of the image, which can enhance the similarity between multispectral images. Second, the nonlinear diffusion function, which is insensitive to the nonlinear intensity differences, is used to construct the multiscale space so as to detect the feature points on PC maps. In order to reduce the impact of noncommon feature points, structure consistency constraint is introduced to eliminate noise points and retain salient feature points. Third, we use the orientation of the multiorientation and multiscale log-Gabor filters responses to represent the structure feature of multispectral images. Structural descriptors are established according to the orientation of the maximum mean log-Gabor filter responses. Finally, feature correspondences are calculated according to the distance of feature descriptors. The main contributions of this article are as follows.

1) The log-Gabor filter is extended to construct the PC map. The nonlinear scale space of PC maps is established by the nonlinear diffusion filter, and extreme value is detected in the nonlinear scale space, which makes the feature points be invariant to illumination and contrast.
2) The noncommon feature points may reduce the matching performance. A structure consistency constraint is introduced to retain the salient feature points, which significantly decreases the possibility of false matches.
3) Based on the log-Gabor filter responses, feature descriptors were established according to the orientation of the maximum mean log-Gabor responses. The proposed method is more robust against nonlinear intensity variation.

The remainder of this article is organized as follows. Section II briefly reviews the related work. Section III presents the proposed method in detail, including PC extraction, feature point detection, and feature descriptor construction. Section IV describes the experiments and results, and Section V concludes this article.

II. Related Work

This section briefly reviews the multispectral image matching methods. In general, the research of local features involves the feature detectors and feature descriptors. These feature-based methods rely on extracting highly repeatable features between images [30]. Noncommon feature points will result in the low repeatability of detected features, which may lead to more wrong matches. Typical feature point detection methods, such as Harris [13], SIFT [14], and FAST [31], search the image point locations with strong intensity or gradient variations. However, these detectors usually have difficulty in detecting highly repeatable feature points between multispectral images for the difference of gradient, which substantially degrades the matching performance [24]. Compared with the image gradient, PC model is more robust to changes in illumination and contrast, many researchers have used PC detector [25]–[27] for feature detection. Ye *et al.* [25] proposed a feature detector (MMPC-Lap) and a feature descriptor named local histogram of orientated PC for remote sensing image matching, which is invariant to illumination and contrast variation. Ma *et al.* [27] combine the frequency domain (PC) and the spatial domain (SAR-SIFT operator) to detect image features. The extracted features are robust because it depends on the image structure. Fan *et al.* [32] proposed a uniform nonlinear diffusion-based Harris feature extraction method using the multiscale Harris operator

to improve the accuracy of detection and matching. In addition to the nonlinear intensity difference of pixel intensity between multispectral images, the lack of sufficient local texture details will also reduce the repeatability of feature points [33]. Hence, the detection of feature points need to fully consider the overall structures of multispectral images.

Once feature points are detected, feature descriptions need to be established for matching. Feature descriptors are numeric vectors that encode the characteristics of the local region of the feature points [34]. Traditional feature descriptors are formulated on intensity or gradient domains, such as the SIFT [14], speeded-up robust feature (SURF) [15], and binary robust independent elementary feature [35], perform well on single spectrum images, but they behave poorly when dealing with multispectral data [18]. This is because the pixel values and gradients between different multispectral images often have a nonlinear relationship, which dampens the matching capability of descriptors. To increase the robustness to significant nonlinear intensity changes, research works have been done to improve the performance of the descriptors designed for visible images (e.g., SIFT, SURF) to encode multispectral images. Approaches, such as the orientation-restricted SIFT [36] and multimodal SURF [37], suggested modifications to SIFT or SURF for better matching performance by imposing scale and orientation restrictions. To alleviate gradient magnitude discrepancy, normalized gradient SIFT [23] and partial intensity invariant feature descriptor (PIIFD) [38] normalize gradients to construct the feature descriptors, a certain extent improved matching precision.

However, the feature changes caused by the discrepancies of image local details dampen the ability of these gradient-based methods, which represents a common feature. In addition, the intensity mapping maybe linear, nonlinear, and erratic [4]. Some methods [16], [39] attempted to describe the distributions of straight line segments between visible and infrared images, but these methods are not robust enough due to inaccurate line detection accuracy. Edge feature has proven to be an effective means of describing image features [40], [41]. Edge histogram descriptor (EHD) [42] established feature descriptions by calculating the edge orientation responses of multioriented Sobel spatial filter. Edge-oriented histogram (EOH) [19] descriptor uses the edge distribution of four directional edges and one nondirectional edge to construct the feature description. Edge features are detected by the Canny detector. However, it is difficult to select an appropriate threshold to ensure that the edges extracted from the multispectral image are the same. PC edge histogram (PCEHD) [43] uses PC to detect feature points and edges structure of infrared and visible images, and the edge orientation histogram of feature points is calculated.

Some recent work attempts to solve the problem of nonlinear distortion with image structural information. The PC image embodies the image structure information, and a number of researchers have proposed methods for similarity measures or feature descriptors based on PC [11], [25]. The convolution of the image by the log-Gabor filter can obtain the geometric structure information [44], which is less sensitive to significant intensity changes. Therefore, it is widely used in image feature detection and visual information comprehension [45], [46]. The log-Gabor histogram descriptor (LGHD) [20] uses multiscale and multioriented log-Gabor filters to replace the EOH filters, to deal with the problem of significant nonlinear intensity differences in multispectral images. The LGHD can get richer and more robust feature representation in multispectral images but suffers from high dimensionality and low efficiency. Nunes and Pádua [47] proposed the multispectral feature descriptor, which leverage fewer log-Gabor filters to construct the descriptor, so its computational efficiency is better than LGHD. Compared with gradient-based methods, these descriptors are more robust for multispectral image matching because they use image structure instead of intensity information [48].

The data-driven deep learning (DL) schemes [49] have been successfully applied to various remote sensing tasks, such as hyperspectral image classification, change detection, object detection, and image segmentation [50]–[53]. Since DL can automatically learn high-level features in images, it has been applied to remote sensing image matching recently. Most image matching methods using DL are based on a Siamese network [54], [55]. In addition, the GANs are applied to image matching and registration [56], [57]. One key point of these methods is to transform one image into another image by the trained GANs to eliminate the significant differences between multispectral images. Learning-based methods can cope with complex feature matching satisfactorily. But when these learning-based methods are applied to multispectral image matching, it need a large number of training data to prevent overfitting. Also, it is difficult to train the network for matching due to the diversity of multispectral data.

III. Methodology

The flowchart of the proposed method is presented in Fig. 2. The proposed method mainly consists of following four steps.

1) The log-Gabor filter is extended to construct the PC maps.
2) Build the nonlinear scale space of PC maps by a nonlinear diffusion filter, and extreme value is detected in the nonlinear scale space to obtain the feature points with invariant to illumination and contrast for multispectral image matching.
3) Establish a structural descriptors using log-Gabor filters convolution sequence to describe the feature points obtained from the previous step.
4) SAD distance is used as a matching measure, and the bilateral matching method is used to establish the feature correspondences.

