Multispectral Image Alignment With Nonlinear Scale-Invariant Keypoint and Enhanced Local Feature Matrix

Qiaoliang Li, Suwen Qi, Yuanyuan Shen, Dong Ni, Huisheng Zhang, and Tianfu Wang

***Abstract*—The scale space-based method has been recently studied for multispectral alignment; however, due to the significant intensity difference between the image pairs, there are usually not enough keypoint correspondences found, and the robustness of the alignment tends to be compromised. In this letter, we attempt to improve the performance from the following two aspects: 1) to avoid the boundary blurring of Gaussian scale space, we adopt nonlinear scale space to explore more keypoints with potential of being correctly matched, and 2) a robust feature descriptor is proposed, and the resulting feature matrix is matched using the previously proposed rotation-invariant distance to obtain more correct keypoint correspondences. Experimental results for multispectral remote images indicate that the proposed method improves the matching performance compared to state-of-the-art methods in terms of correctly matched number of keypoints, aligning accuracy, and rate of correctly matched image pairs. It is also revealed in this letter that, if the descriptor is carefully designed, the local features are distinctive enough for produce good matching even when the main orientation is not present.**

***Index Terms*—Image alignment, nonlinear scale-invariant keypoint, rotation-invariant distance (RID), scale invariant feature transform (SIFT).**

I. Introduction

IMAGE alignment is a vital step in many remote sensing tasks, such as change detection, map updating, image fusion, and environmental surveillance, and it is considered a challenging subject within remote sensing [1]. The traditional method for remote sensing image alignment can be placed into the following two categories. 1) Intensity-based methods: These methods use the pixel intensities to estimate the similarity measures between the image pairs, and the transformations are thereby obtained. The well-known similarity measures are cross-correlation, maximum likelihood [1], and mutual information [2]. These methods require carefully determined initial matching positions, and they suffer from occlusions, image distortions, and illumination differences. 2) Feature-based methods: Such techniques extract features such as edges, corners [3], contours [4], and the centroid of a specific region [5] from the images and use the correlation between these features to determine the optimal alignment between the images. In addition to traditional features, a multi-index image alignment method [6] is proposed recently, where a series of semantic indexes are used for image representation [7]. Nevertheless, most of these features are not scale invariant, and therefore, this type of method is usually not competent when dealing with scaling distortions.

The automatic alignment of remote sensing images with large geometric distortion is still a challenge until today. Recently, the scale invariant feature transform (SIFT) [8] has been successfully applied in registration and recognition for visible images owing to its good characteristics of being invariant to image scaling and rotation and partially invariant to change in illumination and viewpoint, and it has been used to align a remote sensing image with scale, rotation, or viewing angle distortions. However, in multispectral sensing, the objects usually have different reflection or radiation characteristics in different bands; therefore, the same area in multispectral remote images could have significant nonlinear intensity difference. To overcome this problem, in automatic registration of remote-sensing images (ARRSI) [9], instead of image intensity, the local phase congruency is introduced to feature the keypoints and to get more correct correspondences. Yi *et al.* has proposed the scale restricted SIFT (SR-SIFT) [10], an enhanced version of SIFT, in which the scale restriction criteria are introduced to improve the matching performance. In [11], a selection strategy using constraints of stability and distinctiveness is proposed to obtain good SIFT features.

II. Previous Work

Previously, aiming to improve the performance of the SIFT-based aligning method for multispectral imagery, the authors have proposed two algorithms, including robust SIFT (RSIFT) [12] and rotation-invariant distance (RID) [13]. In RSIFT, the feature descriptor of each keypoint is refined to overcome the difference in the intensity between remote image pairs, and a main orientation list is created for each keypoint to enhance the robustness of feature matching. In this algorithm, the feature matching is based on Euclidean distance, the underlying assumption of which is that the keypoints at the same location in the image pair have a consistent main orientation. However, for multispectral imagery, this assumption does not always hold true for keypoint pairs. Therefore, to overcome this problem,

Manuscript received July 25, 2014; revised November 19, 2014 and January 28, 2015; accepted March 11, 2015. Date of publication April 3, 2015; date of current version June 5, 2015. This work was supported in part by the Project of the National Science Foundation of China under Grants 61401285 and 81471735 and in part by the Shenzhen Science Project under Grants JCYJ20140418182819179 and GJHS20120621153609166. *(Corresponding author: Dong Ni.)*

The authors are with the Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Department of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: lql@szu.edu.cn; qisuwen@szu.edu.cn; shenyy@szu.edu.cn; nidong@szu.edu.cn; isaac_zhs@126.com; tfwang@szu.edu.cn).

