A Novel Coarse-to-Fine Scheme for Automatic Image Registration Based on SIFT and Mutual Information

Maoguo Gong, *Member, IEEE*, Shengmeng Zhao, Licheng Jiao, *Senior Member, IEEE*, Dayong Tian, and Shuang Wang, *Member, IEEE*

***Abstract*—Automatic image registration is a vital yet challenging task, particularly for remote sensing images. A fully automatic registration approach which is accurate, robust, and fast is required. For this purpose, a novel coarse-to-fine scheme for automatic image registration is proposed in this paper. This scheme consists of a preregistration process (coarse registration) and a fine-tuning process (fine registration). To begin with, the preregistration process is implemented by the scale-invariant feature transform approach equipped with a reliable outlier removal procedure. The coarse results provide a near-optimal initial solution for the optimizer in the fine-tuning process. Next, the fine-tuning process is implemented by the maximization of mutual information using a modified Marquardt–Levenberg search strategy in a multiresolution framework. The proposed algorithm is tested on various remote sensing optical and synthetic aperture radar images taken at different situations (multispectral, multisensor, and multitemporal) with the affine transformation model. The experimental results demonstrate the accuracy, robustness, and efficiency of the proposed algorithm.**

***Index Terms*—Image registration, mutual information (MI), outlier removal, scale-invariant feature transform (SIFT).**

I. Introduction

IMAGE registration is a vital process which determines the most precise match between two images of the same scene, which may have been acquired at the same or different times, by the same or different sensors, and from the same or different viewpoints [1]. This process geometrically aligns the two images—the reference and sensed images. Image registration has been widely used in many fields such as computer vision, pattern matching, medical image analysis, and remote sensing image processing.

Regarding remote sensing applications in particular, it is highly required to achieve an accurate match and to obtain precise geometric correction. For applications such as image fusion and change detection, reaching a subpixel accuracy is of great importance. For example, a registration accuracy of less than 0.2 of a pixel is demanded to achieve a change detection error of less than 10% [2]. Automatic image registration provides a practical means of achieving this goal. Although many algorithms have been recently proposed [3]–[6], automatic image registration is still a challenge due to the presence of particular difficulties within the remote sensing field. The difficulties involved mainly include both geometric deformations (translation effect, rotation and scale distortion, occlusion, and viewpoint difference) and radiometric discrepancies (illumination change and sensor and spectral content difference). Therefore, further research studies are required in order to improve the performance of the existing registration methods.

Image registration methods can be broadly classified into two categories: intensity- and feature-based methods [1], [7]. Feature-based methods initially extract salient features and then match them using similarity measures to establish the geometric correspondence between two images. One of the main advantages of these approaches is that they are fast and robust to noises, complex geometric distortions, and significant radiometric differences. However, they just perform well on the condition that suitable features are extracted and matched by reliable algorithms. The commonly used features include point, edge, contour, and region, and the well-known feature matching methods include invariant descriptor, spatial relation, and relaxation methods [1]. Mikolajczyk *et al.* [8] compared the performance of descriptors for affine transformation, scale change, rotation, image blur, jpeg compression, and illumination change. They found that scale-invariant feature transform (SIFT) [9] performs best for most of the tests.

SIFT is capable of extracting distinctive invariant features from images, and it can be applied to preform reliable matching across a substantial range of affine distortion, change in 3-D viewpoint, addition of noise, and change in illumination [9]. Despite the attractive advantages of SIFT, there exist some problems when it is directly applied to remote sensing images, i.e., the number of the detected feature matches may be small, and their distribution may be uneven due to the complex content nature of remote sensing images [10]. In addition, many outliers exist in feature matches on account of significant differences on the image intensity between the overlay regions of remote sensing images. Therefore, using SIFT alone cannot produce optimal results, as supported by the recent algorithms applying SIFT to remote sensing images [11]–[15]. These algorithms improved the SIFT approach from different aspects. Image segmentation and maximally stable extremal region were utilized to detect more reliable support regions which are used to compute invariant descriptors [12], [15]. In [13] and [14],

Manuscript received March 28, 2013; revised June 27, 2013 and July 31, 2013; accepted September 4, 2013. Date of publication October 3, 2013; date of current version February 27, 2014. This work was supported by the National Natural Science Foundation of China under Grant 61273317, the National Top Youth Talents Support Program of China, and the Fundamental Research Fund for the Central Universities under Grants K50510020001 and K5051202053.

The authors are with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, Xidian University, Xi'an 710071, China (e-mail: gong@ieee.org).

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

Digital Object Identifier 10.1109/TGRS.2013.2281391

correct matching rate increases by introducing novel matching criterions.

Ma *et al.* [11] developed a two-step nonrigid automatic registration scheme for multiangle satellite images. In [11], SIFT control points (CPs) were first selected in a preregistration process, and normalized cross correlation (NCC) CPs were then detected in the preliminary registered image to obtain more CPs with a better spatial distribution. The algorithm performs well for multiangle images [11] whose gray-level characteristics are rather similar. However, it is not suitable for more complicated situations because NCC is sensitive to the intensity changes introduced by varying illumination and noise and by using different sensors [1]. A better choice is local least squares matching with local radiometric adaptation [16]. In order to match precisely multisensor and/or multitemporal images with significant noise and illumination changes, a more suitable technique needs to be further studied.

In contrast to NCC, the mutual information (MI) criterion is particularly suitable for multisensor image registration [1]. MI, which represents a measure of statistical dependence between two images, is one of the most commonly used similarity measures in intensity-based methods. It has been demonstrated in [17] that MI is more robust to noise and produces consistently sharper peaks at the correct registration values than correlation, which is very important for accurate registration. Moreover, MI is also robust to nonlinear intensity relationship between images [18]. It has been successfully applied to the registration of medical imagery [19], [20] and remote sensing imagery [17], [18], [21]–[23]. Considering its robustness and high accuracy, we use it to refine the coarse results obtained by the SIFT approach as a preregistration step. Cole-Rhodes *et al.* [17] proposed a simultaneous perturbation stochastic approximation (SPSA) search strategy instead of exhaustive search to accelerate maximization of MI, which is applied in a multiresolution framework based on wavelet-like pyramid decomposition. However, SPSA is a less efficient search strategy to refine the coarse results obtained by the preregistration process because it converges slowly when close to the solution as described in [17], and the fact is that the coarse results are generally in the neighborhood of the solution. Thévenaz *et al.* [20] developed a powerful optimizer for maximization of MI for medical image registration. This powerful optimizer benefits strongly from good initial conditions that it exhibits superlinear convergence when close enough to the solution (ground truth) and reaches a high precision. Here, we explore its good performance and make a modification for the registration of remote sensing imagery.

The advantages of both SIFT and MI have also been employed through different schemes in [24]–[27]. Suri *et al.* [24] proposed a robust multisensor synthetic aperture radar (SAR) image registration method that estimated first the rough registration parameters using MI, and then, the parameters were introduced for conjugate feature selection during the SIFT matching phase to increase the number of correct matches. Heo *et al.* [25] proposed a new stereo matching method that combined MI as an appearance measure and the SIFT descriptor as a geometric measure. Heo *et al.* [26] introduced a robust stereo matching method that produces accurate depth maps by combining MI, SIFT, plane-fitting, and log-chromaticity color space. Moreover, Mekky *et al.* [27] presented a hybrid registration technique for dental panoramic X-ray images using wavelet-based hierarchical approach that applied MI-based registration at low-resolution level and utilized SIFT-based registration at high-resolution level with rough parameters obtained from MI. The novel coarse-to-fine registration scheme that we proposed will be described in detail in Section III.