A. PC Based on the Log-Gabor Filter

PC model is a local energy-based model of feature perception, which assumes that the perceived features are located at the points where the Fourier components are maximal in phase. To extend the PC algorithm to 2-D images, Kovesi [26] developed a measure of PC via 2-D log-Gabor filters, which is robust to noise. The 2-D log-Gabor filter [46] is constructed by using a Gaussian function in multiple directions, as defined by

$$L_{n,o}(\omega,\theta)=\exp\left(\frac{-(\log(\omega/\omega_n))^2}{2(\log(\kappa/\omega_n))^2}\right)\exp\left(\frac{-(\theta-\theta_{n,o})}{2\sigma_\theta^2}\right) \tag{1}$$

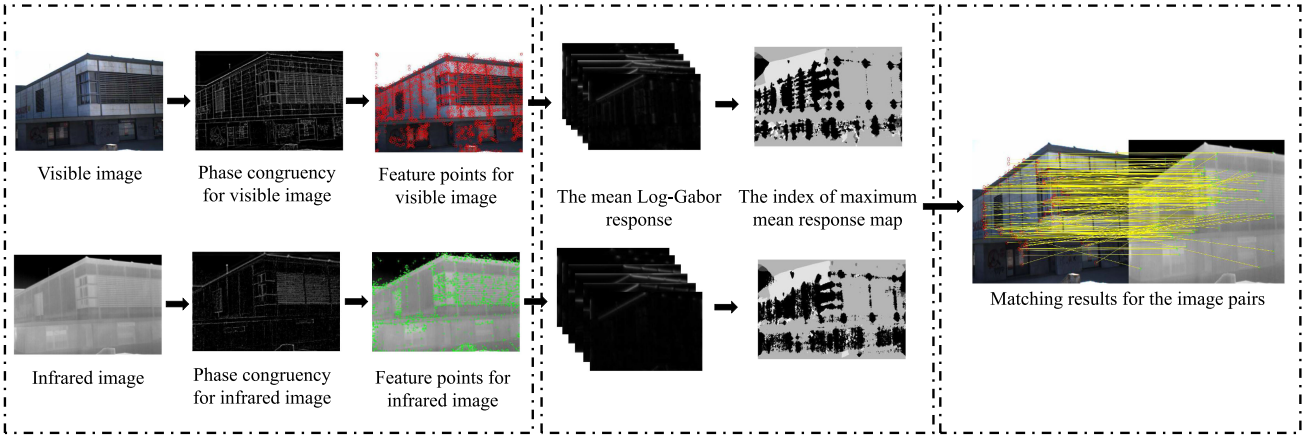


Fig. 2. Illustration of matching by using the proposed method.

where (ω, θ) denote the frequency and angle of the filter, n and o are the scale and orientation of the log-Gabor filter, $(\omega_n, \theta_{n,o})$ are the center frequency and center orientation, κ is the width parameter for the frequency, and σ_θ is the width parameter of the angle.

The log-Gabor filter is a frequency domain filter, and its corresponding spatial domain filter can be obtained by inverse Fourier transform. In the spatial domain, a 2-D log-Gabor filter can be represented as follows:

$$L_{n,o}(x,y) = L_{n,o}^{\text{even}}(x,y) + i \cdot L_{n,o}^{\text{odd}}(x,y) \tag{2}$$

where $L_{n,o}^{\text{even}}(x,y)$ and $L_{n,o}^{\text{odd}}(x,y)$ stand for the even-symmetric (sine) and the odd-symmetric (cosine) log-Gabor filters on scale n and orientation o, respectively. The symbol i is an imaginary unit.

The input image is convolved with each log-Gabor filter to generate responses under different scales and orientations, and their response components are recorded to describe features in the next step. The convolution is defined by

$$\begin{aligned}&[E_{n,o}(x,y), O_{n,o}(x,y)]\\ &\quad = \left[I(x,y) * L_{n,o}^{\text{even}}(x,y), I(x,y) * L_{n,o}^{\text{odd}}(x,y)\right]\end{aligned} \tag{3}$$

where $E_{n,o}(x,y)$ and $O_{n,o}(x,y)$ are the convolution response of $L_{n,o}^{\text{even}}(x,y)$ and $L_{n,o}^{\text{odd}}(x,y)$ at scale n and orientation o.

The amplitude response $A_{n,o}(x,y)$ and the phase response $\phi_{n,o}(x,y)$ of the image $I(x,y)$ at scale n and orientation o are given by

$$A_{n,o}(x,y) = \sqrt{E_{n,o}(x,y)^2 + O_{n,o}(x,y)^2} \tag{4}$$

$$\phi_{n,o}(x,y) = \arctan\left(\frac{E_{n,o}(x,y)}{O_{n,o}(x,y)}\right). \tag{5}$$

PC at point (x,y) is calculated as the ratio of weighted and noise compensated local energy summed over all the orientations to the total sum of filter response amplitudes overall orientations and amplitudes, and are expressed as

$$\text{PC}(x,y) = \frac{\sum_n \sum_o W(x,y)\lfloor A_{n,o}(x,y)\Delta\phi_{n,o}(x,y) - T\rfloor}{\sum_n \sum_o A_{n,o}(x,y) + \varepsilon} \tag{6}$$

$$\begin{aligned}&\Delta\phi_{n,o}(x,y)\\ &= \cos\left(\phi_{n,o}(x,y) - \overline{\phi}(x,y) - \left|\sin\left(\phi_{n,o}(x,y) - \overline{\phi}(x,y)\right)\right|\right)\end{aligned} \tag{7}$$

where $\text{PC}(x,y)$ is the magnitude of the PC. $W_o(x,y)$ is a weight function. $A_{n,o}(x,y)$ is the amplitude component at the filter scale n and direction o at (x,y). $\overline{\phi}(x,y)$ is the weighted mean phase. T is a noise threshold and ε is a small constant to avoid division by zero. The symbols $\lfloor\rfloor$ denotes that the enclosed quantity is equal to itself when its value is positive, or zero otherwise. The value of PC is a dimensionless quantity within 0 and 1, enhance visualization to express structure features by the following formula:

$$\text{PC}(x,y) = \frac{(\text{PC}(x,y) - \min(\text{PC}(x,y)))}{\max(\text{PC}(x,y)) - \min(\text{PC}(x,y))}. \tag{8}$$

Using (6) and (8), we can achieve the PC map of images, and are used to detect the feature points in Section III-B.

B. Salient Feature Point Detection

According to the principle of PC, edges structure features are perceived at the point where the Fourier components are in phase with each other. The PC map is more like the enhanced edges of an image [29]. In view of the characteristic of the PC map, the nonlinear diffusion function is used to construct the nonlinear scale space of PC maps, and the extreme value is detected in the nonlinear scale space to obtain the feature points with invariant to illumination and contrast. Compared with the Gaussian function, the nonlinear spread function has proven to preserve the edges structure and details better, as well as improving the localization accuracy of features while suppressing noise [28], [58]. The nonlinear diffusion function can be expressed as

$$\frac{\partial L}{\partial t} = \text{div}\left(c(x,y,t) \cdot \nabla L\right) \tag{9}$$

where div and ∇ are the divergence and gradient operators, respectively. L is the luminance of an image, $c(x, y, t)$ is the conductivity function and t is the scale parameter. Through setting proper $c(x, y, t)$, we can make the diffusion adaptive to the local structure. The function $c(x, y, t)$ is defined as

$$c(x, y, t) = g\left(\nabla L_\sigma(x, y, t)\right) \tag{10}$$

where ∇L_σ is the gradient of a Gaussian smoothed original image L and σ is representative of the amount of blur. In order to preserve the edge of the image while ensuring a faster diffusion rate [58], the following formula is selected as g:

$$g = \frac{1}{1 + \frac{|\nabla L_\sigma|^2}{\kappa^2}} \tag{11}$$

where κ is the contrast factor that controls the level of diffusion. The greater the κ value chosen, the less edge information will be preserved.

Similar to SIFT [14], the constructed nonlinear scale space is divided into O octaves, and each octave contains S sublevel. The resolution of all layers in the scale space is the same as the original image. The scale relationships among the layers are described as follows:

$$\sigma_i(o, s) = \sigma_0 2^{o+\frac{s}{S}}$$
$$o \in [0, ..., O-1], s \in [0, ..., S-1], i \in [0, ..., N] \tag{12}$$

where o and s represent the index of octave O and sublevel S, respectively. σ_0 is the initial value of the scale parameter. N is the total number of nonlinearly filtered images. To capture more salient feature points, the maximum octave is set to 6, the number of sublevels per scale level is set to 4, and the control factor κ is set to empirically be 0.0003. Having constructed the nonlinear scale space, we calculate the scale-normalized Hessian matrix in the scale space. The 2-D feature points are located by calculating the local maximum of the 3×3 neighborhood in the scale-normalized Hessian matrix.

