Citation:
S. Chen, S. Zhong, B. Xue, X. Li, L. Zhao and C. -I. Chang, "Iterative Scale-Invariant Feature Transform for Remote Sensing Image Registration," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 4, pp. 3244-3265, April 2021, doi: 10.1109/TGRS.2020.3008609.

DOI:
https://doi.org/10.1109/TGRS.2020.3008609

Iterative Scale-Invariant Feature Transform for Remote Sensing Image Registration

Shuhan Chen, *Graduate Student Member, IEEE,* Shengwei Zhong, Bai Xue, *Member, IEEE*,
Xiaorun Li, *Member, IEEE*, Liaoying Zhao, *Member, IEEE*, and Chein-I Chang, *Life Fellow, IEEE*

***Abstract*— Due to significant geometric distortions and illumination differences, developing techniques for high precision and robust multisource remote sensing image registration poses a great challenge. This article presents an iterative image registration approach, called iterative scale-invariant feature transform (ISIFT) for remote sensing images, which extends the traditional scale-invariant feature transform (SIFT)-based registration system to a close-feedback SIFT system that includes a rectification feedback loop to update rectified parameters in an iterative manner. Its key idea uses consistent feature point sets obtained by maximum similarity to calculate new alignment parameters to rectify the current sensed image and the resulting rectified sensed image is then fed back to update and replace the current sensed image as a new sensed image to reimplement SIFT for next iteration. The same process is repeated iteratively until an automatic stopping rule is satisfied. To evaluate the performance of ISIFT, both the simulated and real images are used for experiments for the validation of ISIFT. In addition, several data sets are particularly designed to conduct a comparative study and analysis with existing state-of-the-art methods. Furthermore, experiments with different rotation are also performed to verify the adaptability of ISIFT under different rotation distortions. The experimental results demonstrate that ISIFT improves performance and produces better registration accuracy than traditional SIFT-based methods and existing state-of-the-art methods.**

***Index Terms*— Image registration, iterative scale-invariant feature transform (ISIFT), random sample consensus (RANSAC), scale invariant feature transform (SIFT), similarity metric.**

NOMENCLATURE

FS_j	jth feature point set.
ISIFT	Iterative scale-invariant feature transform by rectification.
ISIFTD	ISIFT with a direct feedback rectification.
IS_M	ISIFT with mutual information.
IS_NM	ISIFT with normalized mutual information.
IS_RM	ISIFT with regional mutual information.
IS_RIM	ISIFT with rotationally invariant regional mutual information.
MI	Mutual information.
NMI	Normalized mutual information.
PS_j	jth parameter set obtained by the FS_j.
R	Reference image.
RANSAC	Random sample consensus.
RIRMI	Rotationally invariant regional mutual information.
RMI	Regional mutual information.
RMSE	Root mean square error.
RMSE_D	RMSE by ISIFTD.
RMSE_M	RMSE by IS_M.
RMSE_N	RMSE by IS_NM.
RMSE_R	RMSE by IS_RM.
RMSE_RI	RMSE by IS_RIM.
RMSE_S	RMSE by scale-invariant feature transform.
S	Sensed image.
SAD	Spectral angular distance.
SIFT	Scale-invariant feature transform.
SV	Similarity value.

I. INTRODUCTION

IMAGE registration is a fundamental process of aligning two or more remote sensing images of the same scene and transforming them into one image in a single coordinate system. This process geometrically aligns the two images (referred to as the reference image and the sensed image) acquired by either the same sensor in different times or different sensors from various viewpoints of the same

Manuscript received December 2, 2019; revised April 20, 2020 and June 27, 2020; accepted July 6, 2020. The work of Shuhan Chen and Xiaorun Li were supported by the National Nature Science Foundation of China under Grant 61671408, and in part by the Joint Fund Project of Chinese Ministry of Education under Grant 6141A02022350 and Grant 6141A02022362. The work of Chein-I Chang was supported by the Fundamental Research Funds for Central Universities under Grant 3132019341. *(Corresponding author: Xiaorun Li.)*

Shuhan Chen and Xiaorun Li are with the Department of Electrical Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: 11410057@zju.edu.cn; lxr@zju.edu.cn).

Shengwei Zhong is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: zhongsw_91@foxmail.com).

Bai Xue is with Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250 USA (e-mail: baixue1@umbc.edu).

Liaoying Zhao is with the School of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: zhaoly@hdu.edu.cn).

Chein-I Chang is with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information and Technology College, Dalian Maritime University, Dalian 116026, China, also with Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD 21250 USA, and also with the Department of Computer Science and Information Management, Providence University, Taichung 02912, Taiwan (e-mail: cchang@umbc.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.3008609

scene [1]. Therefore, image registration is a crucial preprocess in many remote sensing applications, such as change detection [2]–[4], image fusion [5][7], superresolution reconstruction [8], [9], image retrieval [10][12], and image caption [13], [14]. Accordingly, designing and developing an accurate, effective, and robust image registration process to yield high registration accuracy is critical to image data analysis since it has significant impact on follow-up data processing. Over the past decades, many research efforts have been directed for developing various methods for remote sensing image registration which can be generally categorized into feature-based methods [15]–[34], area-based methods [35][38], and joint area–feature-based methods [39][42], [47], [48], [58][61]. In the following, we briefly discussed these works and detailed reviews are provided in Section II.

The first category comprises works based on local invariant features due to their robustness to significant geometric and illumination differences and focuses on how to extract enough, uniform distribution and stationary features, construct robust and distinctiveness descriptors, and match feature with outlier elimination. Many feature-based methods are proposed and take advantage of extracted features to obtain better performance for registering remote sensing images with significant geometric differences.

The second category consists of methods developed based on similarity metrics because of their invariance to nonlinear intensity differences. They attempt to construct robust similarity metrics to evaluate the similarity of overlapping regions of image pairs to be registered. Thus, the similarity also reflects the quality of the registration parameters to some extent. The more accurate the correspondence between the similarity metric and the registration error, the better the registration parameters selected based on the similarity metric which is expected to reflect the registration parameters to a greater extent. However, these methods require prior knowledge of the initial parameters or are used for images with the same resolution.

The third category includes methods that align images by combining intensity information with local features. By taking advantage of the robustness of features to scale variances and the high precision of similarity metrics to grayscale differences, this type of methods integrates information obtained from local features and similarity metrics to be able to effectively register images with large geometric and grayscale differences. There are many existing works that use a feature-based method as a coarse registration to calculate parameters, which can be used as the initial condition to perform fine registration using an area-based method. Usually, such fine registration is achieved by solving an objective function or tuning local features based on their similarity values (SVs) to obtain the final registration parameters.

Different from the above-mentioned methods, all of which are indeed feedforward open systems, this article develops an iterative registration system, called iterative scale-invariant feature transform (ISIFT), to further improve the registration accuracy. However, a simple repeatedly implementing image registration system in an iterative process does not work. This is because such a direct iterative image registration system does not guarantee that the resulting registration error will be reduced through an iterative process. To address this issue, a judicious updating rule must be developed to ensure that the registration error will not be increased after each iteration. The proposed ISIFT particularly designs an intelligent strategy that combines spatial consistency and intensity similarity to produce better consistent feature point sets to rectify the current sensed image by a feedback loop via an iterative process.

There are several novelties derived from ISIFT which can be described as follows.

The first and the foremost is to develop a close-feedback registration system to provide a new rectified sensed image that can be used to update and replace the current sensed image via a feedback loop as to improve the registration accuracy. Unlike many modified and extended versions of SIFT which are still feedforward open systems, the proposed ISIFT implements a consistent feature point set selection strategy coupled with a spatial intensity similarity method to obtain consistent feature point sets and then compares the resulting rectified image against the current sensed image to determine whether the current sensed image should be updated and replaced. If it does, the SIFT-rectified image will be fed back to replace the current sensed image for next iteration. Such a feedback loop is referred to as rectification feedback loop. Otherwise, the current sensed image will remain unchanged and be used again for the next iteration. So, ISIFT is not a simple direct iterative process but rather an iterative process of implementing SIFT with an intelligent automatic updating strategy that uses feedback provided by the spatial information obtained from an intensity similarity metric. Technically, ISIFT combines two separate processes, a feature extraction method and a spatial consistency registration based on intensity similarity into a close-feedback system to update currently being processed sensed images iteratively. The use of a feature extraction method combined with an intensity similarity metric method to find consistent feature point sets to rectify the currently being used sensed image via feedback loops is considered as a major novelty derived from ISIFT and believed to be the first work ever reported in the image registration literature.

Another important novelty is the iterative process carried out by ISIFT which provides progressive profiles of how SIFT-matched feature point sets are changed iteration by iteration. During such an iterative process, some feature sets are stable and stay unchanged, but some are not. Such profiles of iterative changes in feature matching point sets offer a rare view of how SIFT works from one iteration to another.

Since the rectified sensed image from a direct feedback of a SIFT does not always work, the third novelty is to include a similarity metric to measure the similarity between the rectified image and the reference image to determine whether their similarity difference is less than the difference between the current sensed image and reference image. If it does, the rectified image is then fed back to replace the current sensed image. Otherwise, the current sensed image will remain the same and unchanged for next iteration. So, ISIFT is not a straightforward iterative process. It includes a custom-designed feedback rule which utilizes intensity similarity information to determine whether the current sensed image needed to be updated by feedback.

The fourth novelty is to develop a spatial–intensity registration method which uses RANSAC to select spatial-consistent feature point sets followed by a similarity metric to further select one feature point set with maximum similarity from the RANSAC-obtained spatial-consistent feature point sets. Although many feature-area-based methods have been proposed in the past, most of them are based on a coarse-to-fine strategy or a similarity metric-based local feature matching. Our proposed spatial consistency based on intensity similarity method develops a new and novel registration strategy which can improve the registration results. Most importantly, this strategy is very suitable for an iterative process. Specifically, the used similarity metric not only can be used to select spatial-consistent feature point sets but also can be used to determine whether the iterative process should be continued or terminated.

As a summary, two great benefits can be offered by ISIFT. One is the joint use of spatial consistency and intensity similarity to select consistent feature point sets which are then used to rectify the currently sensed image to reduce registration errors. The other is the use of feedback resulting rectified sensed image to improve the registration accuracy iteratively by using the SV between two consecutive iterations as an effective measure for registration accuracy and as a stopping rule to determine whether an iterative process should be terminated.

The remainder of this article is organized as follows. The related work is reviewed in Section II. The proposed new iterative image registration method, ISIFT is described in Section III. The experimental setting is detailed in Section IV followed by a comparative study and analysis of extensive experiments along with their results and discussions in Section V. Finally, a brief conclusion and contributions are summarized in Section VI. In addition, the appendix is included to discuss various similarity metrics and provide a list of acronyms used in this article.

II. RELATED WORKS

This section provides a review of previous works relevant to the work presented in this article.

A. Feature-Based Methods

One of the most widely used feature-based registration methods is the scale-invariant feature transform (SIFT)-based method [15] which uses the difference of Gaussian (DoG)-based scale space functions and the distributions of gradients to detect and describe local features followed by a similarity metric or distance to find matching features by a spatial-based or statistical model-based method to eliminate mismatching features. In order to obtain high robustness and accuracy, existing algorithms usually focus on the distribution and significance of local features, the robustness and distinguishability of feature descriptors, and the stability and accuracy resulting from feature matching.

From the perspective of local feature extraction, most methods work on extracting stable, salient, and reasonably distributed local features. To have the feature quality guaranteed, [16] developed a uniform robust scale-invariant feature (UR-SIFT) method based on the stability and distinctiveness constraints which select local feature among the initial SIFT-selected features in the full spatial and scale resolutions. UR-SIFT divides an image into uniform cells as to achieve uniform spatial distribution, which can be used for registering images with large-scale differences. In order to further distribute features in each cell, [17] modified UR-SIFT set a minimum Euclidean distance, r to the number of pixels that should be maintained when the features are selected from input images. Besides, [18] developed optical-SAR SIFT (OS-SIFT) to utilize two methods, multiscale ratio of exponentially weighted averages (ROEWAs) operator and multiscale Sobel detector, to calculate gradients of a SAR image and an optical image, respectively. Instead of building a Gaussian scale space as SIFT does, OS-SIFT constructed two Harris scale spaces for both images. Furthermore, Chang *et al.* [19] employed a bilinear interpolation method to down-sample the image during the construction of a scale space and used a more accurate coordinate transformation to solve accuracy problems which are caused by shrinking the sensed image without interpolation. Sedaghat and Mohammadi [20] used a novel competency criterion, which is based on a weighted ranking process using three quality measures, including robustness, spatial saliency, and scale parameters, and performed in a multilayer gridding schema to improve the quantity and distribution of local features.

As for the local feature description, some SIFT-modified feature descriptors were developed to divide grids and construct a gradient direction histogram in a local region to account for statistics. Bay *et al.* [21] proposed a SIFT-like descriptor, called speeded-up robust features (SURFs) descriptor, which used Haar's wavelets to extract local features. Both SIFT and SURF used the grid Cartesian coordinate system as opposed to gradient location and orientation histogram [22] and DASIY [23], both of which used the grid layout in the polar coordinate system. In order to make local geometric distortions robust, Sedaghat and Ebadi [24] proposed an adaptive binning scale-invariant feature transform which used an adaptive histogram quantization strategy to compute feature locations and gradient orientations to be robust and resistant to a local view distortion so that the discriminability and robustness are significantly improved. Chen *et al.* [25] developed a descriptor, partial intensity-invariant feature descriptor (PIIFD), to modify and limit gradient orientations between 0 and π during computing PIIFD. Sedaghat and Ebadi [26] proposed a SIFT local feature-based distinctive order self-similarity descriptor which ranked correlation values among data points in a local region to construct a descriptor. Sedaghat and Mohammadi [27] proposed a novel descriptor based on an extended self-similarity measure, called histogram of oriented self-similarity (HOSS), which computed the self-SVs in multiple directions using an oriented rectangular patch. Moreover, a novel index map called rotation index of the maximal correlation (RIMC) incorporated with an adaptive log-polar spatial structure was proposed.

Based on the extracted local feature locations with their descriptors, the best matching candidate for each keypoint can be obtained as the initial matching feature point pairs by comparing the ratio of the closest neighboring distance to that of the second-closest neighboring distance. One challenge of these feature point-matching methods is to remove outliers as to increase correct matches. To eliminate incorrect matches, statistical model-based methods using geometric constraints and methods using spatial information are commonly used. The most frequently used robust statistical model-based estimator for eliminating feature mismatching is random sample consensus (RANSAC) [28], which selected the maximum consistent feature points as desired matching points for calculating the final alignment parameters. Song *et al.* [29] proposed a SIFT-based robust estimation algorithm, called histogram of triangle area representation sample consensus for remote sensing image registration. Wu *et al.* [30] also proposed a fast sample consensus to find an initial correct matching feature set and to iteratively select correct matches to increase the correct matches. Finally, an imprecise point removal strategy is further proposed to increase the accuracy of feature matching. Ma *et al.* [31] developed a locally linear transformation (LLT) algorithm, which formulated the outlier elimination and parameter estimation as a maximum-likelihood estimation of a Bayesian model with hidden variables. In order to solve the problem, a local geometrical constraint was introduced and the expectation–maximization (EM) algorithm was used. Based on local geometrical relationship, Ma *et al.* [32], [33] proposed a locality preserving matching (LPM), which formulated the neighborhood structures of potential true matches between two images as a mathematical model and obtained a simple closed-form solution. Ma *et al.* [34] proposed a method based on spatial consistency with a progressive matching strategy and then a sparse approximation was applied to the estimate of spatial consistency, which preserved inner features and eliminated outliers.