Digital Object Identifier 10.1109/LGRS.2015.2412955

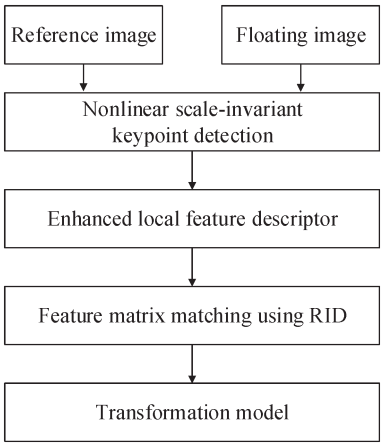

Fig. 1. Flowchart of the proposed method.

recently, we have proposed a distance measure called RID [13], which is an optimal distance independent of the relative main orientation between the keypoint pair.

In the state-of-the-art methods, as well as our previous work, the keypoints are all detected using a SIFT difference of Gaussian (DOG) detector, and this detector relies on the use of the Gaussian scale space, which does not respect the boundaries of objects and smooths to the same degree both details and noise at all scale levels [14], [15]. Therefore, some keypoints with potential of being correctly matched may be undetected by Gaussian scale space. Nonlinear scale space can better preserve edges and texture details in the generated scale-space representation than Gaussian smoothing across scales [16]. Recently, the nonlinear keypoint detector is adopted for natural image alignment [14] and has been proved to be capable to capture more salient keypoints.

In this letter, we present a multispectral image aligning method using a nonlinear scale-invariant keypoint and an enhanced local feature matrix. The proposed method has two novelties: 1) nonlinear scale space-based keypoint extraction and 2) an enhanced scale-invariant feature descriptor. Although the feature descriptors are in the same manner with that presented in [13], the descriptor proposed in this letter is extracted using the scale of the keypoint; therefore, the feature vector is scale invariant and has a rotation-translation characteristic simultaneously. On the other hand, in [13], the feature is extracted for an image patch of fixed size, and the feature is not scale invariant. The detailed procedure of the proposed method is presented in Section III, and the experiments are discussed in Section IV.

III. Proposed Method

The flowchart of the proposed method is shown in Fig. 1. The proposed method consists of three main modules: nonlinear scale-invariant keypoint detection, enhanced local feature descriptor, and feature matching using RID.

A. Nonlinear Scale-Invariant Keypoint Detection

The linear scale space used in SIFT can be considered as a typical nonlinear scale space denoted by

$$\frac{\partial L}{\partial t} = \mathrm{div}\left(c(x,y,t) \bullet \nabla L\right) \tag{1}$$

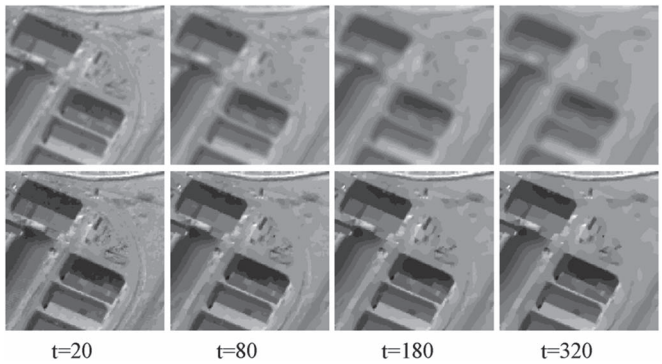

Fig. 2. Comparison between the linear and the nonlinear diffusion scale space for remote sensing images (t is the evolution time). First row: Linear Gaussian scale space. Second row: Nonlinear diffusion scale space. (a) $t = 20$. (b) $t = 80$. (c) $t = 180$. (d) $t = 320$.

with constant diffusion coefficient $c(x, y, t)$, where L is the image and div and ∇ are the divergence and gradient calculation, respectively. In other words, this diffusion coefficient is independent of the image content, resulting in a homogeneous smoothing across the image plane without considering the local boundaries (shown in the first row in Fig. 2). In this letter, a diffusion coefficient c dependent on the local gradient magnitude [16] is introduced to better preserve the local boundaries, and it is defined as

$$c(x,y,t) = \frac{1}{(1 + |\nabla L_\sigma|^2/k^2)} \tag{2}$$

where ∇L_σ is the gradient of a scale version of image L and k is the control factor of the diffusion level. Such version of the diffusion coefficient has been proved to be capable to reduce the blurring of edges while encouraging smoothing of homogeneous regions (shown in the second row in Fig. 2).