Due to the nonlinear diffusion filters preserve more of the PC maps information, the number of feature points obtained by this method is much more than that of the traditional methods, which get feature points on original gray images. This creates many noncommon feature points, and matching of noncommon feature points of images can only bring in the wrong matches, these unreliable matches will bring down the matching performance. Considering PC not only describes the features of image, but also reflects the significance of local structural features [48], we have attempted to introduce PC to remove the noise or unreliable feature points. If the PC information at a feature point (denoted by P_i) is larger than a threshold, i.e., $P_i \geq$ th, then the point is validated. If P_i is below th, then the point is rejected, where th is a preset threshold, and empirical value of 0.05 is suggested. In this way, many impossible matching points are eliminated to avoid unnecessary calculations.

A simple comparison in Fig. 3 is conducted to illustrate the robustness of the proposed method, whereas other methods are used as contrast algorithms to detect the features. One can see that the proposed method efficiently extracts a larger number of distinctive feature points, which illustrates that our method is more robust to nonlinear intensity changes.

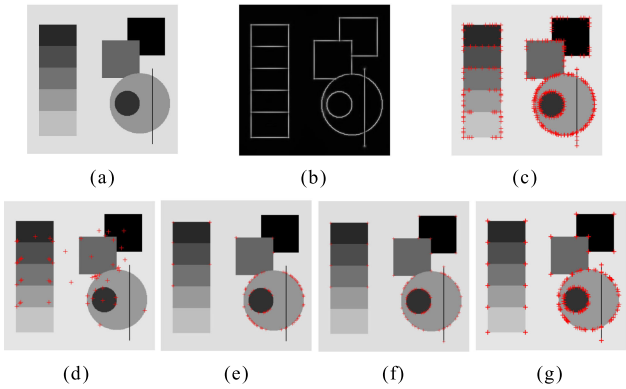


Fig. 3. Feature points detection method comparison. (a) Image with nonlinear intensity changes. (b) PC map. (c) MMSC. (d) SIFT. (e) FAST. (f) Harris. (g) PC detector (PC minimum moment is 0.2).

C. Feature Descriptor Construction

Although the intensity difference is in the various spectral band images, the overall structure and shape features of the scenes maintain a certain degree of similarity. The proposed method leverages the orientation of the maximum mean log-Gabor response to establish feature descriptor based on the global structure of images. The log-Gabor convolution response was obtained in the PC maps calculation stage. The LGHD descriptor leverages all scales and orientation log-Gabor response to calculate the distribution histograms of 4×4 subregions to describe the feature points. Different from their works, we use the mean log-Gabor response of different scales in the same direction to calculate the distribution to increase the distinctiveness and robustness of the feature descriptor against significant intensity variations and decreases the computation complexities of the subsequent processing tasks (e.g., feature matching). For each orientation o, the mean magnitudes of all scales are calculated as

$$A_o(x, y) = \frac{\left(\sum_{n=1}^{N_s} A_{n,o}(x, y)\right)}{N_s} \tag{13}$$

where $o = 1, 2, ..., N_o$ and $s = 1, 2, ..., N_s$ stand for the number of orientations and scales, respectively. A_o is defined as the oriented amplitudes map of orientation o.

Among these different orientations of mean response map, we use the index of the maximum mean response map instead of the maximum mean response map value to create the feature descriptor. Fig. 4 shows the process of constructing the feature description. Concretely, for each feature point, we select a local region with S $\times$ S pixels center at the feature. Based on the MPEG-7 standard, the local area is divided into 5×5 subregions, and build the block distribution histogram by statistics the index value of maximum oriented amplitudes map in all subregions. The feature descriptor is obtained by incorporating the distribution histograms of every block and normalized to unit length.

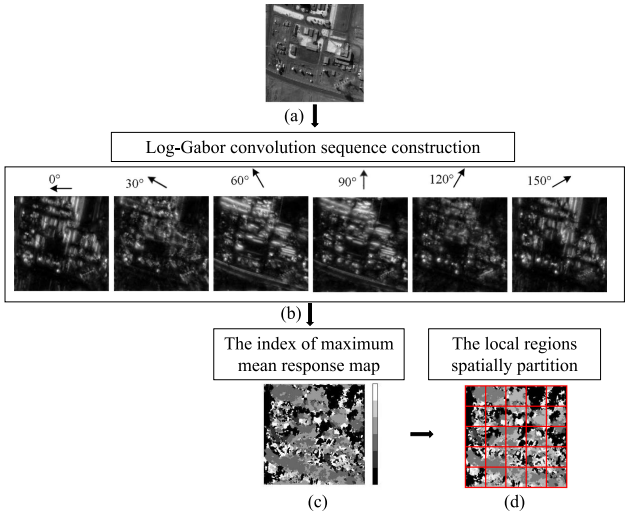


Fig. 4. Process of descriptor construction. (a) Input local region. (b) Average log-Gabor response map in multiple directions. (c) Generate the maximum response orientation index map. (d) Spatial structure division base on the MPEG-7 standard.

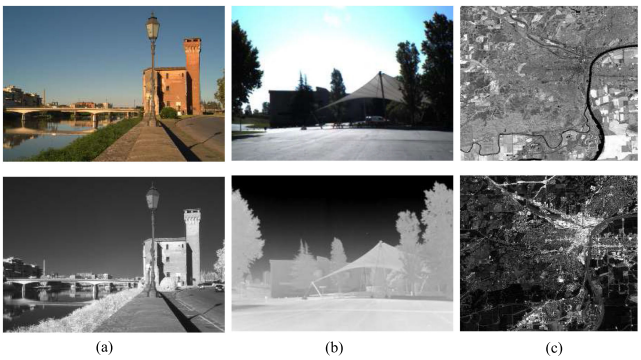


Fig. 5. Examples of multispectral images from different datasets. (a) VIS-NIR dataset. (b) VIS-TIR dataset. (c) SAT dataset.

The matching process is achieved using the bilateral matching based on the unilateral best-bin-first (BBF) method, regard the feature point pairs with minimal SAD distance as potential matches [38].

IV. Experimental Results and Analyses

A. Datasets and Evaluation Criteria

To verify the matching performance of the proposed method, three different types of multispectral image datasets were selected for qualitative and quantitative evaluation. Samples from these datasets are shown in Fig. 5. Fig. 5(a) come from VIS-NIR datasets [1]. This dataset consists of 477 visible and infrared scene image pairs, which include 9 categories as follows: country, field, forest, mountain, old building, street, urban, and water. In our experiments, we select 90 images for testing (top 10 per category). Fig. 5(b) is from the VIS-TIR datasets [20]. This dataset consists of 44 visible and thermal infrared outdoor image pairs, with dimensions of 639×431 pixels. The average spectral distance is approximately 0.5 μm for VIS-NIR datasets and 9 μm

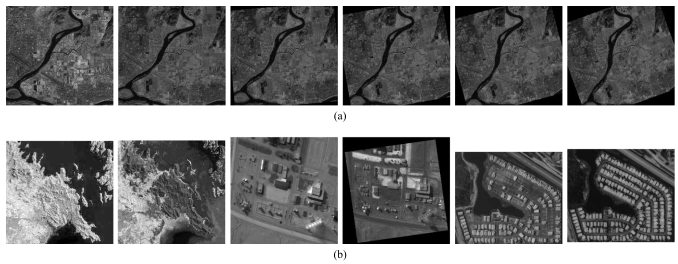


Fig. 6. Samples of the fourth dataset. (a) Dataset with different levels of rotation changes. (b) Samples of multispectral image pairs in the fourth datasets.

for VIS-TIR datasets, which means VIS-TIR datasets are more challenging than VIS-NIR datasets. Fig. 5(c) is from the SAT dataset. The SAT dataset contains four pairs different spectral of remote sensing satellite imagery. Each image pair was composed of a reference and a sensed image of the same area, which were, respectively, acquired from different spectra resulting in relatively high-intensity changes. All the images were rectified and aligned, so that matches could be obtained in horizontal lines. There is no additional preprocessing or enhancement to highlight features or increase in contrast.