B. Area-Based Methods

An area-based method is a nonconvex optimization technique. One of the most successfully used area-based methods is mutual information (MI)-based [35] methods which make use of statistics on intensity correlation of overlapping regions between a reference image and a sensed image. However, such MI-based methods do not take spatial information into account. Thus, many methods modified from MI were proposed to include spatial information. Spatial MI (SMI) [36] integrates MI, which is used to obtain the local optimal transformation with spatial information based on phase congruency, which is invariant to illumination and contrast condition. Region MI (RMI) [37] effectively took care of intensity of pixels in a neighborhood region by using high-dimension statistical analysis. In [38], rotationally invariant regional MI (RIRMI) was developed by combining MI with regional information obtained from the statistical relationship with rotationally invariant description within the overlapping region as to improve the robustness of intensity difference and geometric distortion. All the above-mentioned spatial information-based MI methods mainly focused on how to improve their robustness and distinctiveness simultaneously.

C. Joint Area–Feature-Based Methods

There are two ways to implement joint area–feature-based methods by fusing feature-based and area-based methods. One is to perform preregistration using feature-based matching methods followed by area-based methods to refine the registration accuracy. For example, Gong *et al.* [39] used SIFT to obtain the initial matching parameters and implemented a fine-tuning process by maximizing MI via a modified Marquardt–Levenberg search strategy in a multiresolution framework. Zhao *et al.* [40] implemented a fine registration process by maximizing the RMI via a chaotic quantum particle swarm optimization. The other is also to perform preregistration but differently from the first one by combining feature-based and area-based methods to refine the registration accuracy. Ye and Shan [41] proposed a coarse-to-fine automatic registration scheme in which preregistration was first implemented by the scale restriction SIFT [42] and then followed by Harris' corner detection integrated with a local self-similarity descriptor to yield a more accurate piecewise transformation using normalized correlation coefficients for fine registration.

III. ITERATIVE SIFT

The idea of ISIFT was briefly discussed in Section I. This section presents its ideas and details of how the classical SIFT is extended to an iterative version of SIFT. The entire iterative process is described in Fig. 1 and accomplished by three stages. In the first stage, SIFT is implemented to extract and describe feature points via spectral angular distance (SAD) to measure the matching between two feature sets. It is then followed by the second stage which uses RANSAC to obtain candidate feature sets based on the initial SIFT-matched feature sets and a similarity metric to find a feature set corresponding to the maximum SV as the final feature set. Finally, ISIFT is completed by the third stage which feeds back the sensed image rectified by the transformation matrix obtained by the final feature set to replace the current sensed image as a new sensed image for next round iteration provided that an updating rule is satisfied. The details of implementing each of the above-mentioned three stages are described stage-by-stage as follows.

A. Stage 1: Feature Matching by SIFT

The goal of the first stage is to find the initial matching feature points. Let two images to be registered be denoted by the reference image **R** and the sensed image **S**, both of which are input to SIFT [15] to extract and describe local features, denoted by F_R and F_S, respectively.

For feature extraction, the potential feature points are detected by searching over all their scales and spatial locations. The detection of scale-space extrema is implemented

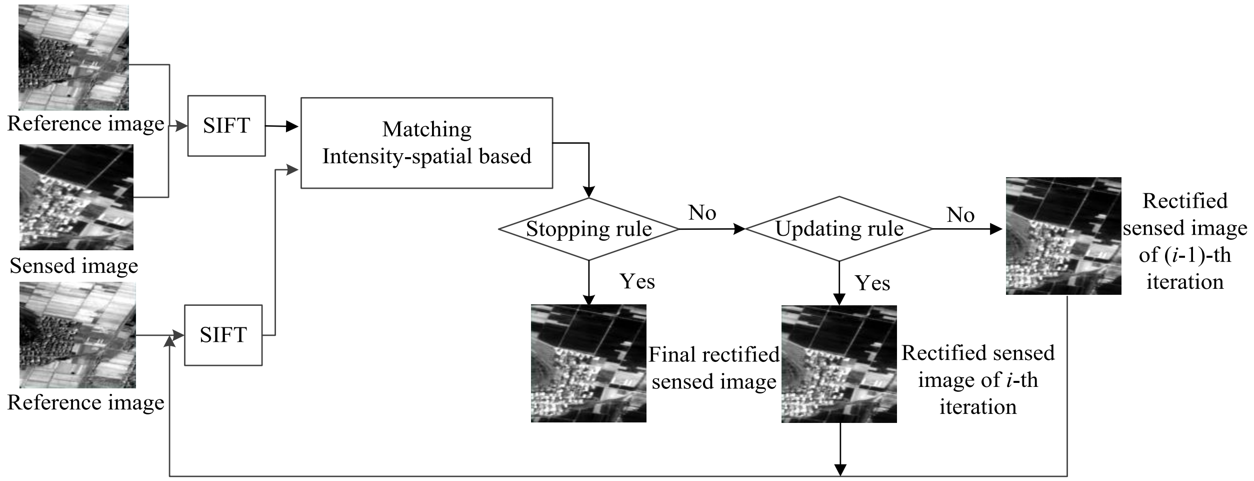


Fig. 1. Graphic diagram of implementing ISIFT based on maximum similarity and random sample consensus.

by detecting the local maxima and minima of $D(x)$, which is defined as the convolution of the DoG function with an image, i.e., F_{S} or F_{R}. In order to exact locations of features, a detailed fit to the nearby data samples is performed. Since there are low-contrast points or poorly localized edge points among these candidate feature points, their contrast values and principal curvature ratios are used to reject the unstable feature points.

After stable candidate features are detected, a dominant orientation is assigned to each keypoint based on local image gradient direction histogram, thereby achieving invariance to image rotation. Then, it is followed by a descriptor vector which is computed for each feature point. Such a descriptor is a 3-D histogram of gradient magnitudes and orientations. The gradient orientation angle is quantized into eight orientation bins and the location is quantized into a 4 × 4 location grid to form a 128-D descriptor.

As the last step, the initial matching process is performed by SAD [43] between feature descriptor vectors defined by

$$\theta = \cos^{-1}\left(\frac{A_i^T B_i}{\|A_i\| \|B_i\|}\right) \tag{1}$$

where A_i is the ith feature descriptor vector in the sensed image, B_i is the ith feature descriptor vector in the reference image, and θ is the angle between the two vectors. The ratio of the first minimum angle to the second minimum angle [44] denoted by θ_{ratio} is used to improve the reliability of the initial match features. If the initial match features with angle distance ratios are greater than a threshold θ_{ratio}, they are rejected.

B. Stage 2: Consistent Feature Point Set Selection by Intensity–Spatial Information

The main idea of the second stage is to use both spatial consistency and intensity similarity metrics to select consistent feature point sets to calculate parameters where SAD is used to obtain the initial matching feature point sets. Consistent feature point sets are selected by RANSAC [28] and then a transformation matrix is calculated for each consistent feature set. If the number of each feature set is less than 3, it will be eliminated. Finally, SV is calculated based on each transformation matrix and a consistent feature set. The feature set corresponding to the maximum SV is selected and used in the current iteration.

More specifically, because of the existence of repetitive features or the limitation of feature descriptors, they may have mismatched features among candidate feature points. Thus, a reliable outlier removal procedure should be implemented to eliminate outliers. RANSAC is a widely used model-based parameter estimation approach to select consistent feature point sets. However, the traditional RANSAC that calculates geometric model parameters to be used to select maximum consistent feature point set as the final selected set is only based on spatial geometric information and does not account for gray level intensity information. Although this method can obtain relatively better results in most cases, its performance degrades if there are many outliers. To address this issue, we improve RANSAC by combining intensity and spatial information to select consistent feature point sets where intensity information is measured by a similarity metric and spatial information is obtained by RANSAC. Specifically, we improve RANSAC by adding an extra piece of information, which is gray-level intensity information to find a set of consistent feature point sets, denoted by $S_{\text{feature}} = \{\mathrm{FS}_j\}$, where FS_j is the jth feature point set, then further use these feature point sets $\{\mathrm{FS}_j\}$ to calculate registration parameter set $S_{\text{parameter}} = \{\mathrm{PS}_j\}$, where PS_j is the jth parameter set obtained by the jth feature point set, FS_j. Let SV_{FS_j} denote the SV between the reference image $\mathbf{R}$ and the rectified sensed image $\mathbf{S}_{\text{rect}}$ which is obtained by geometric rectification using the jth feature point set, FS_j in the registration parameter set S_{feature}. Then $\mathrm{SV}_{\mathrm{FS}_{j^*}}$ is the maximum SV obtained by the feature point set FS_{j^*} where

$$j^* = \arg\left\{\max_{\mathrm{FS}_j \in S_{\text{feature}}} \mathrm{SV}_{\mathrm{FS}_j}\right\}. \tag{2}$$

An affine transformation is a transformation that preserves collinearity (i.e., all points lying on a line initially still remain on a line after transformation) and the ratio of distance (e.g., the midpoint of a line segment remains the midpoint after transformation). Here, the sensed image is rectified by an affine transformation [41] defined by

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{3.1}$$

where (x_1, y_1) represents the coordinate of feature in the reference image, (x_2, y_2) represents the coordinate of feature in sensed image, (a_1, a_2, b_1, b_2) represents the rotation and scale differences, and (c_1, c_2) is the translation between the sensed image and the reference image. In this case, the parameter set is denoted by $\mathrm{PS} = \{a_1, a_2, b_1, b_2, c_1, c_2\}$.

Because of low operating altitude of consumer unmanned aerial vehicles (UAVs), there exist perspective geometric deformations for aerial remote sensing registration. A projective model [45], [46] is introduced as the most appropriate global transformation model, which can handle the affine transform (translation, rotation, and scale) and perspective transform. Thus, we use this projective model to rectify the aerial imagery. Here, the sensed image is rectified by a projective transformation defined by

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \tag{3.2}$$

where (x_1, y_1) represents the original coordinate, (x_2, y_2) represents the transformed coordinate, (a_1, a_2, b_1, b_2) represents the rotation and scale differences, (c_1, c_2) is the shift difference, and (a_3, b_3) is the perspective difference between the sensed image and the reference image. The parameter set is denoted by $\mathrm{PS} = \{a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2\}$.

It is worth noting that in order to define an affine transformation or projective transformation, we need to pick three noncollinear correspondences (affine model) or four noncollinear correspondences (perspective model) from the sensed image and the reference image separately. In practice, for the registration of remote sensing images, it is generally suggested to obtain more than three pairs of points for affine model (four pairs of points for perspective model) and estimate the coefficients through the least squares method.

Among many similarity metrics, MI is robust to nonlinear intensity differences to some extent and has been successfully applied to multispectral or multisensor image registration [47], [48]. Because it can measure significant intensity difference or nonlinear gray difference, some modified similarity metrics based on MI have been proposed for image registration. The four MI-based similarity metrics, NMI, RMI, and RIRMI, to be used in ISIFT are described in detail in the Appendix.

C. Stage 3: Rectification Loop Based on Registration Parameters

Finally, the third stage uses the $\{\mathrm{PS}_j\}$ calculated from the second stage for the sensed image to create a new rectified sensed image which will be used to replace the current sensed image for the next round iteration. Upon implementing the third stage, a stopping rule is also custom designed to determine when the iterative process should be terminated.

In order to implement a rectification feedback loop, the SV between two consecutive iterations is compared to determine whether updating registration parameter set $\{\mathrm{PS}_j\}$ is needed. Since the alignment has randomness caused by a random selection of feature point sets from $S_{\text{feature}} = \{\mathrm{FS}_j\}$, a direct feedback of the iterative results does not necessarily give better registration results. To resolve this dilemma, a similarity metric is used to calculate registration accuracy between two consecutive iterations to determine whether the iterative results should be fed back. Such similarity metric plays three roles. First, it uses gray-level intensity information to obtain a consistent feature point set FS_{j^*} with the maximum similarity among sets of feature points $S_{\text{feature}} = \{\mathrm{FS}_j\}$ according to (3.1) or (3.2). Specifically, for an initial matched feature set, we randomly select three feature pairs and calculate the transformation matrix. The remaining feature sets are substituted into the transformation matrix to select the candidate sets for the current matrix. This random selection is repeated N times (setting 100 times) to obtain N candidate sets. If the number of candidates in a candidate set is larger than 3 [29], the candidate set is retained and used to calculate the transformation matrix (affine) to rectify the sensed image. For projective transformation matrix, the number is 4. During the rectification process, the sensed image is transformed based on the reference image coordinate and using the bicubic interpolation to resample the sensed image based on the gray-level intensity of the sensed image in the last iteration. For each rectified sensed image, the SV of its overlapping region with the reference image is first calculated. The transformation matrix corresponding to the maximum SV is then selected. Second, a similarity metric can be used to quantify the accuracy of the aligned results between two consecutive iterations and then initiate a close-loop rectification. Finally, a similarity metric is used to determine when the iterative process should be terminated. Specifically, when the SV difference between two consecutive iterations is less than a prescribed threshold, the iterative process is terminated.

Updating and stopping rules for ISIFT are given as follows.

1) Updating Rule: The proposed updating rule is based on the difference between two consecutives $\mathrm{SV}_{\max}^{(k)}$ and $\mathrm{SV}_{\max}^{(k-1)}$ defined by

$$\Delta_{\max}^{(k)} = \mathrm{SV}_{\max}^{(k)} - \mathrm{SV}_{\max}^{(k-1)} \tag{4}$$

where $\mathrm{SV}_{\max}^{(k)} = \mathrm{SV}_{\mathrm{FS}_{j^*}}^{(k)}$ and $\mathrm{SV}_{\max}^{(k-1)} = \mathrm{SV}_{\mathrm{FS}_{j^*}}^{(k-1)}$ are obtained by using $\mathrm{FS}_{j^*}^{(k)}$ and $\mathrm{FS}_{j^*}^{(k-1)}$ in (3), respectively. If $\Delta_{\max}^{(k)} > \tau$, then the rectified sensed image obtained by the kth parameters replaces the rectified sensed image obtained by the $(k-1)$th parameters.

2) Stopping Rule: The stopping rule is determined by the fact that the two maximum SVs, $\mathrm{SV}_{\max}^{(k)}$ and $\mathrm{SV}_{\max}^{(k-1)}$, generated at the kth and the $(k-1)$th iterations are sufficiently close with a tolerance value, τ. If $\Delta_{\max}^{(k)} > \tau$, then an updating is needed. Let n be the number of times that no updating is needed. In this case, $n = n + 1$, else $n = 0$, where τ denotes the threshold of SV difference and

$$n > n_\tau \tag{5}$$

where n_τ is a threshold used to limit the number of times that no update is performed, and the initial value of this number is

Algorithm 1 ISIFT

```
Input: reference image $\mathbf{R}^{(0)}$, sensed image $\mathbf{S}^{(0)}$, the number of iterations, $k$, the number of no updating, $n$, maximum iterative number $N$, threshold $\tau$ and $n_\tau$
Output: Final registration parameters
    1) Initialize: $\mathbf{R} = \mathbf{R}^{(0)}$, $\mathbf{S} = \mathbf{S}^{(0)}$ and $k = 0$, $n = 0$.
    2) At the kth iteration, implement SIFT and random sample consensus on reference image $\mathbf{R}$ and sensed image $\mathbf{S}$ to obtain initial consistent feature point sets $S_{\text{feature}}^{(k)} = \{\text{FS}_j^{(k)}\}_{j=1}^{M_k}$, $M_k$ is the number of consistent feature point sets in the kth iteration. Adopt $S_{\text{feature}}^{(k)} = \{\text{FS}_j^{(k)}\}_{j=1}^{M_k}$ to calculate registration parameters set $S_{\text{parameter}}^{(k)} = \{\text{PS}_j^{(k)}\}_{j=1}^{M_k}$ and calculate similarity $\{\text{SV}_{\text{FS}_j}^{(k)}\}_{j=1}^{M_k}$ using $S_{\text{feature}}^{(k)} = \{\text{FS}_j^{(k)}\}_{j=1}^{M_k}$ for image $\mathbf{R}$ and rectified sensed image $\mathbf{S}_{\text{rect}}^{(k)}$ based on $S_{\text{parameter}}^{(k)}$. The parameter set $\text{PS}_{j^*}^{(k)}$ obtained by using $\text{FS}_{j^*}^{(k)}$ to yield the maximum similarity $\text{SV}_{\text{max}}^{(k)}$ is selected and its corresponding rectified sensed image $\mathbf{S}_{\text{rect\_SV}_{\text{max}}}^{(k)}$ is used as a new sensed image in the next iteration.
    3) Check if $k = 0$, go to step 2 and let $k \leftarrow k + 1$. Otherwise, continue.
    4) Comparing the similarity value $\text{SV}_{\text{max}}^{(k)}$ and $\text{SV}_{\text{max}}^{(k-1)}$ in (4). If $\Delta_{\text{max}}^{(k)} > \tau$, then $\mathbf{S}_{\text{rect\_SV}_{\text{max}}}^{(k)}$ is used to replace $\mathbf{S}$ and $n_\tau = 0$; else $\mathbf{S} \leftarrow \mathbf{S}_{\text{rect\_SV}_{\text{max}}}^{(k-1)}$ and $n \leftarrow n + 1$.
    5) Check if $k$ or $n_\tau$ satisfy the stopping rule described in the Section III-B.
    6) ISIFT is terminated and $S_{\text{parameter}}^{(k)}$ is the final registration parameters matrix.
```

set to $n = 0$. The details of implementing ISIFT step-by-step are described as follows.