The main innovation of this study is a new coarse-to-fine transformation parameter solving strategy, which comprises a preregistration process and a fine-tuning process. Its uniqueness lies in the following three aspects.

1) A modified outlier removal procedure is introduced to eliminate most false SIFT matches to guarantee the preregistration results close to the solution (ground truth).
2) An excellent initial solution selection strategy for maximization of MI is developed by using SIFT with outlier removal.
3) Maximization of MI using a modified Marquardt–Levenberg search strategy in a multiresolution framework is utilized to refine the preregistration results to achieve the most precise registration results.

The proposed algorithm employs the advantages of both feature-based methods and intensity-based ones, namely, the robustness of SIFT and the accuracy of MI. We tested the proposed algorithm on a variety of optical and SAR remote sensing images taken at different situations (multispectral, multisensor, and multitemporal) with the affine transformation model. Experimental results show that the proposed approach is robust, fast, and precise.

The rest of this paper is organized as follows. Section II interprets the motivation of each process of the proposed algorithm. Section III describes the proposed algorithm in detail. The experimental results on four different pairs of images are illustrated in Section IV. Concluding remarks are presented in Section V.

II. Motivation

The proposed registration algorithm, as mentioned previously, involves a preregistration process and a fine-tuning process. In this section, we interpret the motivation of each process of the proposed algorithm. First, we describe the reasonableness and superiority of using the SIFT approach equipped with a reliable outlier removal procedure as the preregistration process. Moreover, the benefits of taking the preregistration results as the initial solution of the optimizer in the fine-tuning process are presented. Finally, we state the advantages of utilizing maximization of MI as the fine-tuning process.

A. Motivation of Using SIFT With Outlier Removal as the Preregistration Process

When the two images to be registered are acquired by different sensors or from different viewing angles, geometric distortions may be significant [11], [14]. In this case, some commonly used registration methods such as cross correlation and MI cannot work well [1], [3]. Therefore, a preregistration process is desirable to alleviate the effects of significant geometric distortions, and the SIFT approach is an excellent choice due to its attractive properties mentioned before. An example is presented in [11], where SIFT was used as the preregistration process to generate an intermediate sensed image which has similar geometric characteristics as the reference image, and with this

(a) (b)

Fig. 1. Matches found (a) using distance ratio and (b) after outlier removal using the modified method.

TABLE I
COMPARISON OF THE PERFORMANCES OF THE OUTLIER REMOVAL METHODS

Method	No. of all correct/false matches	No. of remaining correct/false matches	Correct rate
Hough transform	244/56	180/8	70.5%
RANSAC	244/56	183/1	74.6%
Denser region	244/56	239/9	94.3%
Modified	244/56	241/4	97.1%

process, the subsequent NCC matching performed was accurate and easy. On the condition that the initial image pairs have been coarsely registered, the preregistration process is not necessary.

However, when the SIFT approach is used alone for the registration of remote sensing images, many false SIFT matches arise. These false matches will lead to further incorrect geometric correction. Therefore, a reliable outlier removal procedure is needed to ensure correct registration. In this paper, we adopt a modified outlier removal method where a scale histogram is computed and the denser cluster in the scale histogram corresponds to the true scale difference between the images. Its implementation is described in detail in Section III-B. An example is shown in Fig. 1, where the number of correct and false matches can be calculated statistically through visual inspection. A set of ten image pairs [two pairs of multisensory images (SPOT–Landsat Thematic Mapper), three pairs of multiview images (CHRIS/Proba–CHRIS/Proba), and five pairs of multitemporal images (Radarsat-2–Radarsat-2 and JERS-1–JERS-1)] is tested, and the total numbers of correct and false matches are counted, respectively. An excellent outlier removal method should be able to eliminate most false matches and preserve most correct matches as well. Considering this property, an evaluation criterion similar to the one proposed in [8] is used and computed as

$$\text{correct rate} = \frac{\#\text{remaining correct matches} - \#\text{remaining false matches}}{\#\text{all correct matches}}. \tag{1}$$

The experimental results presented in Table I demonstrate that the modified method is more reliable than other popular methods such as Hough transform [9], [28], RANSAC [29], and denser region [12] because it is based on the global-scale difference between the images and the SIFT keypoints that is scale invariant. This method is capable of removing most of the false SIFT matches on the tested data sets.

B. Motivation of Taking the Preregistration Process as the Initial Solution Selection Strategy

Few works have been done on the selection of initial solution that guarantees the convergence of the optimizers in the prior MI-based methods. In [17], the initial solution was determined by exhaustive search, which is impractical. In [18], correlation length was used to estimate the range of attractive region, and this information was utilized to initialize the solution within attractive region, but it still requires *a priori* knowledge on ground truth and human assistance. The SIFT approach equipped with a reliable outlier removal procedure is a promising initial solution selection strategy for three reasons: 1) its obtained coarse results are in the neighbor of optimal solution (see Table II) so that the optimizer in the fine-tuning process converges fast; 2) it is efficient and fully automatic without human assistance; and 3) the fine-tuning process is robust to a variation of the distance ratio using the SIFT matching and a small number of correct matches obtained in the preregistration process as demonstrated by the following sensitivity analysis.

The SIFT matching in the preregistration process is performed through the nearest neighbor approach. An effective measure for matching is the ratio between the distance of the nearest neighbor and that of the second nearest neighbor, defined as d_{ratio}. The sensitivity measures as a function of d_{ratio} regarding the registration of the pair of images described in Section IV-C are shown in Fig. 2, where the root-mean-square error (rmse) is defined as [30]

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(rx_i^2 + ry_i^2)} \tag{2}$$

where n is the number of matches and (rx_i, ry_i) is a residual for a certain match.

As shown in Fig. 2, for $d_{\text{ratio}} = 0.4$ in particular, although a small number of six matches have been obtained, a subpixel accuracy and a maximum of MI have also been achieved, which indicates that this strategy can provide a reliable initial solution even when a few correct matches are available (a common case for remote sensing images). As shown in Fig. 2(b), the values of rmse and MI are basically constant for values of d_{ratio} ranging from 0.4 to 0.85, which reinforces the robustness of the proposed algorithm. Note that similar results have been

TABLE II
REGISTRATION PARAMETERS (a_{11}, a_{12}, a_{21}, a_{22}, δ_x, AND δ_y), CORRESPONDING VALUES OF MI AND RMSE, AND ASSOCIATED COMPUTATION TIME OBTAINED BY MANUAL REGISTRATION [32], PREREGISTRATION, IS-SIFT [12], SPSA [16], AND THE PROPOSED ALGORITHM FOR THE FOUR DIFFERENT PAIRS OF IMAGES DESCRIBED IN SECTIONS IV-A–D