In addition, the fourth dataset was used for experiments to demonstrate the robustness of the proposed method under different interferences and the practicability of image registration. This dataset consists of two group simulated images and six pairs of multispectral/multisensor remote sensing images with different degrees geometric and radiometric variations, where the two group simulated images are simulated images with different levels of Rotation angle and scales, respectively. These images are displayed in Fig. 6. For example, the simulated images for different levels of applied rotation transformations are illustrated in Fig. 6(a), and Fig. 6(b) presents illustrations of sample image pairs from six pairs of multispectral/multisensor remote sensing images. For better quantitative evaluation, we need to obtain a ground truth geometric transformation between each image pair. In the two group simulated images, the known predefined projective model between image pairs is used for this purpose. The six pairs multispectral/multisensor remote sensing images affine transformation parameters were calculated by manually selecting evenly distributed ten pairs of corresponding control points. The source code of MMSC and these datasets along with detailed data descriptions are available.[1]

The performance of the MMSC method is evaluated using four matching criteria named repeatability, precision, the number of correct matches (NCM), and root-mean-square error (RMSE).

Feature points repeatability depicts the percentage of repeatable features detected in reference and sensed images are used to assess the performance of the detectors. It is defined as

$$\text{Repeatability} = \frac{\text{Correspondences}}{(n_1 + n_2)/2} \tag{14}$$

where n_1 and n_1 are the number of feature points in the reference and sensed images, respectively. Correspondences are the

[1][Online].Availabe: https://github.com/liruixiang00/mmsc.

TABLE I
AVERAGE PRECISION AND NCM VALUES OF DIFFERENT PARAMETER N_s

Metrics	N_s = 2	N_s = 3	N_s = 4	N_s = 5	N_s = 6
Precision	0.2378	0.2945	0.3202	0.2716	0.2975
NCM	77.1	106.2	105.9	88.7	65.2

TABLE II
AVERAGE PRECISION AND NCM VALUES OF DIFFERENT PARAMETER N_o

Metrics	N_o = 3	N_o = 4	N_o = 5	N_o = 6	N_o = 7
Precision	0.2504	0.2744	0.2936	0.3202	0.2976
NCM	99.23	102.3	107.1	105.9	102.6

number of corresponding feature point pairs whose reprojection error is less than three pixels. The reprojection error $thre$ of two feature points of x_1 and x_2 can be expressed by the following equation:

$$\|x_1 - Tx_2\| = thre \tag{15}$$

where T is the geometric transformation parameters between reference and sensed images. The definitions of precision and NCM are given as follows [59]:

$$\text{Precision} = \frac{\text{NCM}}{\text{NTM}} \tag{16}$$

where NCM and NTM are the NCMs and the total number of matches in the matching results, respectively.

In addition to these measures, the RMSE was used to evaluate the registration accuracy quantitatively, and is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_i'')^2 + (y_i - y_i'')^2} \tag{17}$$

where N is the number of final matches, which denotes the matched points which have been filtered by fast sample consensus (FSC) as inliers. (x_i, y_i) and (x_i'', y_i'') are the coordinates of pixels in reference image and the transformed sensed image, respectively.

B. Parameter Analysis

The proposed MMSC method has three key parameters: N_s, N_o, and S. Parameters N_s and N_o are the number of scales and orientations of the log-Gabor filter, respectively. The combination of parameters of log-Gabor filters could influence the performance of the proposed MMSC method. Parameter S is the region size for feature description, which is a key factor in the performance of the descriptor. We evaluate the performance of MMSC when its three parameters are set to different values. For evaluation, the values assigned to N_s parameter were changed from 2 to 6 and the values assigned to N_o parameter were changed from 3 to 7. For set S, two different combinations, including 100 × 100 and 80 × 80 sets, were considered. For each combination of three key parameters, the matching process was performed on the VIS-TIR dataset and the average of the precision and NCM values was estimated. The experimental results are listed in Table I–Table III. From Table I, when N_s reaches 4, the precision of MMSC reaches 32.02%. From Table II, the results for each selected N_o indicate that the larger N_o value until $N_o = 6$ provides better matching results. As shown in Table III, When $S = 80$, MMSC achieves the best performance in both precision and NCM metrics. Taking into consideration matching performance and descriptor length, these parameters are fixed to $N_o = 6$, $N_s = 4$, and $S = 80$ in the following experiments. Thus, the dimension of the feature description is 150, which decreases the computation complexities of the subsequent matching.

TABLE III
AVERAGE PRECISION AND NCM VALUES OF DIFFERENT PARAMETER S

Metrics	S = 80	S = 100
Precision	0.3202	0.2995
NCM	105.9	104.9

TABLE IV
AVERAGE REPEATABILITY AND CORRESPONDENCES ACHIEVED BY FEATURE POINT DETECTORS

Detectors	VIS-NIR		VIS-TIR		SAT		Mean	
	Rep	Corr	Rep	Corr	Rep	Corr	Rep	Corr
Harris	0.5195	790.0	0.2394	82.6	0.2086	225.5	0.3225	366.0
SIFT	0.3244	795.2	0.1312	135.9	0.1541	183.5	0.2032	371.5
SURF	0.4559	980.4	0.1626	152.3	0.2615	448.7	0.2933	527.1
FAST	0.4998	1112.5	0.1750	148.7	0.3220	595.3	0.3326	618.8
MMSC	0.5702	1425.6	0.2422	579.0	0.3569	891	0.3931	965.2

Note: Rep = repeatability rate and Corr = the number of correspondences.

C. Analysis of Detector performance

In this part, to evaluate the performance of the proposed MMSC feature detection, a comparison of it with four feature point detectors is made, which are Harris [13], SIFT [14], SURF [15], and FAST [31]. We conduct our experiments on three multispectral datasets shown in Table IV. Here, the evaluation metrics used repeatability and correspondences. Table III lists the comparison results. The mean column lists the average values computed over all the evaluated datasets. The comparison shows that the proposed MMSC detector obtains on average the best performance. In the VIS-NIR dataset, FAST and Harris have similar repetition rates, but FAST obtains more matching correspondences. Since the thermal infrared images have less common information than the near infrared images, all the detectors have low repeatability on the VIS-TIR dataset. Harris demonstrates better performance on VIS-NIR and VIS-TIR dataset, whereas it performs lower to FAST and SURF on SAT dataset. Compared with other methods, MMSC detector achieves the best repeatability in this dataset, followed by the FAST and Harris detector. This is because the PC is more robust than gradient and intensity information to nonlinear intensity differences. In addition, we introduce structure consistency constraint retain salient feature points to improve repeatability. SIFT obtains overall the worst repeatability in the comparison.