IV. Experimental Setting

In order to substantiate ISIFT, two groups of data sets are used to validate the utility of ISIFT. The first group contains three simulated images and three real data sets, both of which were used for experiments to demonstrate the superior performance of ISIFT to SIFT without using feedback loops. The image pair of each set is composed of a reference image and a sensed image, which were acquired from different spectra. The details of characteristics of data sets are described in Section IV-A. The second group contains 80 images which are randomly selected from two public available data sets [19], [49], [52]. The detailed information of these two data sets is given in Section IV-B. These images were used for statistical evaluation performance and comparison with other state-of-the-art methods. Moreover, ten images are randomly selected from the 80 images in Group 2 of data sets to further verify the performance of ISIFT under different rotational transformations. All these data sets along with detailed data descriptions are made available at http://wiki.umbc.edu/display/rssipl/10.+Download. Since ISIFT is implemented by four different similarity metrics, MI, NMI, RMI, and RIRMI, it yields four versions of ISIFT, called ISIFT with MI (IS_M), ISIFT with NMI (IS_NM), ISIFT with RMI (IS_RM), and ISIFT with RIRMI (IS_RIM).

A. Group 1: Data Sets Used for Performance Evaluation of ISIFT

1) Simulated Images: As shown in Table I, three simulated image sets contain different bands of multisource images. To demonstrate the applicability of ISIFT to spatial resolution difference, three image pairs (interband images of the same sensor) were obtained by selecting one band of the corresponding multispectral images as the reference image and applying a known transformation (2.5-times scale and 20° rotational changes by ENVI [40]) to one band of the corresponding multispectral images as the sensed image. The corresponding spatial resolutions of the simulated sensed images were calculated based on the scale difference tabulated in Table I. These image pairs also cover a variety of spatial resolutions from 5 to 30 m. For the three image sets, the spatial resolution, sensor, image size, band information, and acquisition time are given in Table I. The reference images are shown in Fig. 2(a1), (b1), and (c1) and the corresponding sensed images are shown in Fig. 2(a2), (b2), and (c2), respectively.

2) Real Images: The three real image sets were selected from different sensors and taken during the same or different years. These image sets have changes in intensity, geometric difference, and scene. These image pairs also cover a variety of spatial resolutions from 2 to 30 m. For the three image sets, the spatial resolution, sensor, image size, band information, and acquisition time are provided in Table I. The reference images are shown in Fig. 2(d1), (e1), and (f1). The corresponding sensed images are shown in Fig. 2(d2), (e2), and (f2).

B. Group 2: Data Sets Used for Statistical Evaluation Performance and Comparison

To further conduct the performance evaluation and comparison, we combined two public available data sets, which contain many satellite images and aerial images. The first public available data set contains 107 multispectral and multitemporal remote sensing image pairs with 512 × 512 [19], [49], obtained from the United States Geological Survey (USGS) website [50]. The pixel resolution of these images is 1 m. The second public available data set [51], [52] contains a 12-class scene and a total of 1200 images with 256 × 256, which were manually extracted from the USGS National Map Urban Area Imagery collection for various urban areas around the country. The pixel resolution of this public domain images is 0.3 m.

In order to verify the statistical evaluation performance under two types of the simulated deformation, affine transformation, and perspective transformation, three data sets of images were generated for experiments as follows.

1) *Set* $A(S_A)$: It consisted of 40 image cubes randomly selected from the first public available data set. For each image cube in S_{A}, one band image was randomly selected as a reference image, denoted by $\mathbf{b}_{\text{A}}^{\text{R}}$, and another band image was also randomly selected from the

TABLE I

INPUT IMAGE PAIRS

	Data sets	Satellite	Spectral mode	Image size	Pixel size (m/pixel) ※	Date
Simulated Imagery	1	Landsat ETM+	Band 1	867×794	30	1999
		Landsat ETM+*	Band 3	436×419	75	1999
	2	ASTER	Band 1	867×794	15	2001
		ASTER*	Band 3	436×419	37.5	2001
	3	IRS-P6	Band 2	867×794	5	2006
		IRS-P6*	Band 3	436×419	12.5	2006
Real Imagery	1	ZY-3	Pan	1500×1500	2.1	2004
		ZY-3	Band 4	480×595	5.8	2004
	2	IRS-1C	Pan	1346×1135	5	2006
		SPOT4	Pan	700×590	10	1999
	3	SPOT 4	Band 1	611×1235	20	1998
		Landsat TM	Band 3	648×1230	30	1996

*Simulated satellite image

※Pixel sizes of the simulated images are estimated based on scale coefficient of the applied transformation

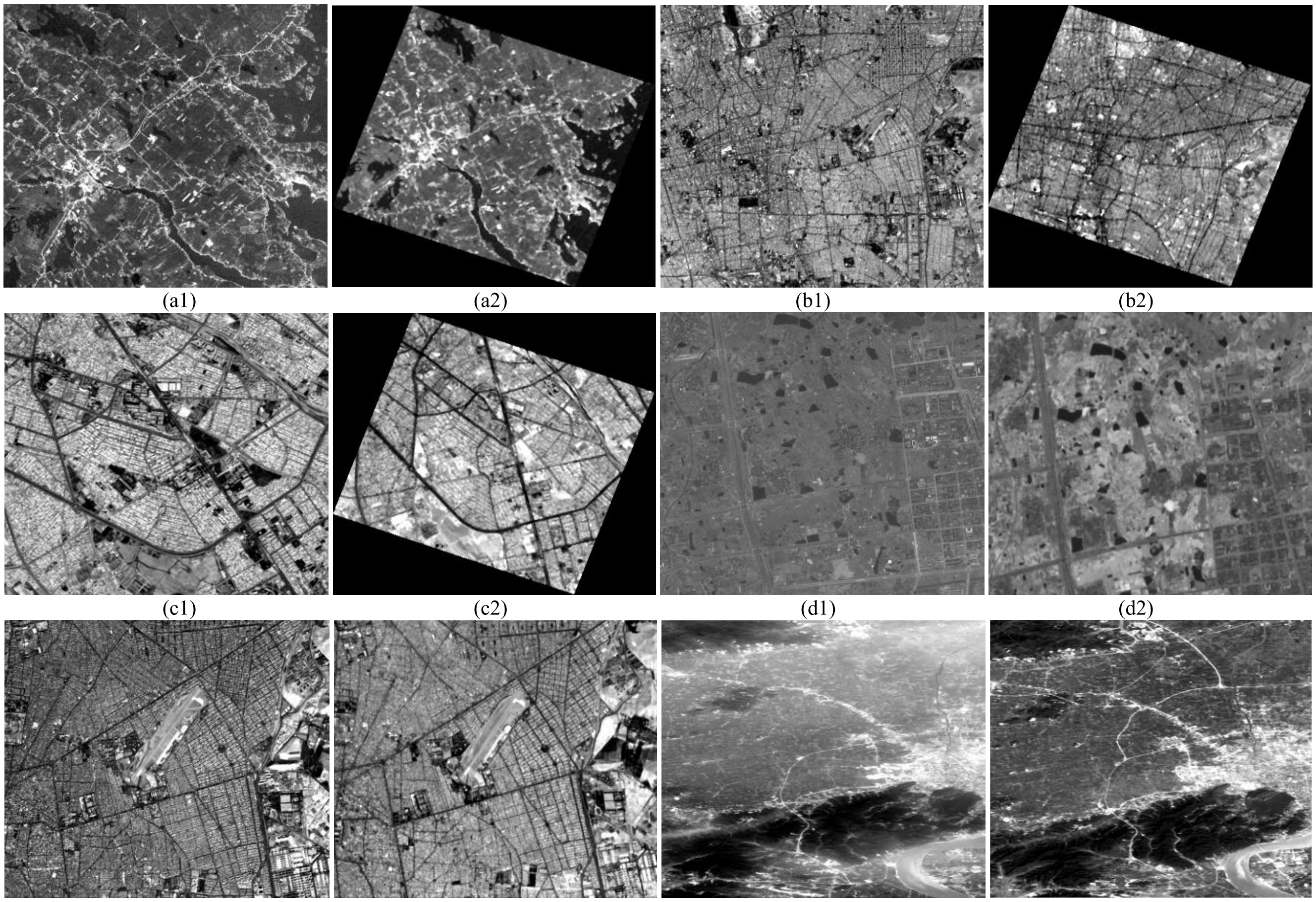

(a1) (a2) (b1) (b2)
(c1) (c2) (d1) (d2)
(e1) (e2) (f1) (f2)

Fig. 2. Input image. (a1), (b1), and (c1) Reference image of simulated image 1, simulated image 2, and simulated image 3. (a2), (b2), and (c2) Sensed image of simulated image 1, simulated image 2, and simulated image 3. (d1), (e1), (f1) Reference image of true image 1, true image 2, and true image 3. (d2), (e2), (f2) Sensed image of true image 1, true image 2, and true image 3.

image cube, denoted by $\mathbf{b}_{\mathrm{A}}^{\mathrm{S}}$, which applied the simulated affine transformations (scaling, rotation, and translation) to create a new sensed image $\hat{\mathbf{b}}_{\mathrm{A}}^{\mathrm{S}}$ to produce an image pair, $(\mathbf{b}_{\mathrm{A}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{A}}^{\mathrm{S}})$. These 40 image pairs were then used to form a new data set, denoted by $\Omega_{\mathrm{A}} = \{(\mathbf{b}_{\mathrm{A}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{A}}^{\mathrm{S}})\}_{S_{\mathrm{A}}}$ for experiments.

2) *Set B*(S_B): In analogy with S_{A}, another data set S_{B} was also generated by 40 image cubes randomly selected from the second public available data set. For each image cube in S_{B}, one band image was randomly selected as a reference image, denoted by $\mathbf{b}_{\mathrm{B}}^{\mathrm{R}}$, and another band image was also randomly selected from the image cube, $\mathbf{b}_{\mathrm{B}}^{\mathrm{S}}$, which applied the simulated projective transformations (scaling, rotation, translation, and perspective) to create a new sensed image, $\hat{\mathbf{b}}_{\mathrm{B}}^{\mathrm{S}}$. As a result, each image cube in S_{B} also produced a pair of two band images, $(\mathbf{b}_{\mathrm{B}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{B}}^{\mathrm{S}})$.

These 40 resulting image pairs (reference image, sensed image) formed a new data set, denoted by $\mathbf{\Omega}_{\mathrm{B}} = \{(\mathbf{b}_{\mathrm{B}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{B}}^{\mathrm{S}})\}_{S_{\mathrm{B}}}$ to be used for experiments.

3) *Set* $C(S_C)$: In order to further validate the robustness of ISIFT under different rotation deformation, a third data set was a mixed data set generated by the above two data sets, S_{A} and S_{B},denoted by $\mathbf{\Omega}_{\mathrm{C}}$. It randomly selected ten image cubes from the joint set of S_{A} and S_{B}, i.e., $S_{\mathrm{A}} \cup S_{\mathrm{B}}$ to form S_{C}. In this case, S_{C} contains of ten image cubes mixed by S_{A} and S_{B}. Now, for each image cube in S_{C}, one band image was randomly selected as a reference image, $\mathbf{b}_{\mathrm{C}}^{\mathrm{R}}$, and another band image was also randomly selected from the same image cube as a candidate sensed image, $\mathbf{b}_{\mathrm{C}}^{\mathrm{S}}$, which was used to create four new rotated sensed images with four different rotations by a transformation (2.5-times scale and different rotational changes, which 20°, 40°, 60°, and 80°) to produce four pairs of $(\mathbf{b}_{\mathrm{C}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{C}}^{\mathrm{S}(20^\circ)})$, $(\mathbf{b}_{\mathrm{C}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{C}}^{\mathrm{S}(40^\circ)})$, $(\mathbf{b}_{\mathrm{C}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{C}}^{\mathrm{S}(60^\circ)})$, and $(\mathbf{b}_{\mathrm{C}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{C}}^{\mathrm{S}(80^\circ)})$. These ten sets of resulting 40 image pairs with each set containing four image pairs produced a new data set, denoted by $\mathbf{\Omega}_{\mathrm{C}} = \{(\mathbf{b}_{\mathrm{C}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{C}}^{\mathrm{S}(20^o)}), (\mathbf{b}_{\mathrm{C}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{C}}^{\mathrm{S}(40^o)}), (\mathbf{b}_{\mathrm{C}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{C}}^{\mathrm{S}(60^o)}), (\mathbf{b}_{\mathrm{C}}^{\mathrm{R}}, \hat{\mathbf{b}}_{\mathrm{C}}^{\mathrm{S}(80^o)})\}_{S_{\mathrm{C}}}$, which would be used to conduct experiments under different rotations.

C. Parameter Settings

Regarding the parameters used for experiments, their specifications are described as follows. Since the smaller the θ_{ratio} in an initial matching process is, the smaller the number of the initial matching feature pairs are and thus, the stricter the condition of constructing correspondence will be. So, as a tradeoff between the number of correct matches and the rate of correct matches, θ_{ratio} was set to 0.7. Another parameter is the maximum number of iterations for RANSAC which was set to 100 [28] based on the calculation equation and experience. In addition, the threshold of the model referred in [15] was set to 0.5. All these parameter settings were done empirically. For the proposed updating rule, the threshold of difference between SVs was set to $\tau = 10^{-4}$. For the proposed stopping rule, the maximum number of iterations for ISIFT was set to 20. The number of continuous nonupdates, n_τ was used for the stopping rule. This parameter is closely related to the running time and registration accuracy of the algorithm. In this article, we set $n_\tau = 5$. That is, if the $n_\tau = 5$, it means that the iterative process is terminated once $n_\tau = 5$.

D. Evaluation Criteria

Root mean square error (RMSE) is used to quantitatively evaluate the alignment accuracy of these algorithms. For an affine transformation, it is defined by

$$\mathrm{RMSE} = \sqrt{\frac{1}{k}} \cdot \left\| \begin{pmatrix} \hat{a}_1 - a_1 & \hat{b}_1 - b_1 & \hat{c}_1 - c_1 \\ \hat{a}_2 - a_2 & \hat{b}_2 - b_2 & \hat{c}_2 - c_2 \end{pmatrix} \begin{pmatrix} x_2^1 & x_2^2 & \dots & x_2^k \\ y_2^1 & y_2^2 & \dots & y_2^k \\ 1 & 1 & \dots & 1 \end{pmatrix} \right\|_2 \tag{6.1}$$

where $a_1, a_2, b_1, b_2, c_1, c_2$ are the real values of the model parameter, $\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2, \hat{c}_1, \hat{c}_2$ are the estimated values of the model parameters, and k is the number of tested pixels.