Pair	Method	a_{11}	a_{12}	a_{21}	a_{22}	δ_x	δ_y	MI	RMSE	Time(sec)
IV-A	Ground Truth	0.9659	-0.2588	0.2588	0.9659	-7.000	250.000	3.470	0.014	
	Manual	0.9657	-0.2583	0.2590	0.9659	-7.031	250.098	3.386	0.147	
	Pre-registration	0.9661	-0.2563	0.2593	0.9669	-7.147	249.930	1.796	0.247	45
	IS-SIFT	0.9648	-0.2586	0.2583	0.9662	-7.019	249.878	3.370	0.132	78
	SPSA	0.9662	-0.2579	0.2579	0.9662	-7.133	250.260	3.210	0.148	92
	Proposed	**0.9658**	**-0.2588**	**0.2588**	**0.9659**	**-6.990**	**249.994**	**3.467**	**0.031**	**61**
IV-B	Manual	1.0009	-0.0198	0.0217	1.0002	-76.737	-11.065	0.521	0.567	
	Pre-registration	1.0058	-0.0216	0.0092	0.9947	-77.329	-7.079	0.377	1.528	3
	IS-SIFT	1.0009	-0.0180	0.0186	0.9954	-77.001	-9.522	0.543	0.795	14
	SPSA	0.9999	-0.0123	0.0123	0.9999	-77.144	-9.137	0.549	0.925	28
	Proposed	**1.0039**	**-0.0209**	**0.0181**	**0.9983**	**-76.565**	**-10.335**	**0.576**	**0.473**	**11**
IV-C	Manual	0.9654	-0.2600	0.2615	0.9652	-44.471	11.629	0.928	0.615	
	Pre-registration	0.9337	-0.2518	0.2700	0.9611	-40.076	10.078	0.611	4.259	12
	IS-SIFT	0.9660	-0.2567	0.2613	0.9645	-45.563	11.832	0.910	0.825	57
	SPSA	0.9648	-0.2628	0.2628	0.9648	-43.771	10.726	0.911	1.479	94
	Proposed	**0.9662**	**-0.2583**	**0.2595**	**0.9656**	**-45.598**	**12.411**	**0.936**	**0.215**	**32**
IV-D	Manual	0.9390	-0.3421	0.3422	0.9395	1092.813	-573.624	0.564	0.924	
	Pre-registration	0.9358	-0.3399	0.3448	0.9294	1103.120	-565.132	0.445	9.443	291
	IS-SIFT	0.9395	-0.3415	0.3420	0.9406	1091.008	-574.004	0.503	1.025	955
	SPSA	0.9405	-0.3416	0.3416	0.9405	1088.224	-574.236	0.487	1.914	1286
	Proposed	**0.9391**	**-0.3425**	**0.3419**	**0.9398**	**1093.67**	**-573.060**	**0.604**	**0.557**	**837**

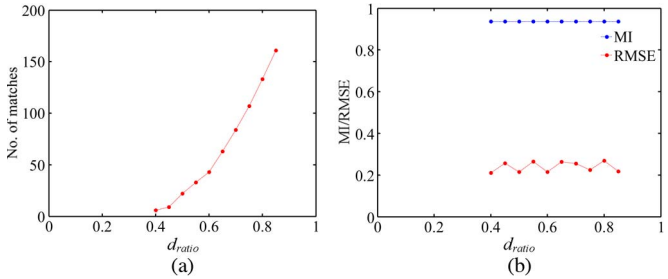


Fig. 2. Sensitivity analysis of the distance ratio in (a) the number of correct matches obtained by the preregistration process and (b) the final values of MI and rmse regarding the registration of the pair of images described in Section IV-C.

obtained for other test images, and a wider convergence range of d_{ratio} has been achieved when there exist more initial correct matches. In what follows, the parameter d_{ratio} is set to 0.6, which is appropriate for the preregistration process to obtain a meaningful initial solution in general. The proposed initial solution selection strategy is also suitable for other MI-based methods.

C. Motivation of Utilizing Maximization of MI to Refine the Preregistration Results

MI represents a measure of statistical dependence between two data sets. In the context of image registration, the MI value of two images is maximal when these two images are perfectly aligned. Therefore, MI can be used as a similarity measure whose maximum indicates the best match between the two images to be registered, and research studies have found that it is particularly suitable for multimodal image registration [17], [19], [21]. In this paper, we adopt it to refine the preregistration results in the fine-tuning process for three reasons. First, MI makes no limiting assumptions on the relation between the corresponding image intensities in different modalities and imposes no constraints on the image content of the modalities involved [19]. Second, MI is robust to noise. Third, MI produces consistently sharp peaks at the correct registration values, which is of great significance for precise registration [17].

The maximization of MI preferably performs in a multiresolution framework to achieve high accuracy and computational efficiency [31], [32]. In the multiresolution framework, most of the iterations execute at the coarsest level where the number of pixels is small due to image decomposition, which significantly reduces computation cost. Another very important advantage is that it strengthens the robustness of optimizers. The underlying reason is that it improves the smoothness of the MI surface so that the optimizers can avoid getting trapped in local optimum.

A multiresolution strategy is successful when the optimizer involved takes particular advantage of good initial conditions [20]. Note that a good initial solution has been achieved through the preregistration process, and we now require a most suitable optimizer that perfectly satisfies the mentioned constraint. All optimizers are not equally suitable, and the best choices are those that can converge in a few criterion evaluations when initialized with good initial conditions [20]. The specifically designed optimizer for the MI criterion in [20] exhibits superlinear convergence when close enough to the solution, a desired property. This optimizer combines the efficiency of the Newton method with the robustness of the gradient method. Here, considering its excellent performance, we make a modification

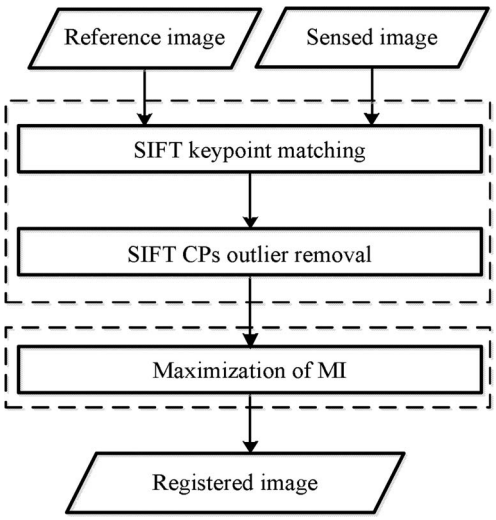


Fig. 3. Main steps of the proposed coarse-to-fine scheme for image registration.

of this optimizer and then apply it to register remote sensing images.

III. Methodology

The purpose of image registration is to find the corresponding geometric relationship between the two coordinate systems of two images. In this context, image registration can be defined as follows. Given a pair of 2-D gray-level images between which there exist some geometric and radiometric differences, let $f_R(x, y)$ and $f_S(x, y)$ represent the reference and sensed images, respectively, where coordinates $(x, y) \in \Delta \subset \mathbb{R}^2$ and Δ is a region of interest. To register these two images is to find the optimal geometric transformation $T_{\boldsymbol{\mu}}(\cdot)$ by which $f_S(T_{\boldsymbol{\mu}}(x, y))$ best matches $f_R(x, y)$ for all (x, y), where $\boldsymbol{\mu}$ is a set of transformation parameters. Here, we select the affine transformation model, which is widely used in the registration of remote sensing images, and it can be written as

$$T_{\boldsymbol{\mu}}(x, y) = \begin{bmatrix} a_{11} & a_{12} & \delta_x \\ a_{21} & a_{22} & \delta_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

where the transformation origin is considered to be the upper left corner of the reference image, $(a_{11}, a_{12}, a_{21}, a_{22})$ represent the rotation, scale, and shear differences, and (δ_x, δ_y) are the shifts between the two images.