D. Analysis of Matching Performance

To verify the matching performance of the proposed MMSC method, we select three multispectral datasets for qualitative and quantitative evaluation. We compare our MMSC algorithm against five state-of-the-art algorithms, i.e., SIFT [14],

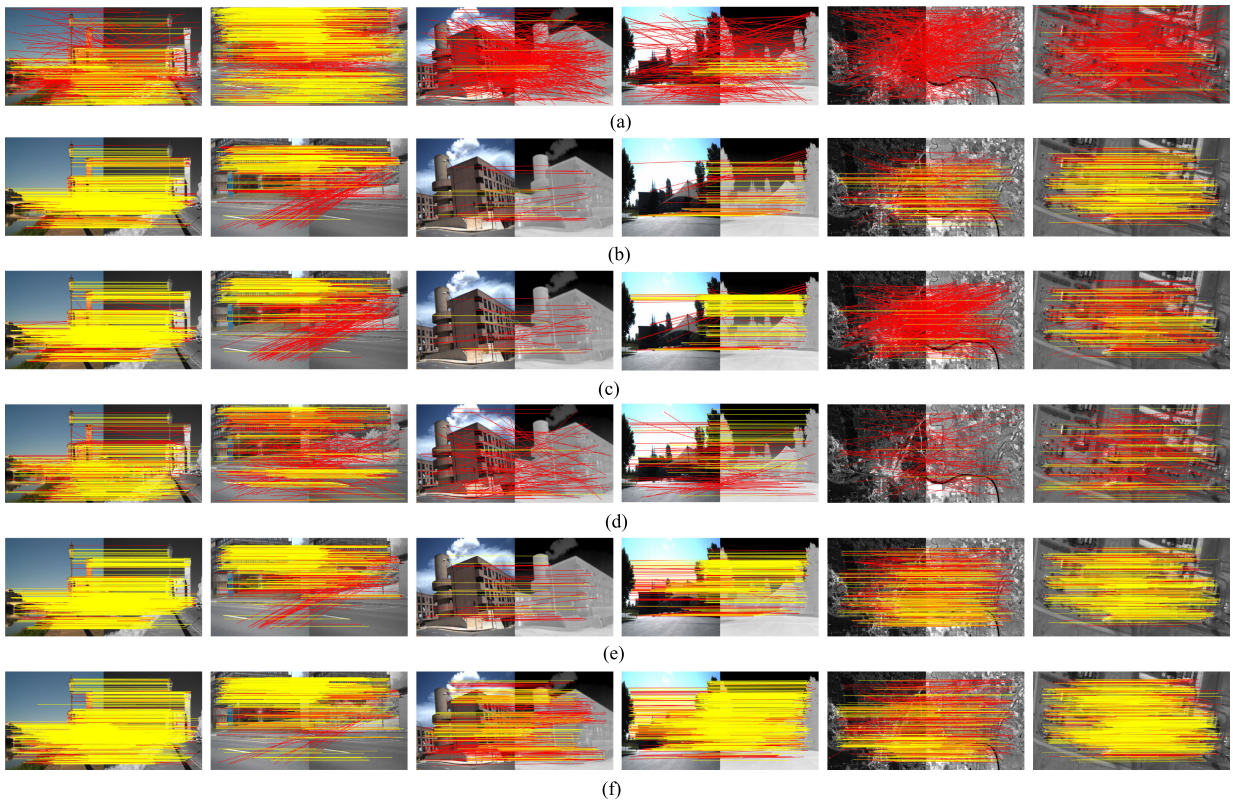


Fig. 7. Comparison of matching performance by the related methods. (a) SIFT. (b) EHD. (c) PCEHD. (d) PIIFD. (e) LGHD. (f) MMSC.

EHD [42], PCEHD [43], PIIFD [38], and LGHD [20]. To be specific, the EHD, PCEHD, and LGHD used the FAST detector to detection feature point, the PIIFD used the SURF detector to detection feature point, the SIFT leveraged the SIFT detector to detection feature point, and then use their descriptor for feature matching. We detected a total of 2500 points for each image of multispectral images. However, image texture and scene content will affect the number of feature points finally detected. The bilateral BBF method was adopted and the threshold of the nearest neighbor criterion of all the methods is set to 1 in this article [38]. The greater the threshold value chosen, the more the NCM will be preserved.

1) Qualitative Results: Fig. 7 shows the matching results of SIFT, EHD, PCEHD, PIIFD, LGHD, and MMSC on different multispectral datasets successively. Red lines indicate wrong matches and yellow lines indicate the correct. The first two columns are the NIR image pairs, which suffer from the large nonlinear intensity change will cause some physical correspondences that are not similar at all. The next two columns are the TIR image pairs, they come from different sensors and during the imaging process, will lose local texture information. The last two columns, they are the SAT image pairs in which the challenge is severe nonlinear intensity differences due to the difference in imaging mechanism.

As shown in Fig 7(a), SIFT algorithm fails to match on the fifth image pairs. Even if the matching is successful, the precision is also low on the third, fourth, and sixth pairs of images. Although PIIFD establishes symmetric feature descriptions to overcome the gradient orientation reversal, nonlinear intensity change and local detail discrepancies can also cause the dissimilarity of feature descriptions. This shows that gradient-based descriptors are more sensitive to nonlinear grayscale differences. The results of EHD and PCEHD are similar, and they are better than the results of the gradient-based descriptor. LGHD only obtains a higher precision on the NIR image pair. In contrast, the proposed MMSC algorithm has the best matching performance on all six image pairs, whose precision is 78.21%, 63.91%, 32.70%, 61.88%, 35.79%, and 64.13% successively. The average precision of MMSC method for the six image pairs are 56.11%, which is 7.30% higher than that of LGHD. The average precision of MMSC is about 3.2 times that of SIFT. This is because the MMSC uses PC instead of image intensity for feature point detection, considers both the number and repeatability of feature points. In addition, the MMSC is constructed from the log-Gabor convolution sequence and is much more robust to nonlinear intensity differences than traditional gradient map. Thus, MMSC not only largely improves the stability of feature detection but also overcomes the limitation of gradient information for feature description.

2) Quantitative Results: The quantitative results of all methods on different multispectral datasets are shown in Table V. In the VIS-NIR dataset, the spectra of the image pairs are close. EHD and PCEHD use the responses of multioriented Sobel filters, and they are better than the results of the SIFT and PIIFD. The LGHD descriptor uses multiscales and multidirectional log-Gabor filters to take the place of the five filters of EHD, which can obtain more stable structural properties of multispectral images than EHD. MMSC match precision is slightly higher

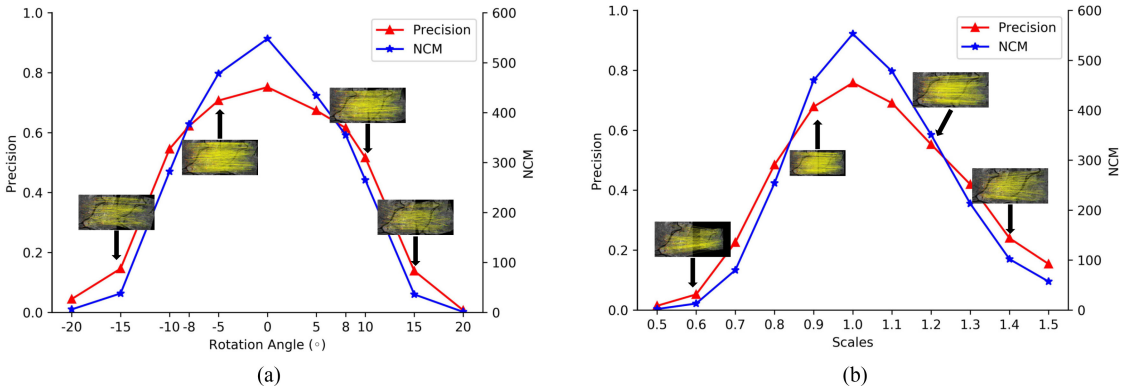


Fig. 8. Matching results of different interferences to MMSC. (a) Rotation change. (b) Scale change.