Having constructed the scale space, we search for maxima in scale and spatial location to locate the keypoints by exploring the response of the scale-normalized determinant of Hessian [17] at each scale level. In our implementation, the maximum octave is set to 6, the number of sublevels per scale level is set to 4, and the control factor k is set to be 0.01. Most of the parameters are set according to KAZE[1] [14] to provide a proper balance between detection performance and computational load.

B. Enhanced Local Feature Descriptor

Based on the previous work [12], [13], we have developed an enhanced feature descriptor compared to the original gradient location and orientation histogram (GLOH) [18]. For each keypoint, we extract a feature vector $\mathbf{V} = \{\mathbf{V^1}, \mathbf{V^2}\}$ using the location, magnitude, and orientation value of the local image gradient. In this vector, the subvectors $\mathbf{V^1}$ and $\mathbf{V^2}$ correspond to the inner and outer rings with radii $R/2$ and R, respectively (see Fig. 3).

Supposing that each ring is divided into N sectors and the gradient orientation is divided into N levels too, $\mathbf{V^1}$ and $\mathbf{V^2}$

[1] Kaze is the Japanese word for wind.

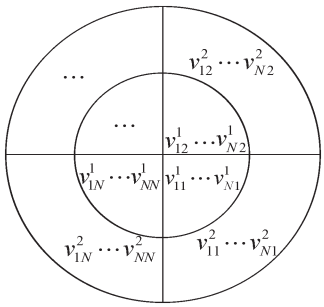

Fig. 3. Correspondence between the descriptor and the feature matrix.

can be reshaped into matrices $\mathbf{H_1}$ and $\mathbf{H_2}$, respectively

$$\mathbf{H_1} = \begin{bmatrix} v_{11}^1 \, v_{12}^1 \, , \ldots , v_{1N}^1 \\ \vdots \quad \ddots \quad \vdots \\ v_{N1}^1 \, v_{N2}^1 \, , \ldots , v_{NN}^1 \end{bmatrix} \quad \mathbf{H}_2 = \begin{bmatrix} v_{11}^2 \, v_{12}^2 \, , \ldots , v_{1N}^2 \\ \vdots \quad \ddots \quad \vdots \\ v_{N1}^2 \, v_{N2}^2 \, , \ldots , v_{NN}^2 \end{bmatrix}. \tag{3}$$

Let $\mathbf{H} = \mathbf{H_1} + \mathbf{H_2}$, and then, each keypoint is associated with a feature matrix $\mathbf{H}$. The matrix will be circularly shifted when the involved image patch is rotated, and we consider it as a "rotation-to-translation" characteristic, which is a prerequisite of RID. The interested reader can be referred to [13] and [18] for the detailed procedure to extract the vector.

1) Robust to Intensity Mapping: Due to the significant difference in the pixel intensity between the multispectral image pair, the gradient orientations of the same area in the image pair may be converse. In this letter, similar to a previous work [12], we modify the gradient orientation of each pixel by

$$R(\varphi) = \begin{cases} \varphi, & \varphi \in [1, 180] \\ 360 - \varphi, & \varphi \in (180, 360) \end{cases} \tag{4}$$

where $R(\varphi)$ denotes the refined value of the original orientation φ. Therefore, the resulting feature would be more robust to intensity mapping compared to GLOH.

2) Scale Invariant With Rotation-to-Translation Characteristic: It has been proved in a previous work [13] that RID can be used for feature matrix matching with a rotation-to-translation characteristic. In this letter, to make full use of the distance measure of RID, we do not assign any main orientation for each keypoint but construct a feature matrix using the same technique as presented in [13]. It must be noted that, in [13], the feature is extracted for an image patch of fixed size; in other words, the feature is not scale invariant. In this letter, each keypoint is detected in scale space, and its scale value can be used to determine the size of the image patch (multiplication of 40 pixels and the scale value); therefore, the feature vector is scale invariant and has a rotation-translation characteristic.