The main steps of the new methodology are shown in Fig. 3. It starts with a preregistration process using SIFT with outlier removal. The preregistration results are used as a good initial solution and are then refined by maximization of MI in a multiresolution framework. Through this coarse-to-fine scheme, we can achieve the optimal transformation parameters. In the following sections, several steps involved are separately described in detail.

A. SIFT Matching

The preregistration process of the proposed methodology starts with the SIFT matching, which contains five major steps: scale-space extrema detection, keypoint localization, orientation assignment, keypoint descriptor, and keypoint matching [9].

1) Scale-space extrema detection: Given an input image $I(x, y)$, the scale space of I, $L(x, y, \sigma)$ is defined as

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{4}$$

where $*$ is the convolution operation in x and y and $G(x, y, \sigma)$ is a variable-scale Gaussian defined as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}. \tag{5}$$

The detection of scale-space extrema is implemented by detecting the local maxima and minima of $D(x, y, \sigma)$, which is defined as the convolution of the difference-of-Gaussian function with the image $I(x, y)$

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{6}$$

where k is a constant multiplicative factor by which two nearby scales are separated. The detection is performed over all scales and locations by comparing each sample point to its eight neighbors in the current scale and nine neighbors in the scale above and below. It is selected as a keypoint candidate if it is larger or smaller than all of these neighbors.

2) Keypoint localization: This step performs a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. The keypoint candidates with a value of less than 0.03 (suggested by Lowe [9] based on experiments) are considered as having low contrast and are discarded due to their unstableness. Similarly, the keypoint candidates with a ratio between the principal curvatures greater than 10 (suggested by Lowe [9] based on experiments) are considered to be poorly localized along an edge and are also discarded because their descriptors are not distinctive.
3) Orientation assignment: Each retained keypoint is assigned a consistent orientation based on local image gradient directions. It performs based on an orientation histogram, which is formed from the gradient orientation of sample points within a region around the keypoint. The orientation histogram has 36 bins covering the 360° range of orientations. Peaks in the orientation histogram denote dominant direction of local gradients and are assigned to the corresponding keypoints. By this step, the descriptor of each keypoint can be represented relative to the orientation and hence can achieve invariance to image rotation.
4) Keypoint descriptor: The image gradient magnitudes and orientations are computed around the keypoint location with a 16 × 16 sample array and are weighted by a Gaussian window. A 4 × 4 array of orientation histograms with eight orientation bins in each array is then created by summarizing the contents over 4 × 4 sample subregions. The descriptor of each keypoint is formed by containing the values of all the orientation histogram entries, achieving a 128-element feature vector.
5) SIFT matching: As mentioned previously, the SIFT matching is performed through the nearest neighbor

approach as proposed in [9]. The nearest neighbor is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector. An effective measure for matching is the ratio d_{ratio}, which denotes the ratio between the distance of the nearest neighbor and that of the second nearest neighbor as defined before. The nearest neighbor matching is efficiently implemented using the best-bin-first algorithm. More details about SIFT can be found in [9].

B. Modified Outlier Removal

Even after the removal of false initial matches through a ratio threshold, some incorrect matches may still exist. Therefore, a reliable outlier removal process is necessary. A consistency check method proposed in [33] was used to exclude false contour matches, and we now modify it to remove false SIFT matches. Given two sets of m matched keypoints $\{R_i\}$ and $\{S_i\}$ belonging to the reference and sensed images, respectively, let $\overline{R_iR_j}$ be the distance between keypoints R_i and R_j, and $\overline{S_iS_j}$ be the distance between keypoints S_i and S_j. A distance ratio D_{ij} is defined as

$$D_{ij} = \frac{\overline{R_iR_j}}{\overline{S_iS_j}}. \tag{7}$$

Ratios of D_{ij} are computed based on all $m(m-1)/2$ possible combinations, and the number of D_{ij} in the intervals is counted in a statistical way, forming a scale histogram. The denser cluster in the scale histogram corresponds to the true scale difference between the images. The keypoint pairs that contribute to the cluster are accepted as correct matches, while the ones that are scattered and away from the cluster are considered as incorrect matches and are then eliminated. The outlier removal process is performed in an iterative fashion: discard the most likely mismatches first, and then compute the rmse based on the remaining matches; the iteration stops when the rmse is below a certain threshold (one) or the maximum number of 20 iterations is achieved, which can guarantee that the procedure obtains a set of available matches. Note that, in order to solve the transformation parameters, the iteration stops in advance if only three matches remain when there are too few initial matching candidates.

A set of reliable SIFT matches can be obtained after the outlier removal procedure. Then, the least square method is used to solve the coarse transformation parameters based on these matches. The coarse results provide an excellent initial solution for the subsequent fine-tuning process.

C. MI

The concept of MI represents a measure of statistical dependence between two data sets. Let V^C be the continuous domain on which $f_R(x,y)$ and $f_S(x,y)$ are defined. The coordinates (x_i, y_i) are samples of V^C, and the discrete set of these samples is defined as V. L_R and L_S are defined as the discrete sets of intensities of the two images. MI is defined as

$$S(\boldsymbol{\mu}) = \sum_{r\in L_R}\sum_{s\in L_S} p(r,s;\boldsymbol{\mu}) \cdot \log_2\left(\frac{p(r,s;\boldsymbol{\mu})}{p_R(r;\boldsymbol{\mu})p_S(s;\boldsymbol{\mu})}\right) \tag{8}$$

where $p_R(r;\boldsymbol{\mu})$ and $p_S(r;\boldsymbol{\mu})$ are the marginal probability distributions, $p(r,s;\boldsymbol{\mu})$ is the joint probability distribution, $r \in L_R$, and $s \in L_S$. These distributions can be obtained by

$$p(r,s;\boldsymbol{\mu}) = \frac{h(r,s;\boldsymbol{\mu})}{\sum_{r\in L_R}\sum_{s\in L_S} h(r,s;\boldsymbol{\mu})} \tag{9}$$

$$p(r;\boldsymbol{\mu}) = \sum_{s\in L_R} p(r,s;\boldsymbol{\mu}) \tag{10}$$

$$p(s;\boldsymbol{\mu}) = \sum_{r\in L_s} p(r,s;\boldsymbol{\mu}) \tag{11}$$

where h is the joint discrete Parzen histogram given by

$$h(r,s;\boldsymbol{\mu}) = \frac{1}{\varepsilon_R\varepsilon_S}\sum_{(x_i,y_i)\in V} w\left(r/\varepsilon_R - f_R(x_i,y_i)/\varepsilon_R\right) \cdot w\left(s/\varepsilon_S - f_S\left(T_{\boldsymbol{\mu}}(x_i,y_i)\right)/\varepsilon_S\right) \tag{12}$$

where $T_{\boldsymbol{\mu}}(\bullet)$ is the geometric transformation with associated parameters $\boldsymbol{\mu}$ and where ε_R and ε_S are two positive scaling factors related to card(L_R) and card(L_S), respectively, in order to control the width of the Parzen window w. In this paper, B-spline functions $\beta^n(x)$ are used as the Parzen window due to their attractive properties [20]. They are piecewise polynomials of degree $n \geq 0$ and can be recursively defined as the convolution of the B-spline of degree $n-1$ with β^0 [20]

$$\begin{aligned}\beta^n(x) &= (\beta^{n-1} * \beta^0)(x) \\ &= \int_{-\infty}^{\infty} \beta^{n-1}(x) * \beta^0(x-t)dt, \quad n > 0\end{aligned} \tag{13}$$

where β^0 is a centered unit square pulse. The cubic B-spline function β^3 is selected as model kernel since it could provide a high-quality model [20]. The Parzen histogram is built based on the partial overlapping domain of $f_R(x,y)$ and $f_S(T_{\boldsymbol{\mu}}(x,y))$, namely, the gray values (r,s) of those pairs of pixels which lie in the same position.