For a projective transformation, it is defined by

$$\mathrm{RMSE} = \sqrt{\frac{1}{k}} \left\| \begin{pmatrix} \hat{a}_1 - a_1 & \hat{b}_1 - b_1 & \hat{c}_1 - c_1 \\ \hat{a}_2 - a_2 & \hat{b}_2 - b_2 & \hat{c}_2 - c_2 \\ \hat{a}_3 - a_3 & \hat{b}_3 - b_3 & 0 \end{pmatrix} \begin{pmatrix} x_2^1 & x_2^2 & \dots & x_2^k \\ y_2^1 & y_2^2 & \dots & y_2^k \\ 1 & 1 & \dots & 1 \end{pmatrix} \right\|_2 \tag{6.2}$$

where $a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2$ are the real values of the model parameter, $\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{b}_1, \hat{b}_2, \hat{b}_3, \hat{c}_1, \hat{c}_2$ are the estimated values of the model parameters, and k is the number of tested pixels. In this case, RMSE can be used as an evaluation criterion of quantified registration accuracy. As for the ground truth (the simulated images) or the reliable reference geometric transformation parameters, they were calculated by manual registration using ENVI [53].

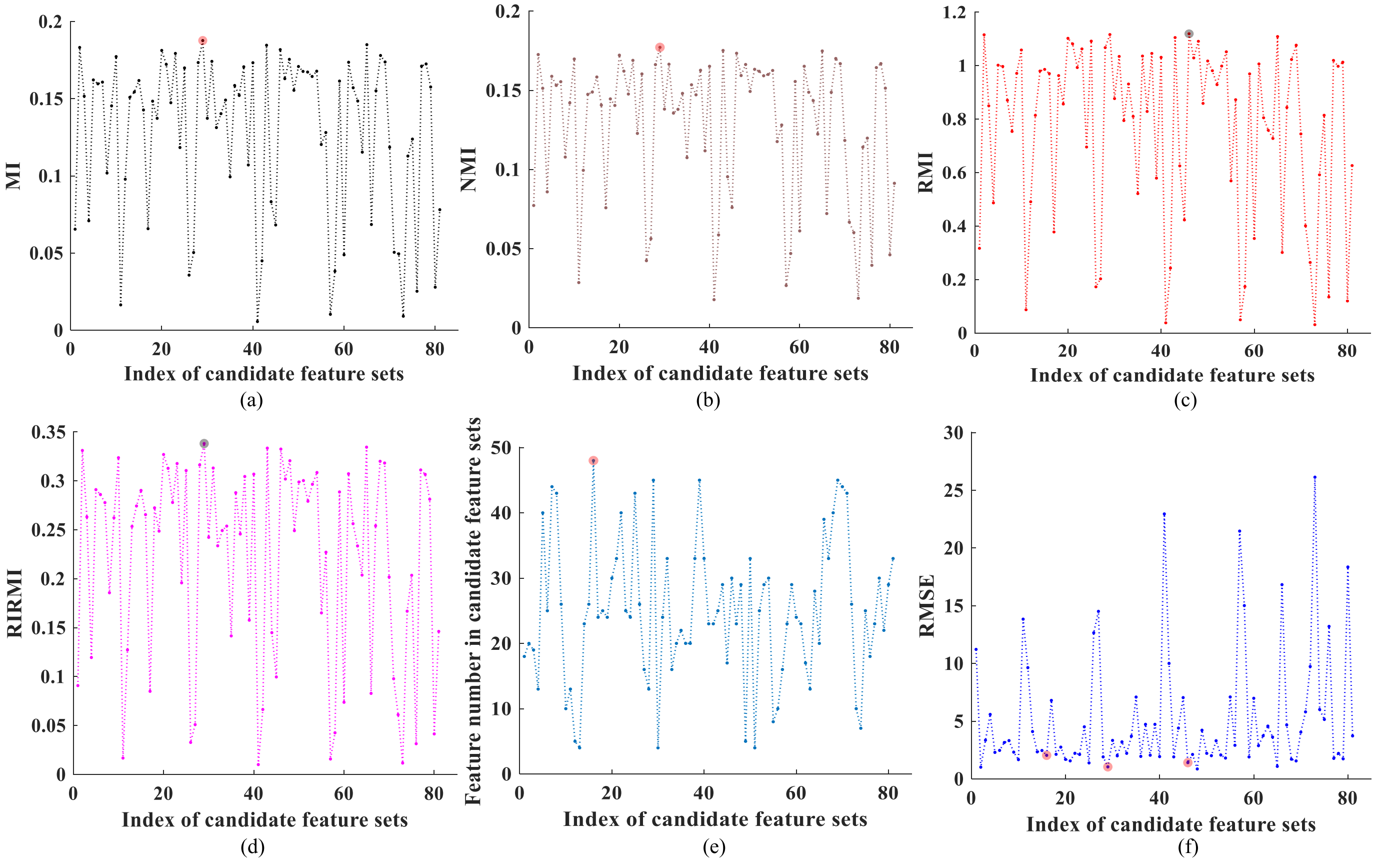
In addition, for the experiment of performance evaluation based on Group 1 data sets, the pairwise linking of feature point pairs pairwise between the reference and sensed images, checkboard mosaicked image, and red–green fusion registration image [54] are provided for visual inspection of the registration results.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed ISIFT was implemented in MATLAB. ISIFT with a direct feedback rectification loop (ISIFTD) without using an update rule is developed in comparison with ISIFT.

More specifically, both SIFT and ISIFTD select consistent feature point sets in each iteration using RANSAC to maximize the number of candidate sets, which are further used to calculate registration parameters to rectify the sensed image. However, there is a key difference between ISIFTD and ISIFT as described by the first and fourth novelties in Section I. ISIFTD always uses the feedback loop to directly update and replace the sensed image by the rectified sensed image generated in the previous loop without using an updating rule to determine whether such feeding back is needed. In contrast, ISIFT includes an updating rule to always check and validate the need of the feedback from the rectified image prior to next iteration. Specifically, ISIFT selects consistent feature point sets in each iteration using RANSAC and SV to maximize the SV of candidate sets. At the end of the kth iteration, ISIFT compares the SV between the reference image and the kth rectified image obtained from the current kth iteration, denoted by $\mathrm{SV}^{(k)}$ against the SV between the reference image and current sensed image, denoted $\mathrm{SV}^{(k-1)}$. If $\mathrm{SV}^{(k)}$ is larger than $\mathrm{SV}^{(k-1)}$, then the kth rectified image is fed back to replace the current sensed image for next round iteration. Otherwise, the kth rectified image will not be fed back, and the current sensed image will remain the same and further be used for next round iteration. Such validation process is referred to as an updating rule.

Compared to ISIFT, ISIFTD directly feeds back the rectified image to replace the current sensed image without an updating rule for validation. Consequently, ISIFTD does not necessarily guarantee that the kth rectified image will be a better sensed image than the current sensed image for next round registration. The experiments conducted in the following sections demonstrate this scenario.



(a) (b) (c) (d) (e) (f)

Fig. 3. SV plots for simulated image 1. (a) MI plots (index is 29 and SVMI = 0.1876). (b) NMI plots (index is 29 and SVRMI = 0.1771). (c) RMI plots (index is 46 and SVRMI = 1.119).

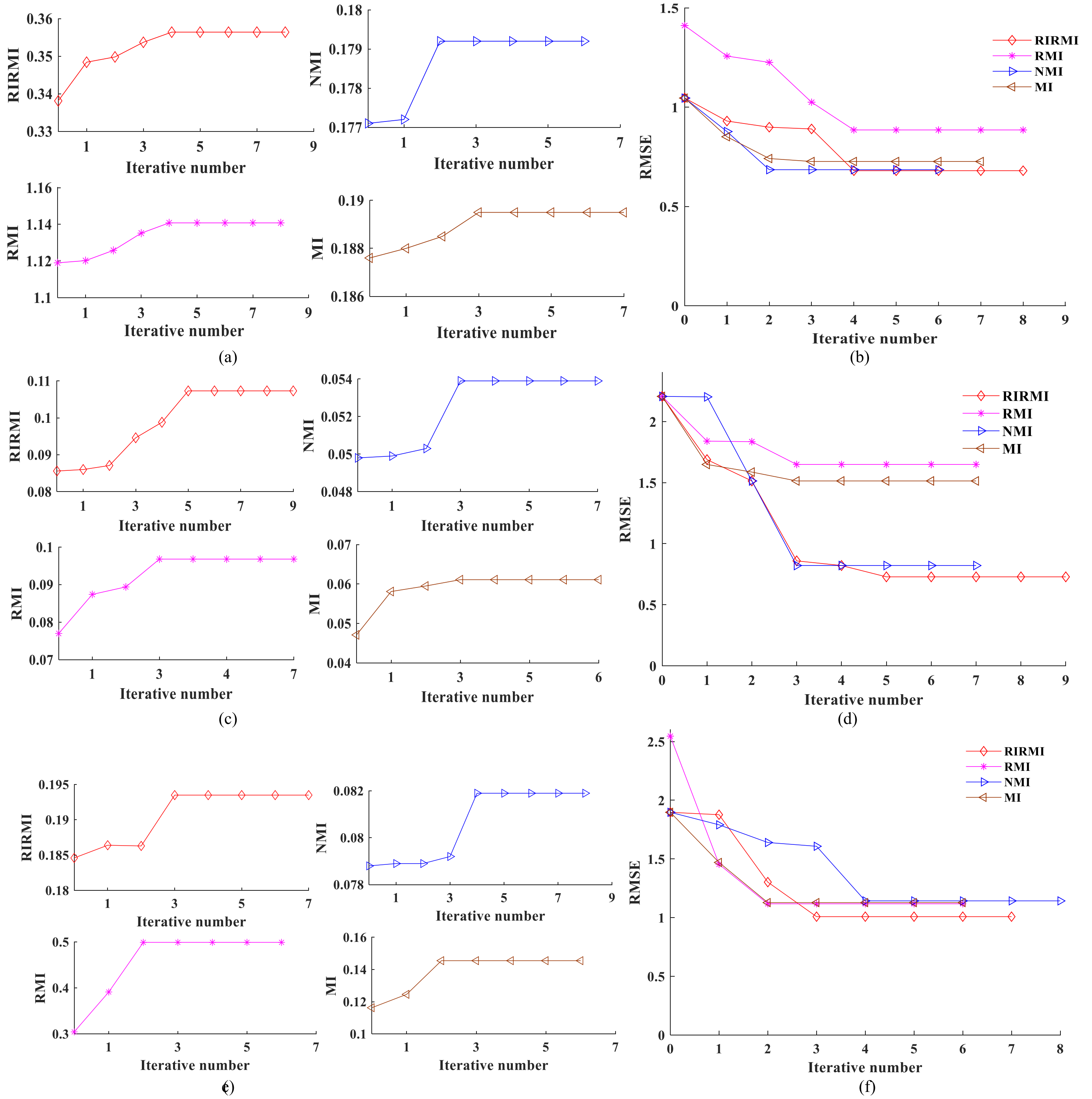
To evaluate ISIFT, ISIFT using four different similarity metrics, i.e., IS_M, IS_NM, IS_RM, and IS_RIM, was implemented to compare ISIFTD and SIFT, for performance analysis. So, a total of six methods were tested for comparison. Specifically, experiments were performed for ISIFT with and without feedbacks to illustrate that ISIFTD did not always succeed in improving registration accuracy. This was because it did not feedback the best rectified image that could improve the current sensed image. In addition, to further conduct quantitative analysis, the RMSEs were calculated and the visual displays of the final fused registered images resulting from six methods using selected consistent feature point sets are also plotted for comparison.

A. Performance Evaluation of ISIFT

1) Simulated Image Experiments: For simplicity, the RMSE calculated by six methods, IS_RIM, IS_RM, IS_NM, IS_M, ISIFTD, and SIFT are denoted by RMSE_RI, RMSE_R, RMSE_N, RMSE_M, RMSE_D, and RMSE_S. For ISIFTD, RMSE was calculated to select consistency feature point sets from the candidate feature points.

Fig. 3 shows the SV calculated by RIRMI, RMI, NMI, and MI along with the total number of consistent feature sets and RMSEs in each candidate feature set for simulated image 1 without any iteration yet. In all the figures of Fig. 3, the horizontal axes are the indexes of candidate feature sets and the vertical axes in Fig. 3(a)–(d) are SVs produced by RIRMI, RMI, NMI, and MI, while the vertical axis in Fig. 3(e) is the total number of features in each candidate feature set. Fig. 3(f) shows the RMSE versus the index of candidate feature sets where RMSE was calculated by each candidate feature set. In above experiments, the total number of candidate sets was set to 100 and only the number of features in each candidate set larger than 3 was retained. However, it should be noted that the resulting number of each candidate set varies with different images, for example, 81 in Fig. 3 but 45 obtained from the real image pair in Fig. 7.

As shown in Fig. 3, all the candidate feature sets are shown by light solid dots, which are connected by dotted lines. The final feature sets highlighted by large solid circles were selected by IS_RIM, IS_RM, IS_NM, and IS_M and their corresponding RMSE. The RMSE calculated by ISIFT are relatively smaller than or equal to the RMSE calculated by RANSAC. In order to demonstrate the effectiveness of the selection strategy in ISIFT, we analyze its results in detail in Fig. 3. The feature set index corresponding to the maximum SV is abbreviated as j^*_{MI}, j^*_{NMI}, j^*_{RMI}, j^*_{RIRMI}, and j^*_{RANSAC}. Intuitively, Fig. 3(a), (b), and (d) shows that $j^*_{\mathrm{MI}} = j^*_{\mathrm{NMI}} = j^*_{\mathrm{RIRMI}} = 29$ with their calculated RMSE corresponding to 1.046 pixels as shown in Fig. 3(f). In contrast, Fig. 3(c) shows that $j^*_{\mathrm{RMI}} = 46$ with its RMSE corresponding to 1.411 pixels as shown in Fig. 3(f). On the other hand, Fig. 3(e)



(a)

(b)

(c)

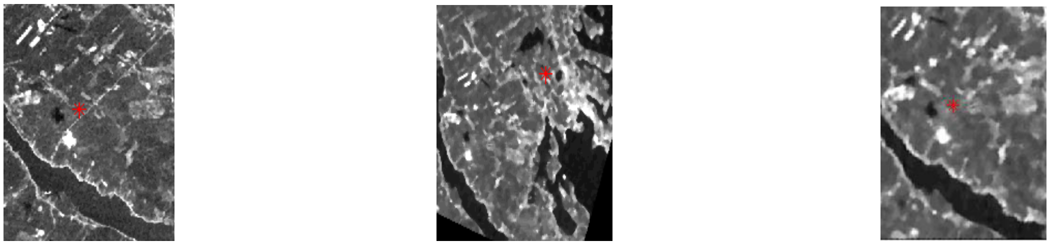
(d)

(e)

(f)

Fig. 4. Iterative plots for simulated images 1–3. (a) SVs for the simulated image. (b) RMSE plots for simulated image 1. (c) SVs for simulated image 2. (d) RMSE plots for the simulated image. (e) SVs for simulated image 3. (f) RMSE plots for simulated image 3.

presents that $j^*_{\mathrm{RANSAC}} = 16$with its RMSE corresponding to 2.045 pixels as shown in Fig. 3(f). Thus, compared to using the number of features as a selection criterion, using SV as a measurement can obtain a better final feature set with a lower RMSE. This shows that the maximum feature set selected by RANSAC could not represent the real geometric relationship between the reference and sensed images. In addition, these results demonstrated that the accuracy of selected feature sets had a direct effect on the registration results. Specifically, when there are large outliers, the maximal RANSAC-selected feature sets may not correspond to high precision (as shown in Fig. 3(f), 2.045 pixels). From Fig. 3, we can also see that the combination of intensity similarity and spatial consistency obtained better results. Although the values of RMSE and similarity metric were not exactly one-to-one correspondence, the Appendix shows that a relatively higher similarity metric value corresponds to a higher precision and so, the result is closer to the minimum RMSE. Based on above analyses, the four metrics can yield good performance for the simulated images.