3) Robust to Detection Bias: It has been proved that the GLOH descriptor [18] is, to some extent, robust to the location and scale bias of keypoint detection. Although our descriptor uses a feature matrix instead of a feature vector, the feature extraction procedure is similar to the original GLOH. The feature matrix could be regarded as a reorganized version of the GLOH feature vector, the robustness of which is still preserved.

C. Feature Matching Using RID

For each keypoint in one image, the keypoint with minimum RID to it in the other image will be set as the expected target.

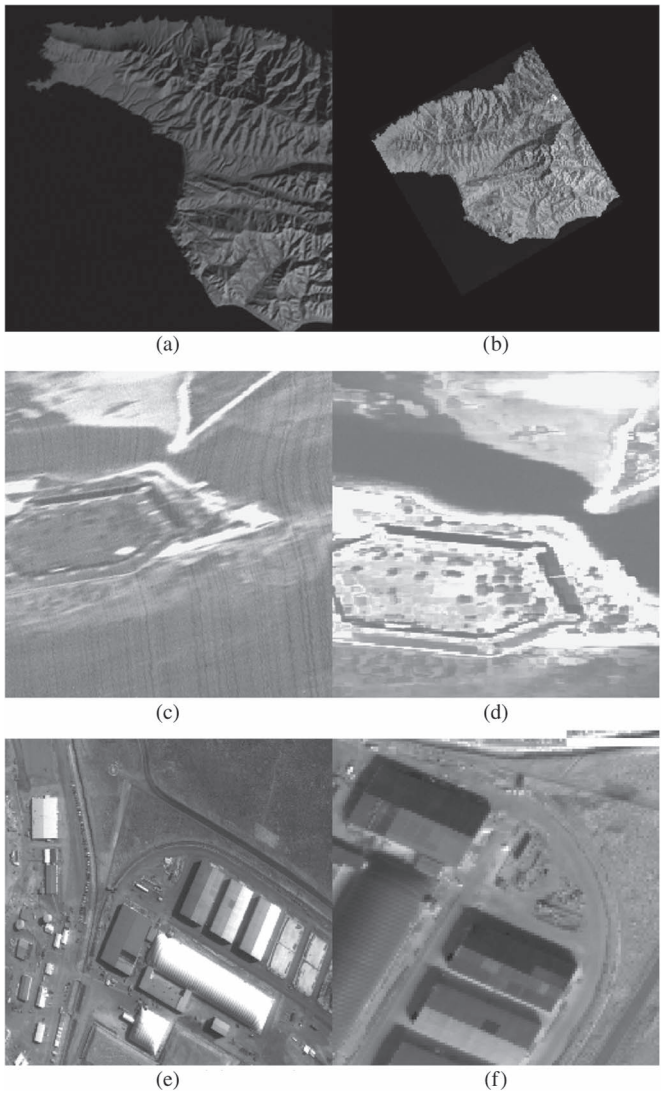

Fig. 4. Remote sensing image pairs. (a) Thematic mapper (TM) date 1984 band 7. (b) TM date 1986 band 7. (c) CASI image band 8. (d) CASI image band 36. (e) Daedalus band 2. (f) Daedalus band 4.

After all keypoints' targets have been found, the matched keypoint pairs will be created. The sector number *N* is set to be 8 to get balance between matching performance and computation complexity. Afterward, the matched keypoint pairs are filtered by the random sample consensus (RANSAC) [8] algorithm, and the projective transformation model between the image pair is estimated using the coordinates of the matched keypoints. Finally, the floating image is transformed by the geometric model, and cubic interpolation is executed to obtain the final aligned image.

IV. Experiments and Results

A. Image Set

In order to evaluate the performance of the proposed method, four image sets are used. The first three are multidate, hyperspectral, and multiband remote sensing image pairs, and the fourth is a group of Daedalus images. The size of all images is 256 × 256 pixels, while the image intensity is stored in

TABLE I
COMPARISON OF RMSE AND CORRECTLY MATCHED NUMBER OF KEYPOINTS

Method	TEST1		TEST2		TEST3	
Manual	1.08		1.02		1.13	
NCC	1.48	21	*	*	*	*
SR-SIFT [10]	1.36	26	2.56	13	*	*
ARRSI [9]	1.42	18	*	*	*	*
RSIFT [12]	1.33	32	1.89	19	*	*
RID [13]	*	*	1.29	33	*	*
Non-linear keypoint + SIFT + Euclid [14]	1.01	62	*	*	*	*
Linear DOG + Multi-scale matrix + RID (for comparison)	1.21	36	1.24	35	2.09	11
Non-linear keypoint + Multi-scale matrix + RID (proposed)	0.80	79	0.85	71	1.76	43

*: Fails to register the image pair (RMSE>50).