From the definition of MI, it is shown that the geometric correction parameter $\boldsymbol{\mu}$ is the optimal solution when the MI value is maximal. In this context, the problem of image registration is mapped as an optimization problem, which can be expressed as

$$\boldsymbol{\mu}^* = \arg\left(opt\left(S(\boldsymbol{\mu})\right)\right) \tag{14}$$

where S is the MI defined previously and $\boldsymbol{\mu}^*$ is a set of the optimal transformation parameters corresponding to the maximum of MI. With the parameter $\boldsymbol{\mu}^*$, the transformed sensed image $f_S(T_{\boldsymbol{\mu}^*}(x,y))$ is correctly aligned with the reference image $f_R(x,y)$.

D. Multiresolution Framework

To obtain the smoothness of the MI surface, the estimation of joint histogram to compute MI is performed using the Parzen window estimator [34], [35] based on cubic B-splines [36]. The main disadvantage of this high-order interpolation model is its high computation cost. A multiresolution framework is a desired solution to this problem.

The multiresolution framework works iteratively from the coarsest level of the image pyramid to the finest level of the

image pyramid. For all cases, the MI between the whole overlap of subband images of the reference and sensed images is computed at each level and maximized successively, and the search is performed on an interval around the optimal transformation parameters found at the previous level and is refined at the next level. Thus, the accuracy of the search increases from coarse resolution to fine resolution. At each level, the iteration stops when the improvement of the MI value is below a certain threshold (3.5×10^{-4}) or the maximum number of 200 iterations is achieved, which are sufficient for the search to converge to a set of registration parameters with a subpixel accuracy based on our experiments. Here, the number of bins of the joint histograms varies with the levels of the image pyramid. The number of bins at the finest level is set to $\sqrt{2 \times N_x \times N_y}/N$, where N_x and N_y denote the number of rows and columns of the images, respectively, and N is between 5 and 50. Moreover, the number of bins decreases geometrically by a factor of 2 from finer levels to coarser levels, e.g., the number of bins at the second finest level is one half of that at the finest level. The histograms produce a significantly smoother MI surface than the $N_x \cdot N_y$ histograms. The smoother MI surface works better with the optimizer involved, and the decreased number of bins dramatically reduces the computation cost of the search.

E. Modified Marquardt–Levenberg Search Strategy

For the multiresolution framework, a suitable optimizer that converges in a few criterion evaluations when initialized with good starting conditions is desired. A specifically designed optimizer proposed in [20] exhibits superlinear convergence when close enough to the optimum. It is a modification of the traditional Marquardt–Levenberg and can be described by

$$\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} - \left(\mathcal{H}S\left(\boldsymbol{\mu}^{(k)}\right)\right)^{-1} \cdot \nabla S\left(\boldsymbol{\mu}^{(k)}\right) \tag{15}$$

where ∇S and $\mathcal{H}S$ are the gradient and the modified Hessian of MI, respectively. The gradient ∇S is defined as

$$\nabla S = \left[\frac{\partial S}{\partial \boldsymbol{\mu}_1}, \frac{\partial S}{\partial \boldsymbol{\mu}_2}, \cdots\right]. \tag{16}$$

A component of ∇S is given by

$$\frac{\partial S}{\partial \boldsymbol{\mu}} = -\sum_{r \in L_R} \sum_{s \in L_S} \frac{\partial p(r,s;\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} \log_2\left(\frac{p(r,s;\boldsymbol{\mu})}{p_S(s;\boldsymbol{\mu})}\right). \tag{17}$$

The modified Hessian $\mathcal{H}S$ is defined as

$$[\mathcal{H}S(\boldsymbol{\mu})]_{i,j} = \left[\nabla^2 S(\boldsymbol{\mu})\right]_{i,j} \cdot (1 + \delta_{i,j}\lambda) \tag{18}$$

where $\delta_{i,j}$ is the Kronecker symbol, λ is a tuning factor, and $\nabla^2 S$ is the Hessian of S. $\nabla^2 S$ is defined as the matrix of the second derivative of S and can be described by

$$\nabla^2 S = \begin{bmatrix} \frac{\partial^2 S}{\partial \mu_1 \partial \mu_1} & \frac{\partial^2 S}{\partial \mu_1 \partial \mu_2} & \cdots \\ \frac{\partial^2 S}{\partial \mu_2 \partial \mu_1} & \frac{\partial^2 S}{\partial \mu_2 \partial \mu_2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{19}$$

A component of $\nabla^2 S$ is given by

$$\frac{\partial^2 S}{\partial \mu_1 \partial \mu_2} \approx \frac{1}{\log_e(2)} \left(\sum_{s \in L_S} \frac{\partial p_S(s)}{\partial \mu_1} \frac{\partial p_S(s)}{\partial \mu_2} \frac{1}{p_S(s)}\right) - \frac{1}{\log_e(2)} \left(\sum_{r \in L_R} \sum_{s \in L_S} \frac{\partial p(r,s)}{\partial \mu_1} \frac{\partial p(r,s)}{\partial \mu_2} \frac{1}{p(r,s)}\right). \tag{20}$$

IV. Experimental Study

In this section, in order to test the performance of the proposed algorithm, we applied it to various optical and SAR remote sensing images taken at different situations (multispectral, multisensor, and multitemporal), namely, optical-to-optical, optical-to-SAR, and SAR-to-SAR image registrations. Furthermore, results are also provided to compare the proposed method with other representative registration methods in terms of registration accuracy and computation complexity. These experiments were conducted on a computer with an Intel Core 2.33-GHz processor and 2.0 GB of physical memory.

The accuracy of the registration methods are evaluated in three ways in this work. The first is based on comparison of registration parameters with the ground truth (when available) or the reliable reference value [33] obtained by the manual registration via ENVI [37]. Two commonly used evaluation measures, namely, MI [17] and rmse [12], [30], are used as the second way. The rmse measure is normalized to the pixel size [30]. The third way is visual inspection by generating the checkerboard mosaiced images using the corresponding registration parameters [14], [17], [21].

The manual registration is performed as follows: first, we manually identify 100 initial matching points which evenly locate in the whole image so that an ideal distribution could be achieved ($S_{\text{cat}} = 1$, where S_{cat} is a measure quantifying the CP distribution [30]). From the initial matching points, 30 matching points leading to the lowest residuals are selected. Then, the retained matching points are refined to reduce the residual as low as possible. Finally, the registration parameters are solved by the least square method.