Feature point in reference image $\mathbf{F}_{ref}$ Feature point in sensed image in the 1st iteration $\mathbf{F}_{sensed}$ $\mathbf{F}^{1^{st}}_{sensed}$

Fig. 5. Single feature pair for simulated image 1 by IS_RIM.

TABLE II
COMPARISON OF THE FINAL RSMEs BASED ON SIFT, ISIFTD, AND ISIFT (MI, NMI, RMI, AND RIRMI) FOR THE SIMULATED IMAGE PAIRS

	Data sets	RMSE_S	RMSE_D	RMSE_M	RMSE_N	RMSE_R	RMSE_RI
Simulated Image	1	2.0446	9.7368	0.7266	0.6853	0.8853	**0.6802**
	2	2.4593	5.2445	1.5137	0.8211	1.6486	**0.7284**
	3	2.5461	17.751	1.1277	1.1422	1.1175	**1.0080**

The same experiments were also implemented on simulated images 2 and 3. The same conclusions as simulated image 1 could be also drawn. In this case, only the results for simulated image 1 are included here to avoid duplication. Fig. 4 shows the iterative profiles of changes in SV and RMSE produced by ISIFT using 4 different similarity metrics, MI, NMI, RMI, and RIRMI. Specifically, Fig. 5 shows an example of how a single feature point pair was matched during the entire iterative process. As we can see, the matched feature pair initially mismatched prior to iteration and then was corrected in the first iteration. In subsequent iterations, the locations of the corresponding feature points in the sensed image had slight differences. Nevertheless, they had no visual effect. Fig. 5 shows the iteration process with significant rectification in subsequent iterations. If a matched feature pair is shown in the entire process, it indicates that the keypoint is stable and significant. However, most importantly, ISIFT indeed corrected mismatched feature pair after the 1st iteration.

Fig. 6 provides visual inspection of feature point pairs pairwise linking, checkboard mosaicked images, and red–green fusion images for the final iteration results. As shown in Fig. 6, IS_RIM could rectify well and obtain more accurate matching feature pairs than the other three similarity metrics. This confirms the results in Fig. 4. The checkboard mosaicked images are also included to display the registration results of the reference image with the rectified sensed image, whereas the red–green fusion images can be used to see whether the upper and lower layers are ghosted by the overall visual effect.

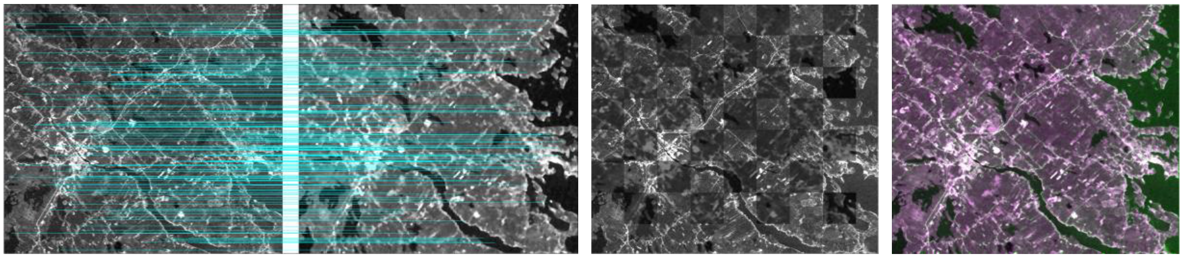
Table II tabulates the final RMSE results of ISIFT, ISIFTD, and SIFT, referred to as RMSE_RI, RMSE_R, RMSE_N, RMSE_M, RMSE_D, and RMSE_S, where RMSE_RI, RMSE_R, RMSE_N, and RMSE_M were better than RMSE_D and RMSE_S. The results also demonstrated that the final RMSE were closely related to the use of a similarity metric.

Combining the quantitative analysis and visual analysis shows that IS_RIM performed better than the other three methods (IS_RM, IS_NM, and IS_M) in the sense of its significant difference in gray values, texture, geometric distortion, and its robustness to similarity variations. Specifically, the other three methods (IS_RM, IS_NM and IS_M) had different position shifts, while IS_RIM had relatively small position shifts.

B. Real Image Experiments

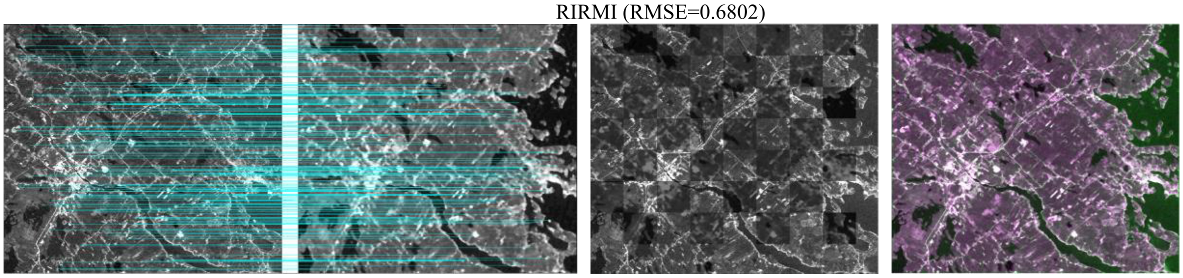
In analogy with the simulated images, real image experiments were also conducted for the six methods. Like Fig. 3, Fig. 7(a)–(d) shows the SV calculated for each candidate feature set for real image 1 by MI, NMI, RMI, and RIRMI, respectively, prior to iteration. Fig. 7(e) shows the total number of features in each candidate feature set for real image 1, while Fig. 7(f) plots RMSEs in each candidate feature set for real image 1 before iteration where RMSEs were calculated by each candidate feature set. Intuitively, Fig. 7(a) shows that $j^*_{\mathrm{MI}} = 41$ with its calculated RMSE corresponding to 2.466 pixels. Fig. 7(b) shows that $j^*_{\mathrm{NMI}} = 36$ with its calculated RMSE corresponding to 2.177 pixels. Fig. 7(c) shows that $j^*_{\mathrm{RMI}} = 11$ with its calculated RMSE corresponding to 1.925 pixels. Fig. 7(d) shows that $j^*_{\mathrm{RIRMI}} = 1$ with its calculated RMSE corresponding to 1.608 pixels. On the other hand, $j^*_{\mathrm{RANSAC}} = 26$ with its RMSE corresponding to 2.419 pixels is shown in Fig. 7(f). Compared to using the maximal number of features as a selection criterion, using maximal SV obtained a better feature set corresponding to a lower RMSE. Also demonstrated, using the maximum consistency feature set could not reflect the real geometric relationship between images. Similarly, the same experiments were also performed on real image 2 and image 3 before iteration. Due to the same observations and conclusions, their results are not included here.

Fig. 8 shows the iteration plots of the SVs and their corresponding RMSEs produced by MI, NMI, RMI, and RIRMI



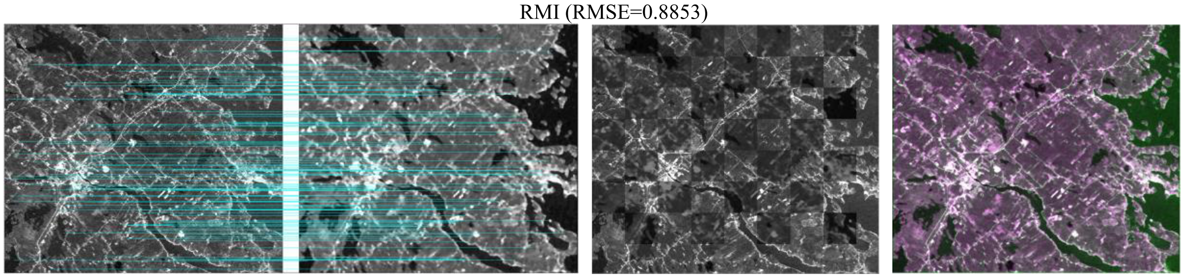
Feature point pairs pairwise linking Checkboard mosaicked images Red-green fusion images

RIRMI (RMSE=0.6802)



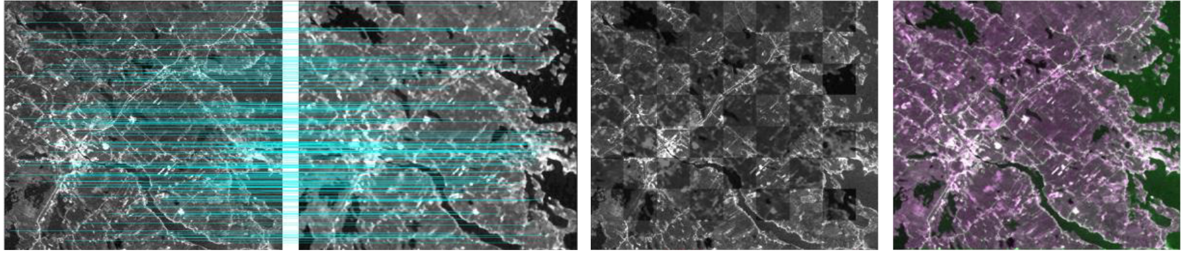
Feature point pairs pairwise linking Checkboard mosaicked images Red-green fusion images

RMI (RMSE=0.8853)



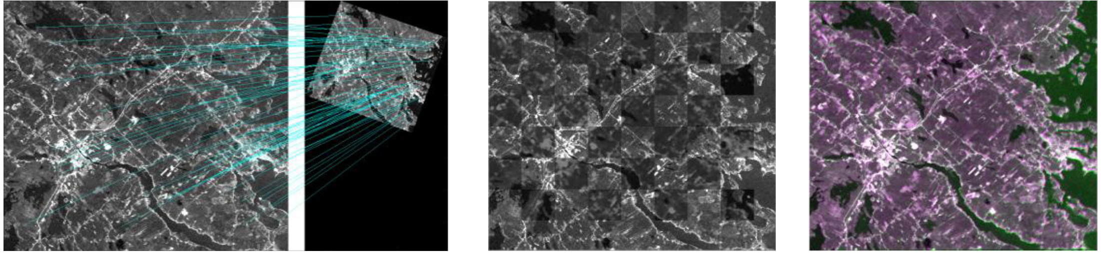
Feature point pairs pairwise linking Checkboard mosaicked images Red-green fusion images

NMI (RMSE=0.6853)



Feature point pairs pairwise linking Checkerboard mosaicked images Red-green fusion images

MI(RMSE=0.7266)



Feature point pairs pairwise linking Checkboard mosaicked images Red-green fusion images

SIFT (RMSE=2.045)

Fig. 6. Matching keypoint images, red–green fusion images, and checkboard mosaicked images based on ISIFT (RIRMI, RMI, NMI, and MI) and SIFT in the final iteration for simulated image 1.

for real images 1–3. Table III tabulates the final RMSEs of ISIFT using RIRMI, RMI, NMI, MI, ISIFTD, and SIFT. For the sake of visual display, Fig. 9 only shows visual inspection of feature point pairs pairwise linking, checkboard mosaicked images, and red–green fusion images of real image 1 for the final iteration. It should be noted that the iterative profiles of changes in SV and RMSE produced by ISIFT based on four similarity metrics plotted in Fig. 8 can be further used to analyze the iteration-by-iteration performance. As demonstrated previously, the registration results were improved as the similarity metric value was increased. Although the results from individual cases may vary on some occasions, the

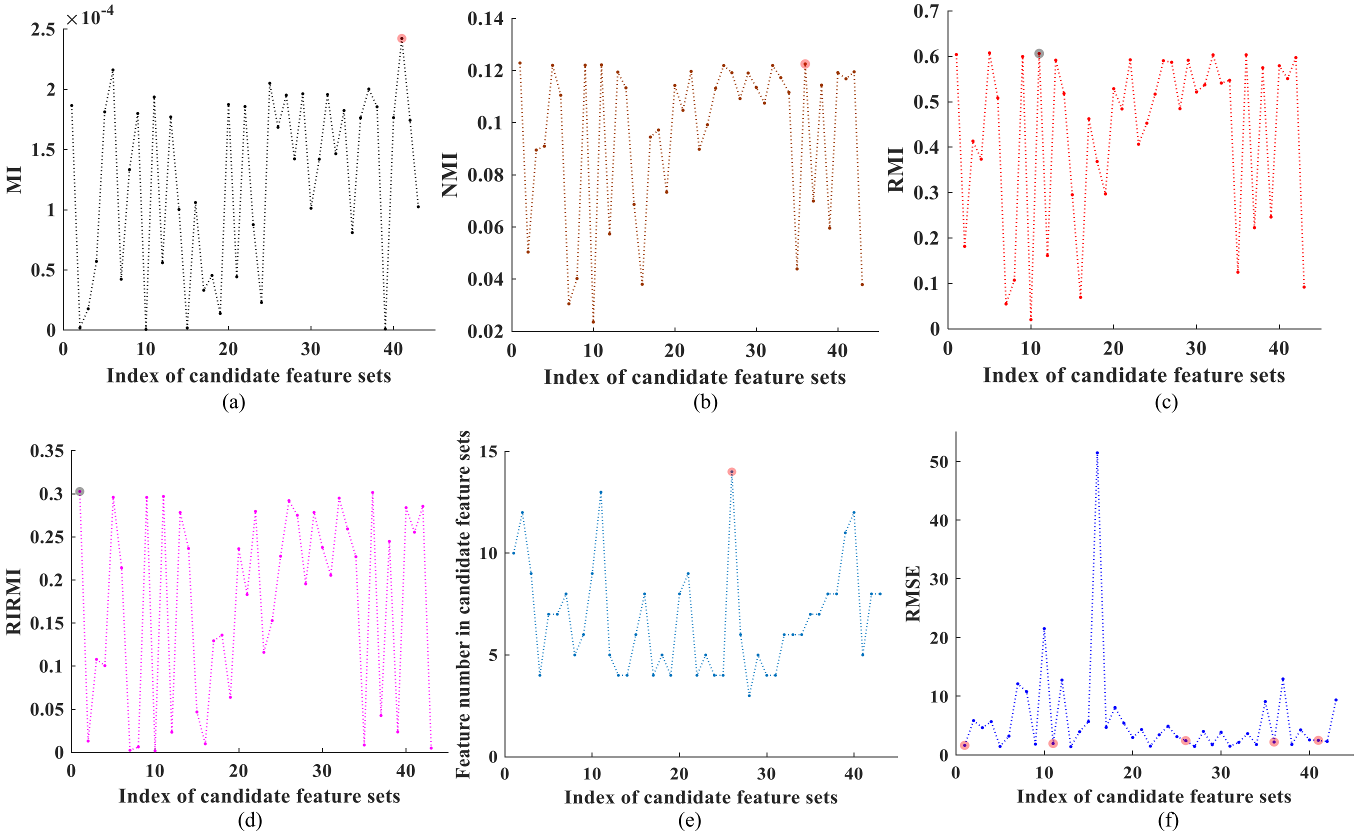


Fig. 7. SV plots for real image 1. (a) MI plots (index is 41 and SVMI $= 2.4 \times 10^{-4}$). (b) NMI plots (index is 36 and $\mathrm{SV_{RMI}} = 0.1225$). (c) RMI plots (index is 11 and SVRMI = 0.6063). (d) RIRMI plots (index is 1 and $\mathrm{SV_{RIRMI}} = 0.3027$). (e) Number of features in each feature set (index is 26 and $N_{\mathrm{Inner}} = 14$). (f) RMSE plots ($\mathrm{RMSE}_1 = 1.608$, $\mathrm{RMSE}_{11} = 1.925$, $\mathrm{RMSE}_{26} = 2.419$, $\mathrm{RMSE}_{36} = 2.177$, and $\mathrm{RMSE}_{41} = 2.466$).