TABLE II
COMPARISON OF CORRECT MATCH RATE AND AVERAGE RUN TIME PER IMAGE (FOR TEST4)

Method	Correct match rate	Run-time (seconds)
NCC	22%	632.4
SR-SIFT [10]	85%	23.4
ARRSI [9]	82%	68.1
RSIFT [12]	89%	25.6
RID [13]	32%	27.6
Non-linear keypoint + SIFT + Euclid [14]	65%	43.1
Linear DOG + Multi-scale matrix + RID (for comparison)	91%	31.8
Non-linear keypoint + Multi-scale matrix + RID (proposed)	93%	48.2

gray format. The two images in each pair have the same image resolution.

1) TEST1: Multidate image pair from Landsat-5; shown in Fig. 4(a) and (b); resolution: 30 m; both in band 7.
2) TEST2: Hyperspectral image pair from the Canadian Aeronautics and Space Institute (CASI) scanner; shown in Fig. 4(c) and (d); resolution: 1.5 m; band 8 and band 36.
3) TEST3: Multiband image pair from the Daedalus scanner; shown in Fig. 4(e) and (f); resolution: 0.5 m; band 2 and band 4.
4) TEST4: 150 image pairs from the Daedalus scanner; resolution: 0.5 m; band 1 to band 6.

B. Evaluation Criterion

Alignment Accuracy: The test-point-based method [9] is used to evaluate the alignment accuracy. A total of 60 test-point pairs are selected manually for each case. From the set of manually selected test-point pairs, 30 pairs are selected for training purposes, and the remaining 30 pairs are selected for evaluating purposes. The root-mean-squared error (RMSE) computed for a transformation model that was fitted using the training pairs is used as the reference baseline standard [9]. The RMSE computed for the remaining pairs is used to judge the aligning accuracy of the registration algorithm being tested. For each test case, the algorithm being tested is executed five times, and the average of the five results is computed as the final result.

Keypoint Number: Since the proposed method is based on keypoints, the robustness is highly dependent on the number of the keypoints. The correctly matched number of keypoints is used as the criterion to evaluate the robustness of the proposed method. Note that correctly matched keypoints are the matched points which have been filtered by the RANSAC [8] algorithm as inliers. Since the alignment accuracy is influenced not only by the keypoint number but also the spatial distribution of keypoint pairs, we can use the correctly matched number of keypoints to make a further comparison when the RMSEs of the two algorithms are close to each other.

Correct Match Rate: This measure is computed as the ratio of the correctly matched number to the total number of the image pairs. Note that we consider one image pair to be successfully matched by the test method only when the resulting *RMSE* is less than four pixels.

C. Experimental Results

The proposed algorithm is compared to other methods, including normalized correlation coefficient (NCC), SR-SIFT [10], ARRSI [9], and KAZE [14], and our previous works RSIFT [12] and RID [13]. The SR-SIFT is an enhanced version of SIFT, in which the scale restriction criteria are introduced to improve the matching performance, while ARRSI is a local phase-congruency-based method. These two state-of-the-art methods are used for comparison to show the advantage of the proposed method. All these methods involved for comparison and the proposed method are built under Matlab 2011. The results of alignment accuracy RMSE and correctly matched keypoint number for TEST1–TEST3 are shown in Table I.