A. Multispectral Imagery

The first pair of images was selected from a set of 12-band aerial images acquired by a Thematic Mapper (TM) simulator on the ER-2 aircraft over a mountainous area with a spatial resolution of 25 m [33]. Both images are shown in Fig. 4. A segment with a size of 512 × 512 pixels from band 1 (0.42–0.45 μm) was selected as the reference image. A segment with the same size selected from band 9 (1.55–1.75 μm) after a simulated rotation (15°) and a shift (−7250) was selected as the sensed image. The information on the image band and transformation parameters is provided in the reference [33]. In this way, the ground truth of the transformation parameters between the two images is known. The ground truth corresponding to the affine transform model (a_{11}, a_{12}, a_{21}, a_{22}, δ_x, and δ_y) is (0. 9659, −0.2588, 0.2588, 0.9659, −7.000, and 250.000).

The results for this pair of images are presented in the first part of Table II. As presented in Table II, the rotation and scaling parameters (a_{11}, a_{12}, a_{21}, and a_{22}) obtained by the proposed method have hardly any error compared with the

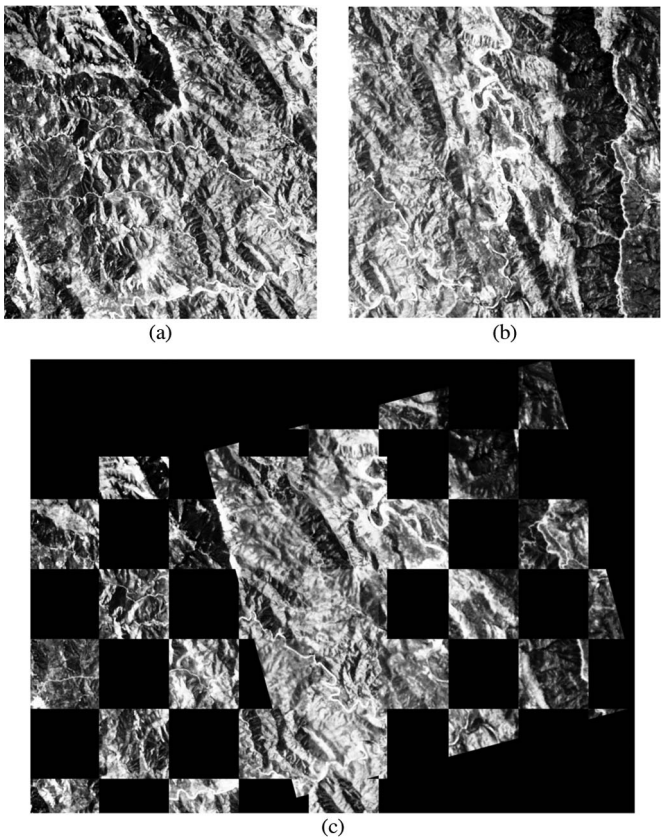


Fig. 4. Aerial images of a mountainous area. (a) Reference image extracted from band 0 of a TM simulator. (b) Sensed image extracted from band 8 of a TM simulator. (c) Checkerboard mosaiced image of (a) and (b).

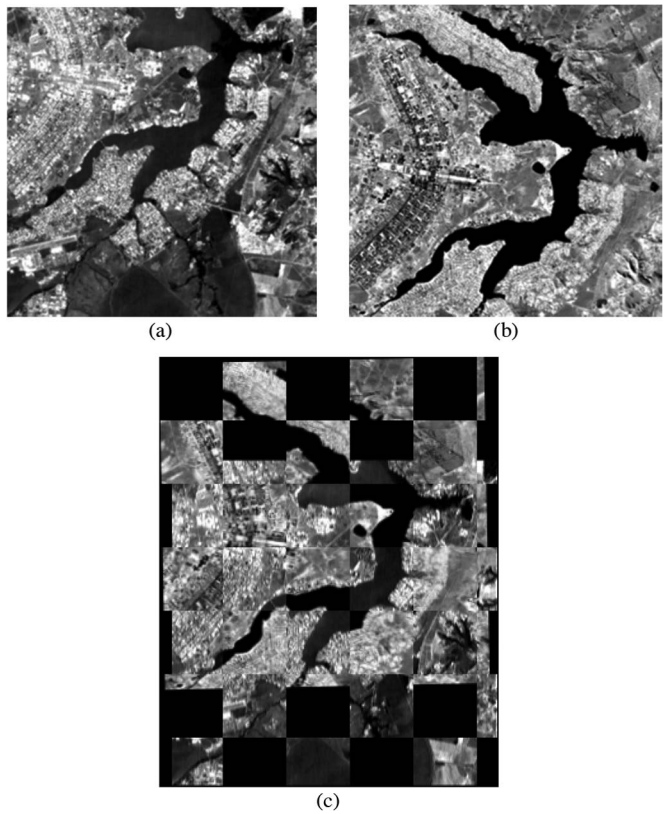


Fig. 5. Images of an urban area in Brasilia, Brazil. (a) Reference image taken by SPOT in 1995. (b) Sensed image taken by Landsat TM in 1994. (c) Checkerboard mosaiced image of (a) and (b).

ground truth, and the shift parameters (δ_x and δ_y) achieve an accuracy of one hundredth of a pixel. These can also be indicated by the values of MI and rmse. Furthermore, the checkerboard mosaiced image [Fig. 4(c)] shows that the image features of the two images such as crest line are precisely overlapped, which demonstrates the accuracy of our obtained registration parameters.

B. SPOT/Landsat

The second pair of images consists of a segment with a size of 256 $\times$ 256 pixels from band 3 (0.78–0.89 μm) of a scene taken by SPOT on August 08, 1995, over an urban area in Brasilia, Brazil (with an initial spatial resolution of 20 m resampled to 30 m), and of a segment with a size of 256 $\times$ 256 pixels from band 4 (0.76–0.90 μm) of a scene taken by the sensor Landsat Thematic Mapper (TM) on June 07, 1994, over the same area (with a spatial resolution of 30 m). Both images [38] are shown in Fig. 5, and the images of Fig. 5(a) and (b) are selected as the reference and sensed images, respectively.

The results obtained for this pair of images are presented in the second part of Table II. For geometric correction, the transformation parameters achieved by the proposed algorithm are more accurate than those achieved by the manual registration. As expected, the registration parameters led to a maximum of MI and a subpixel accuracy regarding rmse. Moreover, from the checkerboard mosaiced image shown in Fig. 5(c), it can be seen that image features of the two images such as edge and region are precisely overlapped.

C. ALOS-PALSAR/Landsat ETM+

The third pair of images consists of a segment with a size of 800 $\times$ 800 pixels from the HH mode (L-band) of a scene taken by the sensor ALOS-PALSAR on June 05, 2010, at the region of Campbell River in British Columbia (with an initial spatial resolution of 15 m resampled to 30 m) and of a segment with a size of 800 $\times$ 800 pixels from band 5 (1.55–1.75 μm) of a scene taken by the sensor Landsat Enhanced Thematic Mapper Plus (ETM+) on June 26, 1999, at the same region (with a spatial resolution of 30 m). Both images are shown in Fig. 6. As is well known, images acquired by SAR sensors (active sensors) with speckle noise have substantial characteristic differences from those acquired by optical sensors (passive sensors), which increases challenges to registration. The Landsat images are courtesy of the U.S. Geological Survey.

The results obtained for this pair of images are presented in the third part of Table II. The proposed method provides more precise registration parameters than the manual registration method. Accordingly, an accuracy higher than 0.30 pixels in terms of rmse has been achieved, which can also be demonstrated by the largest value of MI (0.936). Moreover, visual inspection of the registered image [Fig. 6(c)] validates that the transformed sensed image fits well with the reference image across the whole image overlap.