TABLE V
MATCHING PERFORMANCE OF THE RELATED METHODS

Methods	VIS-NIR		VIS-TIR		SAT	
	Precision	NCM	Precision	NCM	Precision	NCM
SIFT	0.4861	547.5	0.0249	6.3	0.0234	10.0
EHD	0.6891	637.0	0.1748	7.2	0.3728	101.5
PCEHD	0.7187	555.7	0.1416	12.6	0.1051	21.25
PIIFD	0.6077	504.4	0.0941	9.4	0.1519	35.5
LGHD	0.7751	630.0	0.2886	34.2	0.3941	155.5
MMSC	0.7805	748.8	0.3202	105.9	0.4544	248.7

than LGHD. The average NCM of MMSC method in this are 748.8, which is 118.8, 201.3 more than that of LGHD and SIFT, respectively. This is because in the proposed algorithm, the adoption of nonlinear feature points results in significant improve of the number of correctly matched points and the matching performance. In addition, the proposed descriptor describes the feature vectors by the distribution of the maximum mean log-Gabor response, which can capture the more robust structure and shape characteristics of multispectral images.

Compared to NIR images in the VIS-NIR datasets, the TIR images in the VIS-TIR dataset have greater spectral range while losing local texture information with corresponding visible images. Hence, all matching method obtained lower precision. However, the proposed MMSC method performs much better than the other method. EHD leverage multioriented Sobel spatial filter construction edge descriptors, which show better results than gradient-based methods. Although the similarity of image structures decreases as the spectral difference increases, image structures still maintain better consistency than image intensities and gradients. Because the log-Gabor filters are better at retaining the oriented edge characteristics of multispectral images, the LGHD can achieve better matching performance than the EOH. When the local texture of the thermal infrared image is losing, the structural information maintains a certain degree of similarity in visible and thermal infrared images. Using the structural consistency information of the image for feature detection and description makes MMSC have the best matching performance. Comparing our method with LGHD, it can be seen that the precision improves from 28.86% to 32.02% and NCM improves from 34.2 to 105.9.

The spectral ranges of the SAT dataset are much closer than the visible and thermal infrared images of the VIS-TIR. Among the other method, the LGHD and EHD descriptor performs much better than the remaining descriptors, and present similar results. Gradient-based descriptor, including SIFT and PIIFD, showed similar results. They can process images pair with less difference, but cannot get correct matching for multispectral images with great differences. Even if the matching is successful, the NCMs are also small, as shown in Fig. 7(a) and (d). Although the LGHD algorithm achieves second-best matching performance, there are very few NCMs compared with the proposed method, as shown in Fig. 7(e). In the MMSC method, a PC map instead of image intensity for feature detection. Using nonlinear diffusion to detect feature points on the PC maps can improve the repeatability of feature points, much more correct matches can be obtained. Furthermore, our method adopts the average log-Gabor convolution sequence to construct the distribution histogram of maximum log-Gabor response, which make the robustness to intensity changes of the proposed matching method.

E. Influence of Different Interferences to MMSC

To investigate the influence of rotation and scale changes, we test the proposed method on two group simulated images with different scale factors and rotation angle. While the reference image of Fig. 8(a) remains unchanged, the sensed image of Fig. 6(a) is rotated from -20° to 20° with different levels of applied rotation transformations. Similarly, the sensed image of Fig. 6(a) is change from 0.5 to 1.5 scales with different levels of applied scale transformations. Fig. 8 indicates the experiment results in terms of precision and NCM values for MMSC in different interferences. As can be seen, the capability of the proposed method degraded with increasing in the geometric difference level. Therefore, the proposed algorithm is not very effective in directly matching images with large geometric changes. On the other hand, it is shown that MMPC can tolerate rotation change is less than 10°, as shown in Fig. 8(a). It is also revealed that the scale variations that can be tolerated are between 0.8 and 1.2, as shown in Fig. 8(b).

F. Analysis of Registration Performance

In this part, to validate the proposed matching method in image-registration applications, six pairs of images shown in Fig. 6(b) are used for evaluation. We compare our MMSC

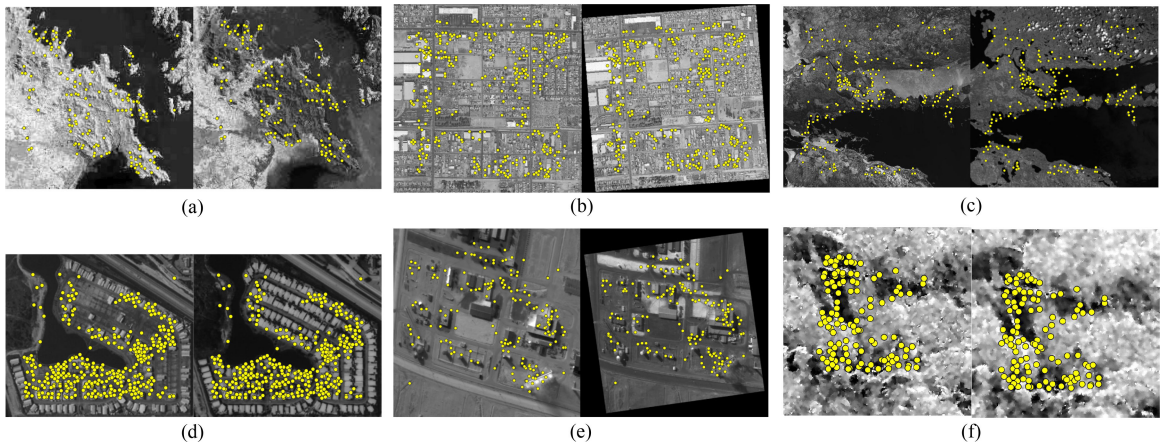


Fig. 9. Detected matched features of the proposed method. (a)–(f) First to sixth pairs of multispectral images in the fourth dataset.

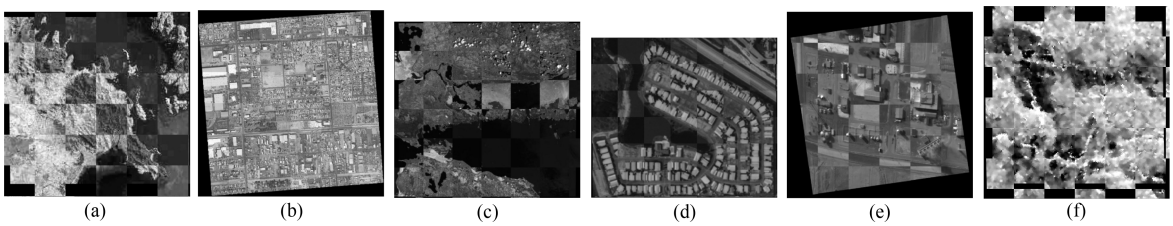


Fig. 10. Checkerboard mosaiced images of the proposed method. (a)–(f) First to sixth pairs of multispectral images in the fourth dataset.