The TEST1 images are taken at different times with the same shot band, and there are monotonic relations of the image intensity between the image pair. These luminance changes can be overcome by NCC, SIFT, and phase-based methods. Accordingly, NCC, SR-SIFT, and ARRSI can accurately align this image pair. Note that the number of correctly matched keypoints is significantly improved by the proposed method due to the keypoint detection and enhanced matching algorithm. The TEST2 images are acquired in a different band, and there are rotation and small scaling displacement between the image pair. Due to the different reflecting nature of the object in two bands, there are irregular relations of the image intensity between the image pair. In this case, NCC and ARRSI fail to align the images, while SR-SIFT gives a satisfied result, confirming the ability of SR-SIFT to overcome this problem. On the other hand, only 13 keypoints are correctly matched by SR-SIFT, indicating that the robustness of the alignment cannot be guaranteed for this method, while 71 keypoints are obtained by the proposed method. The TEST3 images are acquired in two different bands, and there are large translation, rotation, and scaling distortion between the image pair simultaneously. NCC, SR-SIFT, and ARRSI cannot correctly align the images $(\text{RMSE} > 50)$, while the proposed method filtered 43 correctly matched keypoints, resulting in accurate alignment. Compared to our previous work RSIFT [12], the proposed method filtered much more keypoints for all test cases. On the other hand, since the descriptor presented in RID [13] is not scale invariant, it fails to handle TEST1 and TEST3.

It can be concluded that the proposed method outperforms the state-of-the-art methods in multispectral image alignment for these three cases. To further justify the effect of the proposed nonlinear keypoint, we especially compare two algorithms which are configured as "linear DOG keypoint+Multi-scale matrix+RID" and "Non-linear keypoint+Multi-scale matrix+RID" (proposed method). As shown in the Table I, the adoption of nonlinear keypoint results in significant improvement of the number of correctly matched keypoints for TEST1–TEST3, benefiting the accuracy of the registration.

To further validate the proposed method, the TEST4 image group, which is composed of 150 image pairs from the Daedalus scanner, is adopted. According to the results presented in Table II, the correct match rate of the proposed method is improved to 93% from 89% [12].

The average computation time of the proposed method is 48.2 s (in Matlab), which is higher than that of the method in previous work [12], [13] but still lower than that of NCC and ARRSI [9]. Nevertheless, we can improve the speed of alignment by implementing the algorithm in C/C++ languages in the future.

V. Conclusion

To achieve robust alignment for multispectral imagery, we have developed a method using a nonlinear scale-invariant keypoint and an enhanced local feature matrix based on the previous work. Experiment results for remote images show that the proposed method gets much more correct matched keypoints and higher aligning accuracy than the state-of-the-art methods, thus benefiting the remote image fusion and analysis.

References

[1] W. Li and H. Leung, "A maximum likelihood approach for image registration using control point and intensity," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 1115–1127, Aug. 2004.

[2] J. Inglada, V. Muron, D. Pichard, and T. Feuvrier, "Analysis of artifacts in subpixel remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 254–264, Jan. 2007.

[3] Z. Xong and Y. Zhang, "Error analysis of corner and center points for image registration," *Can. J. Remote Sens.*, vol. 37, no. 3, pp. 253–263, Jun. 2011.

[4] 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.

[5] Y. W. Sheng, C. A. Shah, and L. C. Smith, "Automated image registration for hydrologic change detection in the lake-rich Arctic," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 414–418, Jul. 2008.

[6] X. Huang, Q. K. Lu, and L. P. Zhang, "A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 90, pp. 36–48, Apr. 2014.

[7] Y. Q. Tang, X. Huang, and L. P. Zhang, "Fault-tolerant building change detection from urban high-resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1060–1064, Sep. 2013.

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

[9] A. Wong and D. A. Clausi, "ARRSI: Automatic registration of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1483–1493, May 2007.

[10] Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on SIFT," *Electron. Lett.*, vol. 44, no. 2, pp. 107–108, Jan. 17, 2008.

[11] 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.

[12] Q. L. Li, G. Y. Wang, J. G. Liu, and S. B. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 287–291, Apr. 2009.

[13] Q. L. Li, H. S. Zhang, and T. F. Wang, "Multispectral image matching using rotation-invariant distance," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 406–410, May 2011.

[14] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *Proc. 12th ECCV*, Florence, Italy, Oct. 7–13, 2012, pp. 214–227.

[15] V. Vrabel and M. Slodicka, "Nonlinear diffusion problem with dynamical boundary value," *J. Comput. Appl. Math.*, vol. 246, pp. 94–103, Jul. 2013.

[16] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[17] M. Brown and D. Lowe, "Invariant features from interest point groups," in *Proc. BMV Conf.*, Manchester, U.K., Sep. 2–5, 2002, pp. 253–62.

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