D. Radarsat-2/Radarsat-2

In order to validate the robustness of the proposed method, a pair of multitemporal SAR images was considered in this

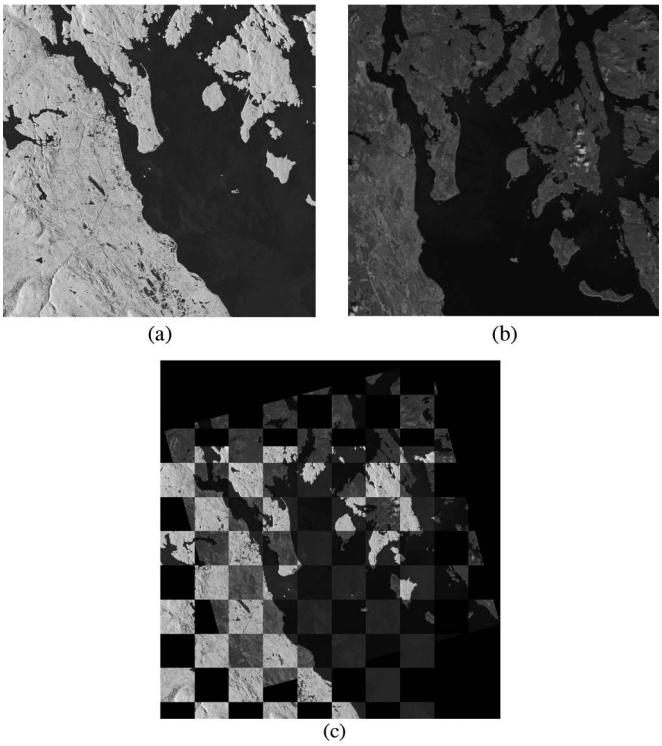


Fig. 6. Images of Campbell River in British Columbia. (a) Reference image taken by ALOS-PALSAR in 2010. (b) Sensed image taken by Landsat ETM+ in 1996. (c) Checkerboard mosaiced image of (a) and (b).

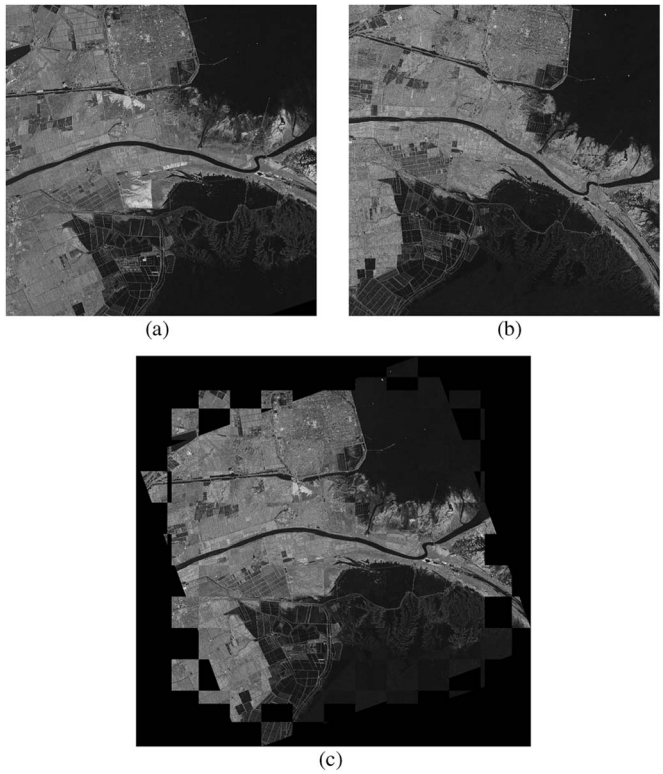


Fig. 7. Images of Yellow River Estuary. (a) Reference image taken by Radarsat-2 in 2008. (b) Sensed image taken by Radarsat-2 in 2009. (c) Checkerboard mosaiced image of (a) and (b).

example, a more complicated situation due to the influence of speckle noise. This pair of SAR images with a size of 5000 $\times$ 5000 pixels was taken by Radarsat-2 (C-band) at the region of Yellow River Estuary in China in June 2008 and June 2009, respectively (Fig. 7). The spatial resolution of this pair of image is 8 m. Note that the two images are a single-look SAR image (the SAR image that is only produced based on one synthetic aperture range) [39] and a four-look SAR image (the SAR image that is produced by averaging on four single-look SAR images) [39], respectively, which means that the influence of speckle noise on the image acquired in 2009 is much greater than that of the one acquired in 2008. Such huge disparity of speckle noise level between the two images may aggravate the difficulties met in the process of registration. There exist many line features in the two images which allow a better visualization of the registration accuracy.

The registration results for this pair of images are presented in the last part of Table II. From the results, one can find that the difference of the parameters (a_{11}, a_{12}, a_{21}, and a_{22}) between our obtained value and the reference value is not higher than 0.004, and the difference of the parameters (δ_x and δ_y) is below a pixel. One can also find that our obtained registration parameters result in the largest value of MI and the lowest value of rmse, which confirm the high registration accuracy. Furthermore, in the mosaic shown in Fig. 7(c), the lines extracted from the reference image and the transformed sensed image are precisely overlapped.

E. Comparison With Other Registration Methods

The comparative results of the applications of these registration methods to the four previously described pairs of images are presented in Table II. It can be observed that the proposed method generally outperforms other registration methods. There are mainly three reasons behind this: 1) the SIFT approach equipped with a reliable outlier removal procedure can guarantee the preregistration results close to the solution as validated by the results presented in Table II; 2) the Parzen window estimator used to compute MI is more accurate than generalized partial volume estimation so that the search is not influenced by spurious optima [35]; and 3) the modified Marquardt–Levenberg search strategy combines the efficiency of the Newton method with the robustness of the gradient method. Regarding other methods, IS-SIFT [12] obtains more correct matches than standard SIFT and improves the quality of these matches. The IS-SIFT method achieves acceptable results in some cases. However, the number of final matches is still inadequate to get precise parameters for very difficult cases such as the SAR-to-SAR registration since the number of initial matches is low for these cases. The registration results achieved by SPSA [17] are also not precise enough. Moreover, the stochastic nature of SPSA is an undesirable property, and it converges slowly when close to the solution [17].

Regarding computational efficiency, the associated processing times of each method are presented in the time column in Table II. For SIFT-based matching, the computational cost is relative to the number of detected keypoints. For example, the number of keypoints detected from the last image pair is much larger than that of those from the third image pair, and hence, the preregistration process for the last image pair requires much higher computational time. IS-SIFT [12] takes additional computational costs in the image segmentation process. For the MI-based methods, the computational cost is proportional to the

number of MI evaluations. SPSA is computationally expensive because it converges slowly when close to the solution as described in [17]. On the contrary, the Marquardt–Levenberg search strategy exhibits superlinear convergence [20] so that a few MI evaluations are required.