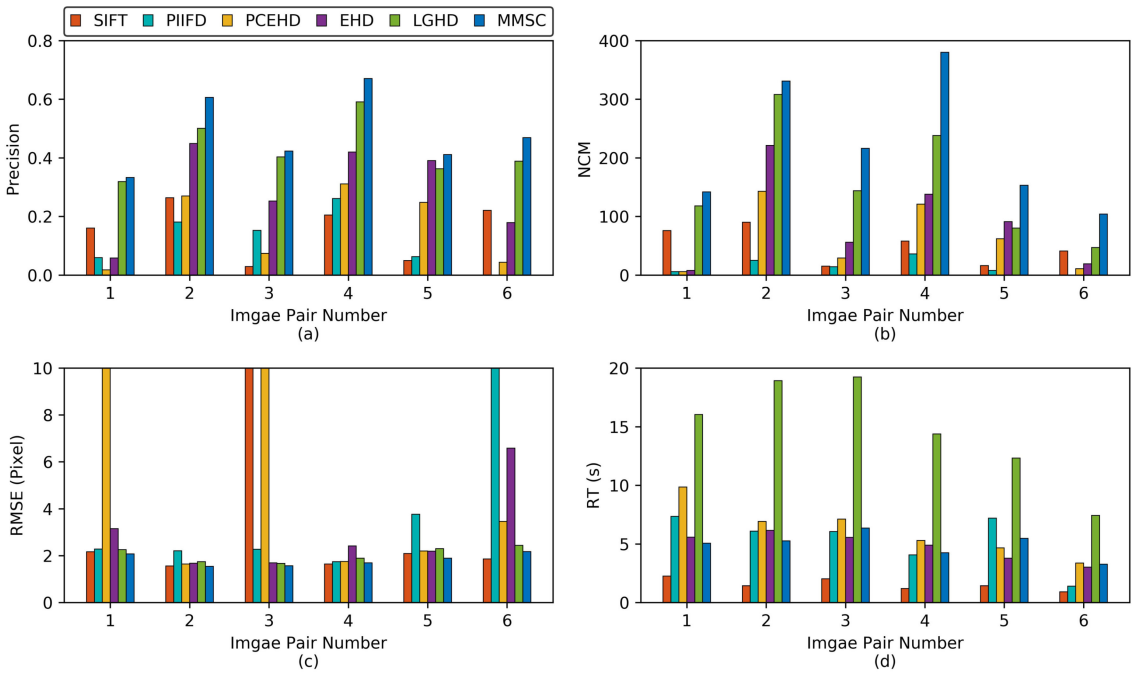


Fig. 11. Registration results for different methods for six pair images in terms of (a) precision, (b) NCM, (c) RMSE, and (d) RT.

algorithm with SIFT, PIIFD, EHD, PCEHD, and LGHD. We first perform feature matching on these image pairs. Then, we use the FSC algorithm [60] to remove the outliers and to estimate the transformation parameters.

The registration results obtained by the proposed method are shown in Fig. 9. In addition, the checkerboard mosaiced images are provided for visual inspection of the registration results in Fig. 10. The comparative results for different descriptors, and for all evaluation criteria, such as precision, the NCM, the registration accuracy (RMSE), and run time (RT), are shown in Fig. 11. As can be seen from Fig. 11(a) and (b), the proposed method outperforms the other methods both in precision and

NCM. This may be because the MMSC uses PC instead of image intensity for feature point detection, and considers both the number and repeatability of feature points. In addition, the MMSC adopts the orientation of the maximum mean log-Gabor filter responses, which are encoded to build feature descriptors instead of simple gradients or similarity measures. LGHD achieves the second-best matching precision. Similar to the MMSC, the LGHD is also based on log-Gabor filters and is robust against nonlinear intensity changes. However, the NCMs by the MMSC algorithm are approximately twice. As can be seen in Fig. 9, the matched features demonstrate the capability of the proposed MMSC method in detecting a sufficient number match point pairs even in multispectral image pairs with geometric and significant nonlinear intensity differences. The RMSE of the registration results of using different methods is given in Fig. 11(c). We can see that the PCEHD algorithm failed in the first and third image pairs, SIFT algorithm failed in the third image pairs, and PIIFD algorithm failed in the sixth image pairs. A possible reason is that they could not get enough correct matched point correspondences. The NCMs obtained by our method are from two to ten times as many as other methods. The average RMSE of MMSC over all image pairs is 1.78 pixels. Experimental results show that our method is robust against the nonlinear intensity differences, and outperforms the comparison methods in terms of the NCM and registration accuracy.

All comparative experiments are implemented on a laptop with 1.8-GHz Intel Core i5 CPU, 8-GB RAM, and MATLAB, and the running times required by the related methods for the six image pairs are shown in Fig. 11(d). As can be seen, the MMSC spends a moderate level of computation time among compared methods, SIFT achieves the best performance, and LGHD is featured with a high computation time. This is because the LGHD descriptor uses all the scale and oriented magnitudes of the log-Gabor filters to calculate the feature vectors. The MMSC descriptor is composed of 150-dimensional vectors, which saves a lot of time in the feature description and feature matching process, as compared with LGHD.

V. Conclusion

In this article, a multispectral image MMSC was proposed. First, the salient feature points were detected from PC maps based on PC and nonlinear diffusion, which were invariant to illumination and contrast. Subsequently, the orientation of the maximum log-Gabor responses was used to represent the structure features of multispectral images. Feature descriptors were established according to the orientation of the maximum log-Gabor responses. In the experimental part, three multispectral datasets were employed for matching test and several state-of-the-art methods were used for comparison. The experimental results showed that our method obtained superior matching performance over the others. However, when calculating the PC maps, it requires relatively higher computation. In the future, we will study how to reduce the algorithm complexity and the eliminate outliers.

References

[1] M. Brown and S. Süsstrunk, "Multi-spectral sift for scene category recognition," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 177–184.

[2] J. Nie, S. Qu, Y. Wei, L. Zhang, and L. Deng, "An infrared small target detection method based on multiscale local homogeneity measure," *Infrared Phys. Technol.*, vol. 90, pp. 186–194, 2018.

[3] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, Jul. 2019.

[4] D. Firmenichy, M. Brown, and S. Süsstrunk, "Multispectral interest points for RGB-NIR image registration," in *Proc. 18th IEEE Int. Conf. Image Process.*, 2011, pp. 181–184.

[5] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, 2019.

[6] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3015157.

[7] S. Cao, X. Zhu, Y. Pan, and Q. Yu, "A stable land cover patches method for automatic registration of multitemporal remote sensing images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3502–3512, Aug. 2014.

[8] 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, 2014.

[9] M. Lecca and M. D. Mura, "Intensity and affine invariant statistic-based image matching," in *Proc. Comput. Model. Objects Represented Images*, 2012, pp. 85–90.

[10] J. P. Kern and M. S. Pattichis, "Robust multispectral image registration using mutual-information models," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1494–1505, May 2007.

[11] 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.

[12] K. Yu, J. Ma, F. Hu, T. Ma, S. Quan, and B. Fang, "A grayscale weight with window algorithm for infrared and visible image registration," *Infrared Phys. Technol.*, vol. 99, pp. 178–186, 2019.

[13] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, vol. 15, no. 50, pp. 10–5244.

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

[15] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.

[16] Y. P. Kwon, H. Kim, G. Konjevod, and S. McMains, "Dude (duality descriptor): A robust descriptor for disparate images using line segment duality," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 310–314.

[17] H. Goncalves, L. Corte-Real, and J. A. Goncalves, "Automatic image registration through image segmentation and sift," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2589–2600, Jul. 2011.

[18] S.-J. Chen, H.-L. Shen, C. Li, and J. H. Xin, "Normalized total gradient: A new measure for multispectral image registration," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1297–1310, Mar. 2018.

[19] C. Aguilera, F. Barrera, F. Lumbreras, A. D. Sappa, and R. Toledo, "Multispectral image feature points," *Sensors*, vol. 12, no. 9, pp. 12661–12672, 2012.

[20] C. A. Aguilera, A. D. Sappa, and R. Toledo, "LGHD: A feature descriptor for matching across non-linear intensity variations," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 178–181.

[21] Q. Li, S. Qi, Y. Shen, D. Ni, H. Zhang, and T. Wang, "Multispectral image alignment with nonlinear scale-invariant keypoint and enhanced local feature matrix," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1551–1555, Jul. 2015.

[22] J. Jin and M. Hao, "Registration of UAV images using improved structural shape similarity based on mathematical morphology and phase congruency," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1503–1514, Apr. 2020.

[23] S. Saleem and R. Sablatnig, "A robust sift descriptor for multispectral images," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 400–403, Apr. 2014.