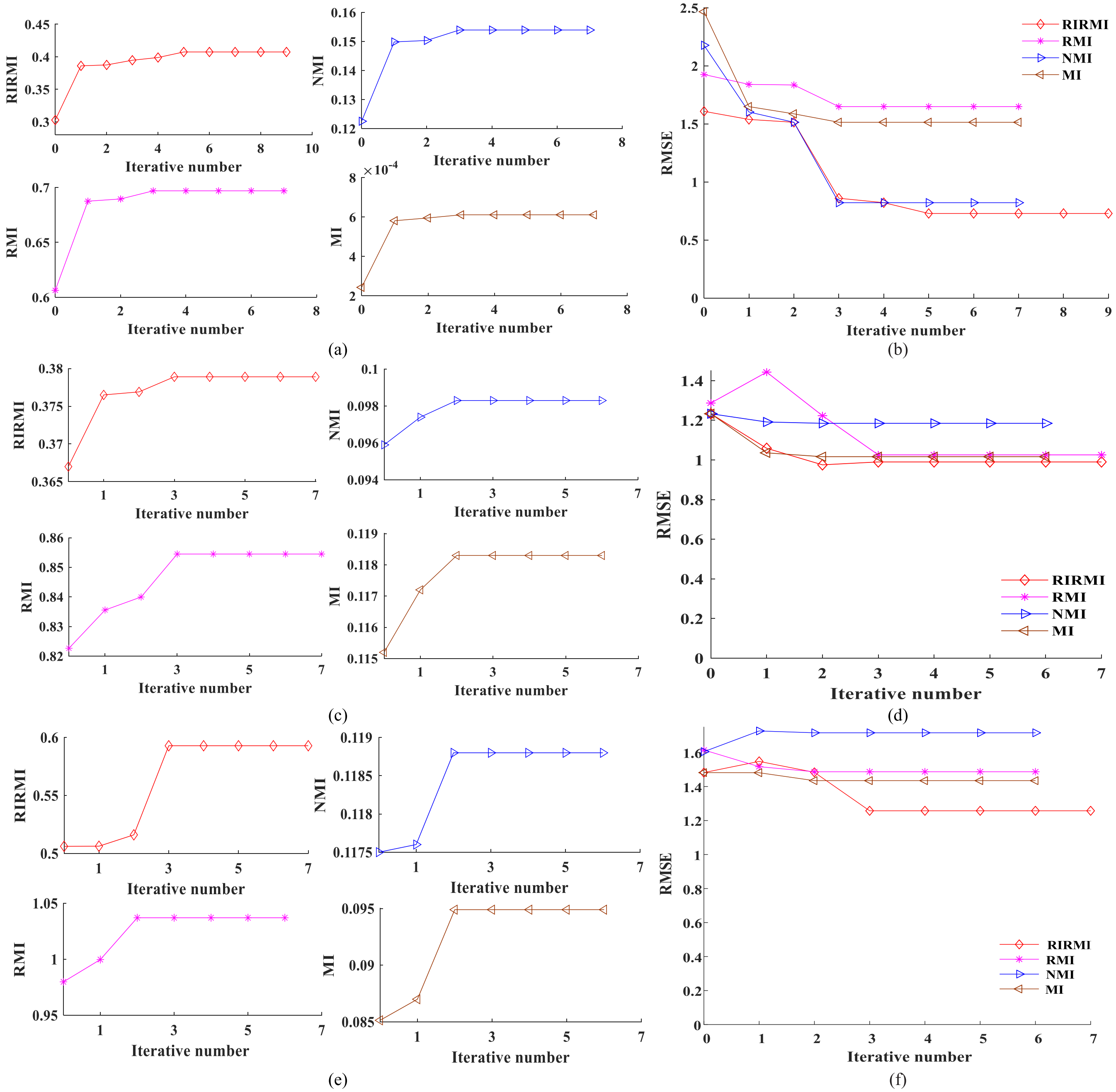
TABLE III

COMPARISON OF THE FINAL RSMES BASED ON SIFT, ISIFTD, AND ISIFT (MI, NMI, RMI, AND RIRMI) FOR REAL IMAGE PAIRS

	Data sets	RMSE_S	RMSE_D	RMSE_M	RMSE_N	RMSE_R	RMSE_RI
Real imagery	1	2.4193	5.2445	1.5137	0.8211	1.6486	**0.7284**
	2	1.8439	2.521	1.0162	1.1847	1.0255	**0.9899**
	3	1.6233	1.5981	1.4347	1.7162	1.4870	**1.2574**

general tendency remained the same. That is, Fig. 8(a) and (b) shows that the registration accuracies in terms of RMSE were decreased as SVs were increased for real image 1. However, for Fig. 8(c) and (d), a smaller RMSE did not necessarily correspond to a larger similarity in a few individual cases. This phenomenon can be explained by the fact that the limitation of using a similarity metric as a criterion, which does not exactly correspond to the registration error. However, the iterative profiles plotted in Fig. 8 demonstrated that the larger the SV was, the smaller the registration error was. The results for three real images also indicated that the maximum SV did not necessarily correspond to a maximum consistency point set.

As seen from the fused sensed images in Fig. 9, ISIFT could also rectify the sensed images and it did for the simulated images. In this case, although the fusion results looked good, the ghosting of overlapped regions reflected the quality of the registration results. Besides, the dislocations of the fused image using the traditional SIFT could be also seen visibly. As shown in Fig. 9, the ghosting of the fusion images, which were obtained by IS_RIM and IS_NM, was slightly reduced compared to those obtained by IS_RM and IS_M. In Table III, the results could explain this phenomenon from a quantitative perspective and be also used to verify the performance of these metrics. For real image 1, the same candidate feature set is used as the initial condition for the six algorithms. Compared to the final RMSEs of ISIFT, ISIFTD, and SIFT, RMSE_RI, RMSE_R, RMSE_N, and RMSE_M were smaller than RMSE_D and RMSE_S. Comparing Tables III to II, the RMSEs of real images were relatively larger than RMSEs of the simulated images. This is because real images have greater complex geometric distortion than the simulated images do. Based on the above quantitative analysis and visual inspection, the proposed ISIFT indeed performed better than SIFT without rectification feedback loops.



(a)

(b)

(c)

(d)

(e)

(f)

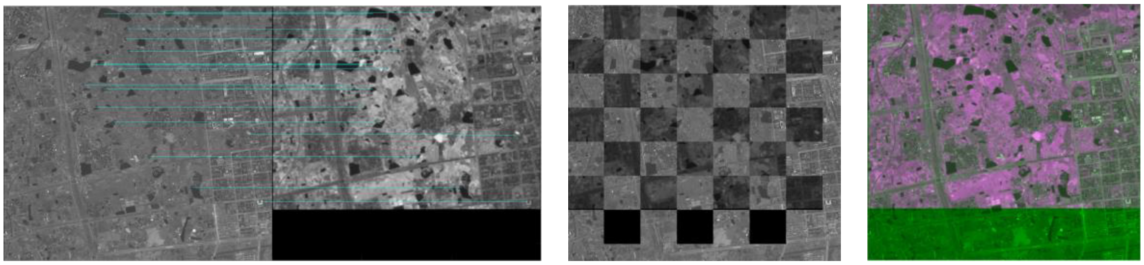
Fig. 8. Iterative plots for real images 1–3. (a) SVs for real image 1. (b) RMSE plots for real image 1. (c) SVs for real image 2. (d) RMSE plots for real image 2. (e) SVs for real image 3. (f) RMSE plots for real image 3.

C. Comparative Analysis on Registration Performance

To perform a comparative study and analysis, two data sets generated in Section IV-B, Ω_A and Ω_B were used for experiments. In addition, to further verify the robustness of test algorithms under different rotations, Ω_C generated in Section IV-B was also used for experiments.

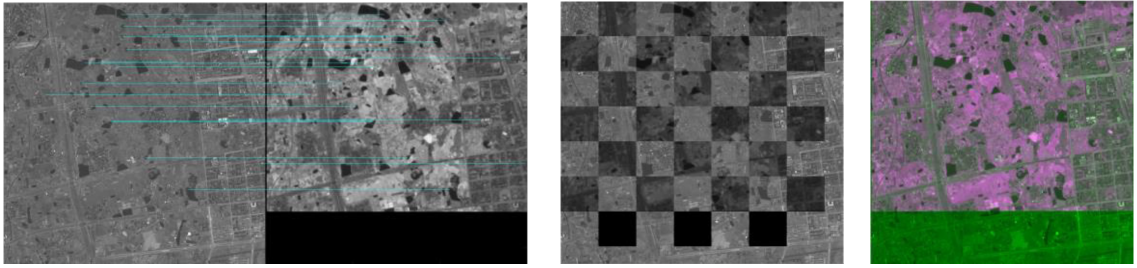
1) Statistical Analysis on RMSE Among State-of-the-Art Algorithms: The previous experiments validated the effectiveness of ISIFT. In this section, we performed experiments to evaluate the comparative analysis of ISIFT to four state-of-the-art methods, RANSAC [28], LLT [31], [55], LPM [32], [56], and PSSC [34], [57]. For fairness of comparison, SIFT was used for the four methods to extract feature points combined with SAD to obtain matching feature pairs. Since there was a wide range of RMSE, the plots in Fig. 10(a) and (b) only showed the RMSE less than 4 pixels, while all other results were grouped into one category ≥ 4 pixels.

The statistical RMSE histograms plotted in Fig. 10(a) and (b) show their comparative results of the five algorithms (IS_RIM, RANSAC, LLT, LPM, and PSSC) where the registration errors shown in Fig. 10(a) using Ω_A data sets with affine deformations shown in Fig. 10(a) had smaller RMSE than that using Ω_B with projective deformation shown in Fig. 10(b). As can be seen from Fig. 10, the number of images resulting from IS_RIM-registered error RMSE less



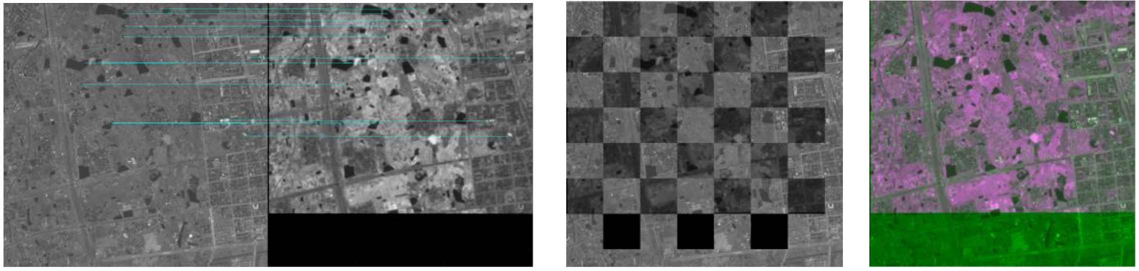
Feature point pairs pairwise linking Checkboard mosaicked images Red-green fusion images

RIRMI (RMSE=0.728)



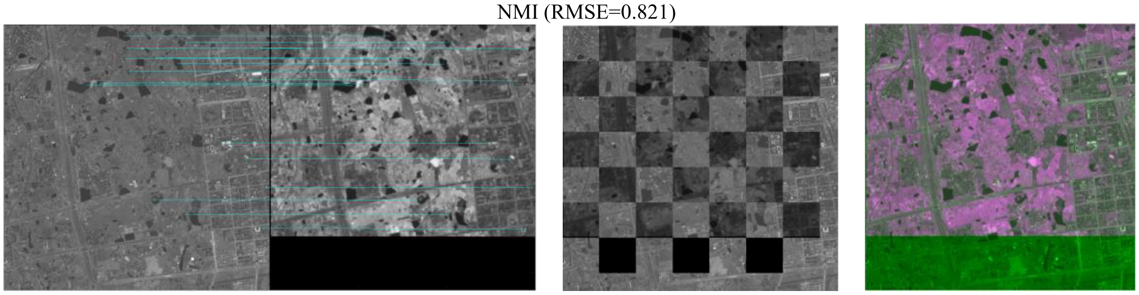
Feature point pairs pairwise linking Checkboard mosaicked images Red-green fusion images

RMI (RMSE=1.649)



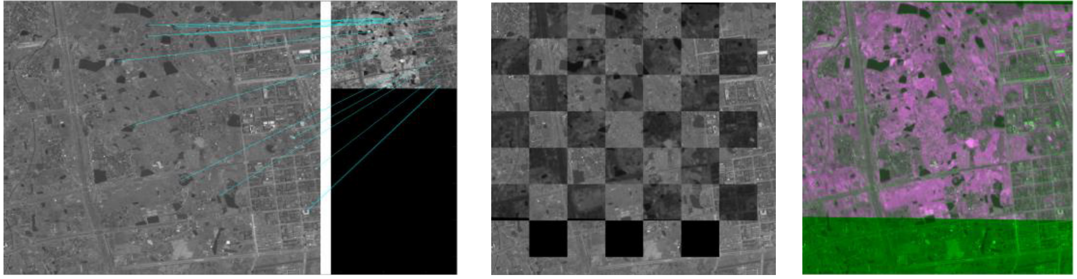
Feature point pairs pairwise linking Checkboard mosaicked images Red-green fusion images

NMI (RMSE=0.821)



Feature point pairs pairwise linking Checkboard mosaicked images Red-green fusion images

MI(RMSE=1.5137)

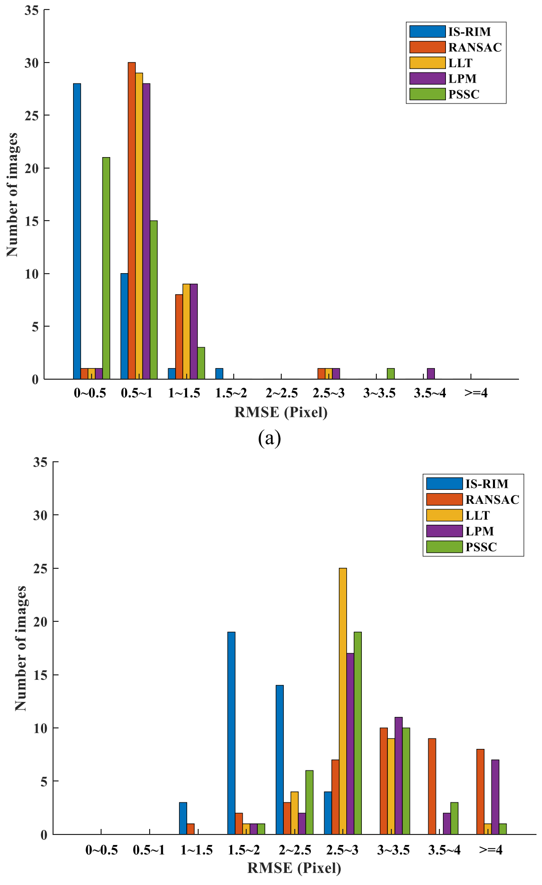


Feature point pairs pairwise linking Checkboard mosaicked images Red-green fusion images

SIFT (RMSE=2.4193)

Fig. 9. Matching keypoint images, red–green fusion images, and checkboard mosaicked images based on ISIFT (RIRMI, RMI, NMI, and MI) and SIFT in the final iteration for real image 1.

than 0.5 pixels was greatly larger than that produced by the other four methods. Moreover, the RMSEs of IS-RIM were almost smaller than 3 pixels. For example, using Ω_A with affine deformations, all the five algorithms could achieve subpixel precision in most of the cases. However, IS_RIM could achieve the best results less than 0.5 pixels followed by the second best PSSC in Fig. 10(a). Comparing the RMSEs in Fig. 10(b) to RMSEs in Fig. 10(a), using Ω_B data sets



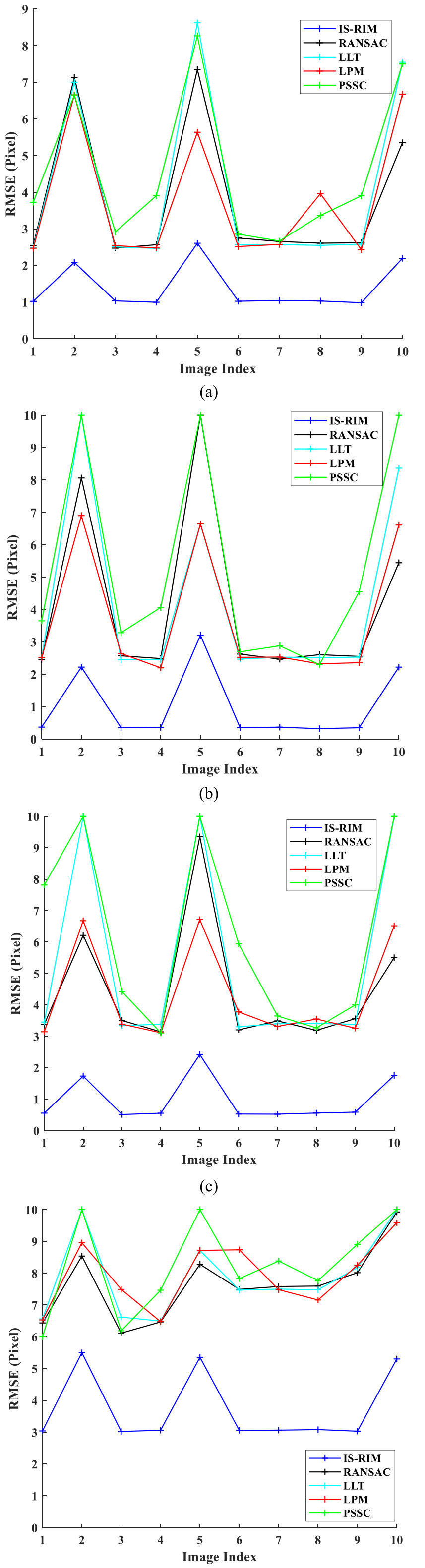
(a)

(b)

Fig. 10. Comparison of the final RMSEs using IS_RIM, RANSAC, LLT, LPM, and PSSC with different transformation. (a) Set-A data sets with affine deformation. (b) Set-B data sets with perspective deformation.

produced much greater RMSE that of using Ω_A. However, in comparison with other four algorithms IS-RIM performed significantly better than other four test algorithms since most images with RMSE produced by IS-RIM were less than 2.5 pixels, while four test algorithms had RSME centered around 2.5–3 pixels as shown in Fig. 10.

There are three reasons for that. First, there was a projective transformation between the image pairs in Ω_B. To register image pairs with a projective deformation, a feature-based algorithm may result in performance degradation in extracting stable features. Second, the spatial resolution of the images in Ω_B is relatively larger than that in Ω_A. The precise spatial locations of feature points of high-resolution images may have shifted by a few pixels in images with high spatial resolution but may not be detectable in images with low spatial resolution. Third, the size of the images in Ω_B was smaller than that used in Ω_A which might result in fewer features. This is because a small width and complex geometric transformation may yield large registration errors. However, even in this case, IS_RIM could still achieve better and stable registration results. These comparative experiments further demonstrated the effectiveness and superiority of IS_RIM to other four test methods for registering both satellite and aerial remote sensing images with affine or projective transformations.



(a)

(b)

(c)

(d)

Fig. 11. Comparison of the final RMSEs based on IS_RIM, RANSAC, LLT, LPM, and PSSC for Set-C data sets with different rotation. (a) 20° rotation. (b) 40° rotation. (c) 60° rotation. (d) 80° rotation.

2) Analysis on Changes in Different Rotations: Fig. 11(a)–(d) plots the final RMSEs produced by the five test algorithms using Ω_C with different rotations. Because the final registration results of all images had a wide range of RMSE, the results with an RMSE greater than 10 pixels were grouped into one category of 10 pixels. Fig. 11(a)–(c) shows the RMSE of IS-RIM resulting from 20°, 40°, and 60° where its registration errors were less than 2 pixels in most cases. When the rotation was 80°, the registration error was increased to 3 or 5 pixels as shown in Fig.11(d). Obviously, the registration errors of the image pairs 2, 5, and 10 were relatively larger than other image pairs. The reason was that these three image pairs were selected from $\boldsymbol{\Omega}_{\mathrm{B}}$ and its higher spatial resolution and smaller image size led to fewer precisely located feature points. Thus, these images incurred larger registration errors than the other seven image pairs. Overall, the registration error generally increased as the rotation change increased. However, compared with the other four algorithms, IS_RIM produced much smaller registration errors for all rotational deformations. These results further demonstrated the better performance and robustness of IS_RIM for different rotations.

VI. CONCLUSION

This article develops a close-feedback iterative SIFT registration system, called ISIFT. Unlike the traditional SIFT-based registration methods which are considered as feed forward open systems, ISIFT includes a rectification feedback loop to rectify and update the sensed image to improve the registration performance. The rectification loop allows ISIFT to select better registration parameters based on the maximal SV and spatial consistency to rectify the current sensed image and the resulting rectified image is then fed back to replace the current sensed image for the next round of SIFT implementation in an iterative manner. As expected, ISIFT performs better positional accuracy than other registration algorithm, including the RANSAC [28], LLT [31], LPM [32], [33], and PSSC [34]. The proposed close-feedback registration system is not necessarily limited to SIFT as a feature extraction method and can be also extended to other feature-based algorithms and similarity metrics to replace SIFT and MI-based similarity metrics, respectively.

Several contributions are summarized as follows. First, the proposed ISIFT using rectification feedback loops is a close-feedback registration system which has never been explored in the literature over the past years. Second, any SIFT-based system can be used as an initial condition to initialize ISIFT. With this interpretation, ISIFT can be regarded as a generalization of traditional SIFT-based methods without using feedback. Third, in addition to introducing an iterative process into SIFT, an automatic stopping rule is also custom-designed to terminate the iterative process carried out by ISIFT. Fourth, a joint spatial-consistency and intensity-similarity feature point selection method is further developed to select consistent feature points. Finally, extensive experiments are performed to conduct a comparative study and analysis among six methods, ISIFT using four different similarity metrics (MI, NMI, RMI, and RIRMI) and spatial consistency, SIFT with direct feedback loops (ISIFTD), and SIFT without feedbacks. The experimental results demonstrate that ISIFT indeed performed better than ISIFTD and SIFT without feedbacks.

Due to the use of an iterative process, the running time is relatively longer compared to other feature-based or area-based methods. However, due to its nature constraints, SIFT is only suitable for images with a certain scale change and illumination differences. Under some extreme situations, when there are no more than three or more pairs of correctly matched feature pairs among the initial SIFT feature pairs, other modified versions of SIFT or other feature extraction methods should be used to replace SIFT to extract local features. In this case, our proposed iterative strategy is still applicable. In addition, the location accuracy of local features also affects the accuracy of SIFT. Our proposed ISIFT provides improvements over SIFT. For future work, the local feature point locations will be modified to reduce the inherent error and improve the local matching accuracy.

APPENDIX

There are many similarity metrics that can be used for ISIFT to calculate the similarity between the reference image and rectified sensed image. In what follows, we describe four similarity metrics which can be used for this purpose where the four metrics are MI, NMI, RMI, and RIRMI.

A. Mutual Information

A classical similarity metric is the MI [58] given by

$$I(\mathbf{X},\mathbf{Y}) = H(\mathbf{X}) + H(\mathbf{Y}) - H(\mathbf{X},\mathbf{Y}) \tag{7}$$

where $\mathbf{X}$ is the reference image and $\mathbf{Y}$ is the sensed image. $I(\mathbf{X},\mathbf{Y})$ is the MI between $\mathbf{X}$ and $\mathbf{Y}$, $H(\mathbf{X})$ and $H(\mathbf{Y})$ denote the marginal entropy values of $\mathbf{X}$ and $\mathbf{Y}$, respectively. The joint entropy $H(\mathbf{X},\mathbf{Y})$ measures the amount of information in the overlapping region, and it is given by

$$H(\mathbf{X},\mathbf{Y}) = \sum_{x,y} -P_{X,Y}(x,y)\log_2 P_{X,Y}(x,y) \tag{8}$$

$$H(\mathbf{X}) = \sum_{x} -P_X(x)\log_2 P_X(x) \tag{9}$$

$$H(\mathbf{Y}) = \sum_{y} -P_Y(y)\log_2 P_Y(y) \tag{10}$$

where the joint probability distribution is $P_{X,Y}(x,y) = P(X_s = x, Y_s = y)$ for $x \in \Omega_X$, $y \in \Omega_Y$ and (X_s, Y_s) is the corresponding pixel pair in the overlapping region. $P_X(x)$ and $P_Y(x)$ are the marginal probability distributions of intensity.

B. Normalized MI

Another well-established registration similarity metric normalized MI [59] is given by

$$\mathrm{NMI} = \frac{H(\mathbf{X}) + H(\mathbf{Y})}{H(\mathbf{X},\mathbf{Y})} \tag{11}$$

which will more accurately reflect the change of parameters and make the registration function smoother. NMI can reduce the sensitivity of MI to overlapped parts of two images and improve robustness.

C. Region MI

By incorporating spatial information in the traditional MI, the RMI [60] measures statistical correlation between two regions. Different from MI, the RMI considers not only the original intensity information but also the intensity information with spatial dependency. The RMI can be represented by

$$\mathrm{RMI} = H_g(\mathbf{C}_A) + H_g(\mathbf{C}_B) - H_g(\mathbf{C}) \quad (12)$$

$$\mathbf{C} = \frac{1}{N}\mathbf{P}_0\mathbf{P}_0^T \quad (13)$$

$$\mathbf{P}_0 = \mathbf{P} - \frac{1}{N}\sum_{i=1}^{N} p_i \quad (14)$$

where $\mathbf{P} = [p_1, p_2, \ldots, p_N]$, $N = (m - 2r)(n - 2r)$ is the number of pixels to be counted in a region, the size of the overlapping region is $m \times n$, r is the radius of local window, and p_i is the pixel's neighborhood vector for the ith pixel. The size of P is $d \times N$ with $d = 2 \times r \times r$. The marginal entropies are $H_g(\mathbf{C}_A)$ and $H_g(\mathbf{C}_B)$, where C_A is the $(d/2) \times (d/2)$ matrix in the upper left of C and C_B is the $(d/2) \times (d/2)$ matrix in the lower left of C.

D. Rotation-Invariant Region MI

In order to obtain a more robust similarity metric, the RIRMI uses a rotation-invariant local ternary pattern to calculate robust intensity–spatial information correlation statistic. The detailed description was given in [38]. The RIRMI can be represented by

$$\mathrm{RIRMI}(\mathbf{X}, \mathbf{Y}) = \log_2\left(\frac{\det(\mathbf{C}_A)\det(\mathbf{C}_B)}{\det(\mathbf{C})}\right)^{1/2} \quad (15)$$

where $\mathbf{C}$ $\mathbf{C}_A$, and $\mathbf{C}_B$ are defined in the same way as defined in the RMI.

E. Comparison Between MI, NMI, RMI, and RIRMI

MI has been widely used in remote sensing [39]. Despite its outstanding performance, the MI-based methods provide a local maximum rather than a global maximum of the entire search space for the correct transformation [61]. As a result, a region of the search space should be predefined where the MI-based registration is implemented, which inevitably reduces its robustness. Thus, there are some limitations. According to the equation of MI, the probability is obtained by histogram statistic which is calculated by original intensity. It is likely that the overlapping region of MI calculation has the same histogram but with the different scene. The reason is that MI only considers original intensity statistic characteristic but ignores spatial information, which may lead to mismatch. In addition, due to the complexity of remote sensing image, similarity metric function (such as MI) may have many local extremums, which cannot correspond exactly to registration accuracy. The SV is not exactly consistent with the registration accuracy. Specifically, comparing the two SVs, large similarity does not exactly correspond to the more accurate registration results. Thus, the robustness and accuracy of similarity metric are important for measuring the registration accuracy. Although the similarity metric value cannot correspond to registration accuracy one-to-one, the more robust the metric is, the more accurate registration results can be measured. As can be seen from the example in this section, although the SV cannot completely correspond to the evaluated geometric difference, robust similarity metrics may have a tendency that the higher the registration accuracy is, the higher the SV will be, and the registration parameter corresponding to the largest similarity measure will be close to the optimal parameter.

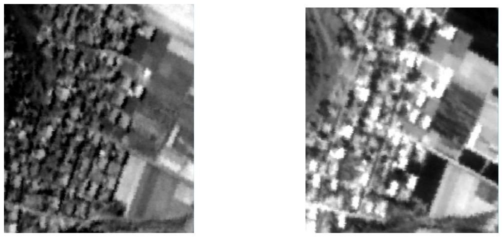

(a) (b)

Fig. 12. Image pairs of different bands for the same region. (a) 78th band spectral image. (b) 55th band spectral image.

In this section, four similarity metrics (MI, NMI, RMI, and RIRMI) are used to select consistent feature point sets and determine alignment parameters in each iteration. Following the basic theory in the previous section, the simulated comparison and analysis are given in this section based on given geometric parameters.

As shown in Fig. 12, there are image pairs (size: 201 × 201 pixels) which are acquired from different band images of the same region. For the image pairs with nonlinear intensity differences and deformation differences, the geometric parameters are set artificially. Then, consistent feature point sets are selected based on different similarity measures for comparison.

Because the similarity metric is used to measure the similarity of the corresponding overlapping region, the overlapping region has translation or rotation. In order to simulate the correspondence of SV and geometric difference, translation and translation–rotation simulation plots are drawn by translating or rotating the sensed image. In detail, the plot of SV versus translation is drawn as follows: a template rectangular window with a size of 100 × 50 is selected from the 78th band image; then, a rectangular window of the 55th band image slides horizontally from the center (101, 101) to the left and right with step 1 pixel, respectively. This process can simulate the translation mismatching among −25 to 25 pixels. Then MI, NMI, RMI, and RIRMI of the rectangular pairs in the reference image and sensed image are calculated and similarity plots are shown in Fig. 13. The plot of SV versus translation–rotation is drawn as follows: in order to simulate the rotation difference, the sensed image (55th band images) is rotated from −18° to 18°. In the overlapping region, the similarity measures MI, NMI, RMI, and RIRMI are calculated for the corresponding rectangular window with a size of 50 × 50. Although rotation also causes translation, the corresponding of rotation and SV are shown here. The similarity plots are shown in Fig. 14.

A robust similarity measure is one with a smooth, convex landscape with respect to mismatch, specifically, one that does not have too many local extrema approaching a global optimum. As can be seen in Figs. 13 and 14, NMI, RMI, and

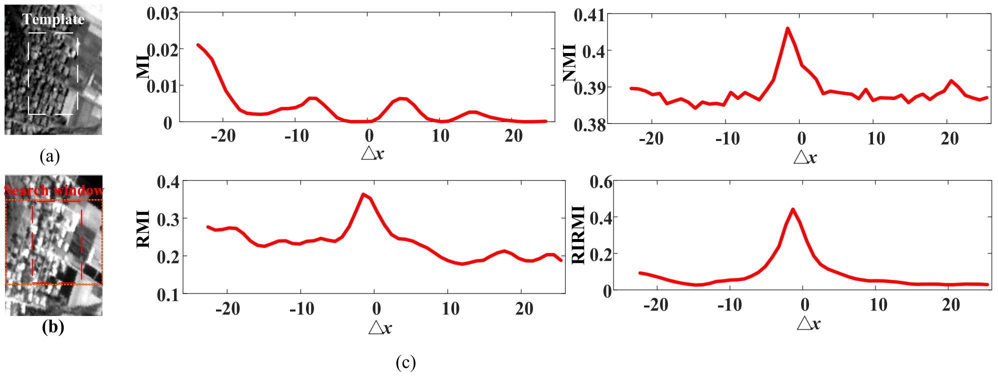


Fig. 13. Comparison among similarity plots of MI, RMI, RMI, and RIRMI for translation transform. (a) Subimage of the 78th band image. (b) Subimage of the 55th band image.