[24] S.-Y. Cao, H.-L. Shen, S.-J. Chen, and C. Li, "Boosting structure consistency for multispectral and multimodal image registration," *IEEE Trans. Image Process.*, vol. 29, pp. 5147–5162, Mar. 2020.

[25] 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, 2018.

[26] P. Kovesi, "Phase congruency detects corners and edges," in *Proc. Aust. Pattern Recognit. Soc. Conf., DICTA*, 2003, vol. 2003, pp. 309–318.
[27] W. Ma, Y. Wu, S. Liu, Q. Su, and Y. Zhong, "Remote sensing image registration based on phase congruency feature detection and spatial constraint matching," *IEEE Access*, vol. 6, pp. 77554–77567, Dec. 2018.
[28] 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.
[29] 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.
[30] S. Saleem, A. Bais, R. Sablatnig, A. Ahmad, and N. Naseer, "Feature points for multisensor images," *Comput. Elect. Eng.*, vol. 62, pp. 511–523, 2017.
[31] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, Jan. 2010.
[32] J. Fan, Y. Wu, M. Li, W. Liang, and Y. Cao, "SAR and optical image registration using nonlinear diffusion and phase congruency structural descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5368–5379, Sep. 2018.
[33] W. Ma, Y. Wu, Y. Zheng, Z. Wen, and L. Liu, "Remote sensing image registration based on multifeature and region division," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1680–1684, Oct. 2017.
[34] Z. Zhang, Q. Zou, Y. Lin, L. Chen, and S. Wang, "Improved deep hashing with soft pairwise similarity for multi-label image retrieval," *IEEE Trans. Multimedia*, vol. 22, no. 2, pp. 540–553, Feb. 2020.
[35] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 778–792.
[36] M. F. Vural, Y. Yardimci, and A. Temlzel, "Registration of multispectral satellite images with orientation-restricted sift," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2009, vol. 3, pp. III-243–III-246.
[37] D. Zhao, Y. Yang, Z. Ji, and X. Hu, "Rapid multimodality registration based on MM-SURF," *Neurocomputing*, vol. 131, pp. 87–97, 2014.
[38] G. Wang, Z. Wang, Y. Chen, and W. Zhao, "Robust point matching method for multimodal retinal image registration," *Biomed. Signal Process. Control*, vol. 19, pp. 68–76, 2015.
[39] C. Zhao, H. Zhao, J. Lv, S. Sun, and B. Li, "Multimodal image matching based on multimodality robust line segment descriptor," *Neurocomputing*, vol. 177, pp. 290–303, 2016.
[40] Q. Guo, M. He, and A. Li, "High-resolution remote-sensing image registration based on angle matching of edge point features," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2881–2895, Aug. 2018.
[41] Q. Zou, H. Jiang, Q. Dai, Y. Yue, L. Chen, and Q. Wang, "Robust lane detection from continuous driving scenes using deep neural networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 41–54, Jan. 2020.
[42] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
[43] T. Mouats and N. Aouf, "Multimodal stereo correspondence based on phase congruency and edge histogram descriptor," in *Proc. IEEE 16th Int. Conf. Inf. Fusion*, 2013, pp. 1981–1987.
[44] H. K. Ragb and V. K. Asari, "Chromatic domain phase features with gradient and texture for efficient human detection," *Electron. Imag.*, vol. 2017, no. 10, pp. 74–79, 2017.
[45] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "DeepCrack: Learning hierarchical convolutional features for crack detection," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1498–1512, Mar. 2019.
[46] J. Arrospide and L. Salgado, "Log-Gabor filters for image-based vehicle verification," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2286–2295, Jun. 2013.
[47] C. F. Nunes and F. L. Pádua, "A local feature descriptor based on log-Gabor filters for keypoint matching in multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1850–1854, Oct. 2017.
[48] X. Liu, Y. Ai, B. Tian, and D. Cao, "Robust and fast registration of infrared and visible images for electro-optical pod," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1335–1344, Feb. 2019.
[49] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3016820.
[50] L. Gao, D. Yao, Q. Li, L. Zhuang, B. Zhang, and J. M. Bioucas-Dias, "A new low-rank representation based hyperspectral image denoising method for mineral mapping," *Remote Sens.*, vol. 9, no. 11, pp. 1145–1165, 2017.
[51] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 193–205, 2019.
[52] L. Zhang, Z. Shao, J. Liu, and Q. Cheng, "Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data," *Remote Sens.*, vol. 11, no. 12, 2019, Art. no. 1459.
[53] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.
[54] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 148–164, 2018.
[55] H. He, M. Chen, T. Chen, and D. Li, "Matching of remote sensing images with complex background variations via Siamese convolutional neural network," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 355.
[56] H. Zhang *et al.*, "Registration of multimodal remote sensing image based on deep fully convolutional neural network," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3028–3042, Aug. 2019.
[57] J. Zhang, W. Ma, Y. Wu, and L. Jiao, "Multimodal remote sensing image registration based on image transfer and local features," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1210–1214, Aug. 2019.
[58] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 214–227.
[59] Z. Shao, K. Yang, and W. Zhou, "Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset," *Remote Sens.*, vol. 10, no. 6, 2018, Art. no. 964.
[60] Y. Wu, W. Ma, M. Gong, L. Su, and L. Jiao, "A novel point-matching algorithm based on fast sample consensus for image registration," *IEEE Geosc. Remote Sens. Lett.*, vol. 12, no. 1, pp. 43–47, Jan. 2015.

Ruixiang Li received the B.S. degree in surveying and mapping engineering from the Henan University of Engineering, Zhengzhou, China, in 2016. He is currently working toward the master's degree in surveying and mapping engineering with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China.

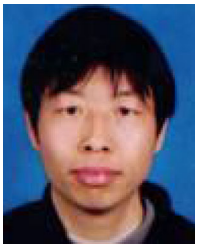
His research interests include multispectral/multimodal image registration and image fusion.

Haitao Zhao received the M.E. degree in surveying and mapping engineering from Wuhan University, Wuhan, China, in 2006, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2013.

He is currently a Senior Engineer with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include POS-supported aerial photogrammetry, hyperspectral remote sensing, and geometric rectification.

Xiaoye Zhang received the bachelor's degree in electronic information engineering and the Ph.D. degree in communication and signal system from Wuhan University, Wuhan, China, in 2010 and 2019, respectively.

He is currently a Postdoctoral Fellow with the Guangdong Diankeyuan Energy Technology Company Ltd., Guangzhou, China and the School of Computer Science, Wuhan University. His research interests include robots, artificial intelligence, and computer vision.

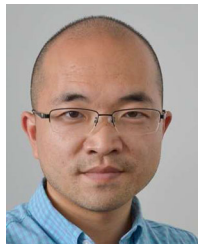


Xiaosan Ge received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2007.

He is currently an Associate Professor with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China. His areas of interest include data fusion of multisensor images and spatial data processing and analysis.

Zhanliang Yuan received the Ph.D. degree in geodesy and surveying engineering from Henan Polytechnic University, Jiaozuo, China, in 2015.

He is currently a Professor and the Deputy Dean at the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China. His area of interest includes fusion of multisource remote sensing data.



Qin Zou (Senior Member, IEEE) received the B.E. degree in information engineering and the Ph.D. degree in computer vision from Wuhan University, Wuhan, China, in 2004 and 2012, respectively.

From 2010 to 2011, he was a Visiting Ph.D. student with the Computer Vision Laboratory, University of South Carolina, Columbia, SC, USA. He is currently an Associate Professor with the School of Computer Science, Wuhan University. His research interests include computer vision, pattern recognition, and machine learning.

Dr. Zou was a corecipient of the National Technology Invention Award of China, in 2015. He is a member of ACM.