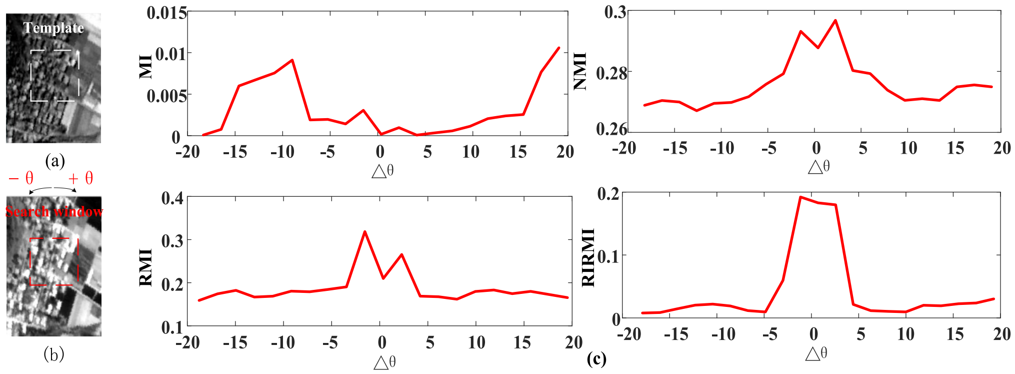


Fig. 14. Comparison among similarity plots of MI, RMI, RMI, and RIRMI for rotation transform with translation. (a) Subimage of the 78th band image. (b) Subimage of the 55th band image. (c) Similarity curves of MI, NMI, RMI, and RIRMI.

RIRMI have the correct change tendency, that is, maximum similarity metric value corresponds to the approximate correct position. Compared with RMI and NMI, RIRMI is not easy to be trapped in local extreme points. Thus, its value can still ensure that the maximum value can be obtained when the two translated or rotated images are matched. However, it is not true for MI. This implies that the RIRMI is more robust than RMI, NMI, and MI to the image pairs with nonlinear intensity differences and geometric distortions [38].

REFERENCES

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

[2] G. Chen, K. Zhao, and R. Powers, "Assessment of the image misregistration effects on object-based change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 19–27, Jan. 2014.

[3] F. Song *et al.*, "Multi-scale feature based land cover change detection in mountainous Terrain using multi-temporal and multi-sensor remote sensing images," *IEEE Access*, vol. 6, pp. 77494–77508, 2018.

[4] Y. Han, S. Jung, S. Liu, and J. Yeom, "Effect analysis in the fine co-registration of very-high-resolution satellite images for unsupervised change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Yokohama, Japan, Jul. 2019, pp. 1558–1561.

[5] B. Ayhan, M. Dao, C. Kwan, H.-M. Chen, J. F. Bell, and R. Kidd, "A novel utilization of image registration techniques to process mastcam images in mars rover with applications to image fusion, pixel clustering, and anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4553–4564, Oct. 2017.

[6] Q. Zhang, Z. Cao, Z. Hu, Y. Jia, and X. Wu, "Joint image registration and fusion for panchromatic and multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 467–471, Mar. 2015.

[7] Y. Zhou, A. Rangarajan, and P. D. Gader, "An integrated approach to registration and fusion of hyperspectral and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3020–3033, May 2020.

[8] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "DeepSUM: Deep neural network for super-resolution of unregistered multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3644–3656, May 2020.

[9] H. Chen, H. Zhang, J. Du, and B. Luo, "Unified framework for the joint super-resolution and registration of multiangle multi/hyperspectral remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2369–2384, 2020.

[10] B. Chaudhuri, B. Demir, L. Bruzzone, and S. Chaudhuri, "Region-based retrieval of remote sensing images using an unsupervised graph-theoretic approach," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 987–991, Jul. 2016.

[11] B. Chaudhuri, B. Demir, S. Chaudhuri, and L. Bruzzone, "Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1144–1158, Feb. 2018.

[12] X. Li, J. Yang, and J. Ma, "Large scale category-structured image retrieval for object identification through supervised learning of CNN and SURF-based matching," *IEEE Access*, vol. 8, pp. 57796–57809, 2020.

[13] B. Qu, X. Li, D. Tao, and X. Lu, "Deep semantic understanding of high resolution remote sensing image," in *Proc. Int. Conf. Comput., Inf. Telecommun. Syst.*, Jul. 2016, pp. 124–128.

[14] Z. Shi and Z. Zou, "Can a machine generate humanlike language descriptions for a remote sensing image?" *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3623–3634, Jun. 2017.

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

[16] 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.

[17] S. Paul and U. C. Pati, "Remote sensing optical image registration using modified uniform robust SIFT," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1300–1304, Sep. 2016.

[18] Y. Xiang, F. Wang, and H. You, "OS-SIFT: A robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3078–3090, Jun. 2018.

[19] H.-H. Chang, G.-L. Wu, and M.-H. Chiang, "Remote sensing image registration based on modified SIFT and feature slope grouping," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1363–1367, Sep. 2019.

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

[21] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

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

[23] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.

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

[25] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, "A partial intensity invariant feature descriptor for multimodal retinal image registration," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1707–1718, Jul. 2010.

[26] A. Sedaghat and H. Ebadi, "Distinctive order based self-similarity descriptor for multi-sensor remote sensing image matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 62–71, Oct. 2015.

[27] A. Sedaghat and N. Mohammadi, "Illumination-robust remote sensing image matching based on oriented self-similarity," *ISPRS J. Photogramm. Remote Sens.*, vol. 153, pp. 21–35, Jul. 2019.

[28] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[29] Z. Song, S. Zhou, and J. Guan, "A novel image registration algorithm for remote sensing under affine transformation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4895–4912, Aug. 2014.

[30] 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 Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 43–47, Jan. 2015.

[31] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.

[32] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, May 2019.

[33] J. Ma, J. Zhao, H. Guo, J. Jiang, H. Zhou, and Y. Gao, "Locality preserving matching," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Aug. 2017, pp. 4492–4498.

[34] Y. Ma, J. Wang, H. Xu, S. Zhang, X. Mei, and J. Ma, "Robust image feature matching via progressive sparse spatial consensus," *IEEE Access*, vol. 5, pp. 24568–24579, 2017.

[35] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[36] J. Liang, X. Liu, K. Huang, X. Li, D. Wang, and X. Wang, "Automatic registration of multisensor images using an integrated spatial and mutual information (SMI) metric," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 603–615, Jan. 2014.

[37] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer, "Image similarity using mutual information of regions," in *Proc. Eur. Conf. Comput. Vis.*, vol. 3023, pp. 596–607, 2004.

[38] S. Chen, X. Li, L. Zhao, and H. Yang, "Medium-low resolution multisource remote sensing image registration based on SIFT and robust regional mutual information," *Int. J. Remote Sens.*, vol. 39, no. 10, pp. 3215–3242, May 2018.

[39] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, "A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4328–4438, Oct. 2014.

[40] L.-Y. Zhao, B.-Y. Lü, X.-R. Li, and S.-H. Chen, "Multi-source remote sensing image registration based on scale-invariant feature transform and optimization of regional mutual information," *Acta Phys. Sinica*, vol. 64, no. 12, 2015, Art. no. 124204.

[41] Y. Ye and J. Shan, "A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences," *ISPRS J. Photogramm. Remote Sens.*, vol. 90, no. 3, pp. 83–95, Apr. 2014.

[42] 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. 2008.

[43] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA, USA: Kluwer, 2003.

[44] L. X. Li, "The registration and fusion of multi-sensor images based on mutual information," M.S. thesis, Dept., Cir. Sys., Univ. Electron. Sci. Technol. China, Chengdu, China, 2013.

[45] 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.

[46] P. Navy, V. Page, E. Grandchamp, and J. Desachy, "Matching two clusters of points extracted from satellite images," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 268–274, Mar. 2006.

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

[48] 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.

[49] *UC Merced Land Use Dataset*. Accessed: Dec. 5, 2018. [Online]. Available: http://140.112.27.140/wp-content/uploads/2018/12/Datasets.zip

[50] USGS. *Earth Explorer*. Accessed: Aug. 28, 2016. [Online]. Available: https://earthexplorer.usgs.gov/

[51] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (ACM GIS)*, 2010, pp. 270–279.

[52] *High-Resolution Orthoimagery Data*. Accessed: Oct. 28, 2010. [Online]. Available: http://weegee.vision.ucmerced.edu/datasets/landuse.html

[53] *ENVI Image Analysis Software (5.2 Version)*. L3HARRIS Geospatial. Accessed: Oct. 15, 2014. [Online]. Available: https://www.harrisgeospatial.com/Software-Technology/ENVI

[54] H. Yang, X. Li, L. Zhao, and S. Chen, "A novel coarse-to-fine scheme for remote sensing image registration based on SIFT and phase correlation," *Remote Sens.*, vol. 11, no. 15, p. 1833, Aug. 2019.

[55] *Matlab Code of LLT*. Accessed: Sep. 29, 2018. [Online]. Available: https://github.com/jiayi-ma/LLT

[56] *Matlab Code of LPM*. Accessed: Mar. 16, 2019. [Online]. Available: https://github.com/jiayi-ma/LPM

[57] *Matlab Code of PSSC*. Accessed: Dec. 31, 2018. [Online]. Available: https://github.com/JiahaoPlus/PSSC

[58] H.-M. Chen, P. K. Varshney, and M. K. Arora, "Performance of mutual information similarity measure for registration of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2445–2454, Nov. 2003.

[59] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, Jan. 1999.

[60] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer, "Image similarity using mutual information of regions," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 596–607.

[61] K. Johnson, A. Cole-Rhodes, I. Zavorin, and J. Le Moigne, "Mutual information as a similarity measure for remote sensing image registration," in *Proc. SPIE*, 2001, pp. 51–61.

Shuhan Chen (Graduate Student Member, IEEE) received the B.S. degree from Ludong University, Yantai, China, in 2011 and the M.S. degree from Liaoning Technical University, Huludao, China, in 2014. She is pursuing the Ph.D. degree in control theory and control engineering from Zhejiang University, Hangzhou, China.

She is doing research as a Visiting Faculty Research Assistant with Remote Sensing Signal and Image Proessing Laboratory (RSSIPL), Department of Computer Science and Electrical Engineering, University Maryland, Baltimore County (UMBC), Baltimore, MD, USA. Her research interests include image registration, hyperspectral image processing, and pattern recognition.

Shengwei Zhong received the B.E. degree in information countermeasure technology, the M.S. and Ph.D. degrees in electronics and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2013, 2015, and 2020, respectively.

She was an exchange Ph.D. Student visiting Remote Sensing Signal and Image Proessing Laboratory (RSSIPL) as a Faculty Research Assistant at the University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA. She is a Post-Doctoral Researcher and a Lecturer with the Nanjing University of Science and Technology, Nanjing, China. Her research interests include hyperspectral image processing, remote sensing image fusion, and applications.

Bai Xue (Member, IEEE) received the B.E. degree in automation from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2015 and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland Baltimore County (UMBC), Baltimore, MD, USA, in 2019.

From 2014 to 2015, he was a Program Exchange Undergraduate Student visiting the Department of Electrical and Computer Engineering, University of Detroit Mercy (UDM), Detroit, MI, USA. He was also a Visiting Research Assistant with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Dalian Maritime University (DMU), Dalian, China, from 2016 to 2017. He is a Research Associate with Remote Sensing Signal and Image Processing Laboratory (RSSIPL), Department of Computer Science and Electrical Engineering (CSEE), College of Engineering and Information Technology, UMBC. His research interests include multispectral/hyperspectral imaging, pattern recognition, medical imaging, and medical data analysis.

Xiaorun Li (Member, IEEE) received the B.S. degree from the National University of Defense Technology, Changsha, China, in 1992 and the M.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1995 and 2008, respectively.

Since 1995, he has been with Zhejiang University, where he is a Professor with the College of Electrical Engineering. His research interests include hyperspectral image processing, signal and image processing, and pattern recognition.

Liaoying Zhao (Member, IEEE) received the B.S. and M.S. degrees from Hangzhou Dianzi University, Hangzhou, China, in 1992 and 1995, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, in 2004.

Since 1995, she has been with Hangzhou Dianzi University, where she is a Professor with the College of Computer Science. Her research interests include hyperspectral image processing, signal and image processing, pattern recognition, and machine learning.

Chein-I Chang (Life Fellow, IEEE) received the B.S. degree in mathematics from Soochow University, Taipei, Taiwan, in 1973, the M.S. degree in mathematics from the Institute of Mathematics, National Tsing Hua University, Hsinchu, Taiwan, in 1975, the M.A. degree in mathematics from the State University of New York at Stony Brook, Stony Brook, NY, USA, in 1977, the M.S. and M.S.E.E. degrees from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 1982, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 1987.

He has been with the University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA, since 1987 and is a Professor with the Department of Computer Science and Electrical Engineering. He has been holding a Chang Jiang Scholar Chair Professorship and the Director of the Center for Hyperspectral Imaging in Remote Sensing (CHIRS) at Dalian Maritime University, Dalian, China, since 2016. In addition, he is also a Chair Professor with National Chiao Tung University, Hsinchu, Taiwan, since 2019. He has authored four books, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* published by Kluwer Academic Publishers in 2003 and *Hyperspectral Data Processing: Algorithm Design and Analysis*, John Wiley & Sons, 2013, *Real Time Progressive Hyperspectral Image Processing*: *Endmember Finding and Anomaly Detection* 2016 by Springer, and *Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation*, Springer 2017. In addition, he also edited two books, *Recent Advances in Hyperspectral Signal and Image Processing*, 2006, and *Hyperspectral Data Exploitation: Theory and Applications*, John Wiley & Sons, 2007 and co-edited with A. Plaza a book on *High Performance Computing in Remote Sensing*, CRC Press, 2007. He holds seven patents on hyperspectral image processing. His research interests include multispectral/hyperspectral image processing, automatic target recognition, and medical imaging.

Dr. Chang is a fellow of SPIE. He received the National Research Council Senior Research Associateship Award from 2002 to 2003 sponsored by the U.S. Army Soldier and Biological Chemical Command, Edgewood Chemical and Biological Center, Aberdeen Proving Ground, Maryland. He was a Plenary Speaker for the Society for Photo-optical Instrumentation Engineers (SPIE) Optics+Applications, Remote Sensing Symposium in 2009. He was the Guest Editor of a special issue of the *Journal of High Speed Networks* on Telemedicine and Applications in April 2000 and a Co-Guest Editor of another special issue of the same journal on the Broadband Multimedia Sensor Networks in Healthcare Applications in April 2007. He is also the Co-Guest Editor of special issues on High Performance Computing of Hyperspectral Imaging for the *International Journal of High Performance Computing Applications* in December 2007, Signal Processing and System Design in Health Care Applications for the *EURASIP Journal on Advances in Signal Processing* in 2009, Multispectral, Hyperspectral, and Polarimetric Imaging Technology for the *Journal of Sensors* in 2016, and Hyperspectral Imaging and Applications for the *Remote Sensing* in 2018.