V. Concluding Remarks

In this paper, we have proposed a novel coarse-to-fine strategy for automatic image registration based on SIFT and MI. This strategy consists of a preregistration process and a fine-tuning process. To start with, the preregistration process is implemented by the SIFT approach equipped with a reliable outlier removal procedure. SIFT results in robust matching by means of its distinctiveness and invariance. Our modified outlier removal method can generally eliminate most incorrect SIFT matches and can retain most correct ones by means of its robustness. The coarse results obtained by the preregistration process provide a near-optimal initial solution for the optimizer in the fine-tuning process, and the preregistration process achieves near real-time performance. Next, the fine-tuning process is implemented by maximization of MI using the modified Marquardt–Levenberg search strategy in a multiresolution framework. The MI similarity measure is particularly suitable for multimodal image registration. A multiresolution strategy can increase the robustness of the algorithm and can significantly improve its computational efficiency. The selected optimizer has the advantages of efficiency and robustness, which exhibits superlinear convergence when close to the solution. In our experiments, we have tested the proposed method on the optical and SAR remote sensing images in the following situations: multispectral, multisensor, and multitemporal. The experimental results show that the proposed method obviously improves the accuracy of registration and achieves good computational efficiency. Regarding the registration of multiview images with the difference in the acquisition angle and the terrain elevation, the affine transformation model applied in this work is not suitable. Other more appropriate transformation models such as thin-plate spline [1] can substitute the affine model in the proposed coarse-to-fine scheme to handle the effects introduced by the acquisition angle and terrain elevation differences [40], which is one of our future works.

Acknowledgment

The authors would like to thank the editors and anonymous reviewers for their valuable comments and helpful suggestions which greatly improved this paper's quality.

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] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Trans. Geosci. Remote Sen.*, vol. 36, no. 5, pp. 1566–1577, Sep. 1998.

[3] H. Gonçalves, J. A. Gonçalves, and L. Corte-Real, "HAIRIS: A method for automatic image registration through histogram-based image segmentation," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 776–789, Mar. 2011.

[4] C. Xing and P. Qiu, "Intensity-based image registration by nonparametric local smoothing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 2081–2092, Oct. 2011.

[5] M. Debella-Gilo and A. Kääb, "Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation," *Remote Sens. Environ.*, vol. 115, no. 1, pp. 130–142, Jan. 2011.

[6] J. Liang, Z. Liao, S. Yang, and Y. Wang, "Image matching based on orientation-magnitude histograms and global consistency," *Pattern Recognit.*, vol. 45, no. 10, pp. 3825–3833, Oct. 2012.

[7] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

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

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

[10] 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.

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

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

[13] 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.

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

[15] L. Cheng, J. Gong, X. Yang, C. Fan, and P. Han, "Robust affine invariant feature extraction for image matching," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 246–250, Apr. 2008.

[16] M. Schneider, S. Suri, M. Lehner, and P. Reinartz, "Matching of high-resolution optical data to a shaded DEM," *Int. J. Image Data Fusion*, vol. 3, no. 2, pp. 111–127, Jun. 2012.

[17] 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.

[18] 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.

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

[20] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Process.*, vol. 9, no. 12, pp. 2083–2099, Dec. 2000.

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

[22] H. Ghorbani and A. A. Beheshti, "Multiresolution registration of multi-temporal remote sensing images by optimization of mutual information using a simulated annealing based Marquardt–Levenberg Technique," in *Proc. Int. Conf. Intell. Adv. Syst.*, 2007, pp. 685–690.

[23] D. Falco, A. D. Cioppa, D. Maisto, and E. Tarantino, "Differential evolution as a viable tool for satellite image registration," *Appl. Soft Comput.*, vol. 8, no. 4, pp. 1453–1462, Sep. 2008.

[24] S. Suri, P. Schwind, P. Reinartz, and J. Uhl, "Combining mutual information and scale invariant feature transform for fast and robust multisensor SAR image registration," in *Proc. 75th ASRPS Conf.*, Baltimore, MD, USA, Mar. 2009, pp. 1–12.

[25] Y. S. Heo, K. M. Lee, and S. U. Lee, "Mutual information-based stereo matching combined with SIFT descriptor in log-chromaticity color space," in *Proc. CVPR*, 2009, pp. 445–452.

[26] Y. S. Heo, K. M. Lee, and S. U. Lee, "Joint depth map and color consistency estimation for stereo images with different illuminations and cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1094–1106, May 2013.

[27] N. E. Mekky, F. E.-Z. Abou-Chadi, and S. Kishk, "A new dental panoramic X-ray image registration technique using hybrid and hierarchical strategies," in *Proc. Int. Conf. Comput. Eng. Syst.*, 2010, pp. 361–367.

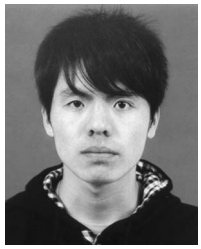
[28] D. Ballard, "Generalized Hough transform to detect arbitrary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-13, no. 2, pp. 111–122, Apr. 1981.

[29] 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.
[30] H. Gonçalves, J. A. Gonçalves, and L. Corte-Real, "Measures for an objective evaluation of the geometric correction process quality," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 292–296, Apr. 2009.
[31] P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 27–41, Jan. 1988.
[32] I. Zavorin and J. L. Moigne, "Use of multiresolution wavelet feature pyramids for automatic registration of multisensory imagery," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 770–782, Jun. 2005.
[33] H. Li, B. S. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 320–334, Mar. 1995.
[34] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.
[35] D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens, "Comparison between Parzen window interpolation and generalized partial volume estimation for nongrid image registration using mutual information," in *Proc. WBIR*, 2006, pp. 206–213.
[36] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part 1—Theory," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 821–833, Feb. 1993.
[37] [Online]. Available: http://www.exelisvis.com/ProductsServices/ENVI.aspx
[38] D. Fedorov, L. M. G. Fonseca, C. Kenney, and B. S. Manjunath, "Automatic registration and mosaicking system for remotely sensed imagery," in *Proc. Image Signal Process. Remote Sens. VIII*, 2003, vol. 4885, pp. 444–451.
[39] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Boston, MA, USA: Artech House, 1998.
[40] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2127–2137, Sep. 2005.

Maoguo Gong (M'07) received the B.S. degree in electronic engineering and the Ph.D. degree in electronic science and technology from Xidian University, Xi'an, China, in 2003 and 2009, respectively.

Since 2006, he has been a Teacher with Xidian University, where he was promoted as an Associate Professor and as a Full Professor in 2008 and 2010, respectively, both with exceptive admission. He is currently a Full Professor with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, Xidian University. In 2012, he was appointed as the Chief Professor (Leader) of the Intelligent Computing Innovative Research Team of Shaanxi Province, China. He is leading or has completed ten projects as the PI, funded by the National Natural Science Foundation of China, the National High Technology Research and Development Program (863 Program) of China, and others. He has published over 50 papers in journals and conferences. He is the holder of eight granted patents as the first inventor. His research interests include computational intelligence with applications.

Dr. Gong is an Executive Committee Member of the Natural Computation Society of the Chinese Association for Artificial Intelligence and a Senior Member of the Chinese Computer Federation. He was the recipient of the prestigious National Top Young Talent of China (selected by the Central Organization Department of China), the New Century Excellent Talent in University (selected by the Ministry of Education of China), the Young Scientist Award of Shaanxi Province, the New Scientific and Technological Star of Shaanxi Province, the Elsevier SCOPUS Young Researcher Award of China, and two Science & Technology Awards of Shaanxi Province (First Level in 2008 and 2010).



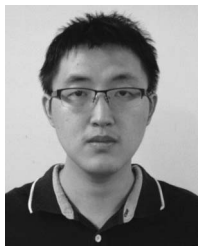
Shengmeng Zhao received the B.S. degree in biomedical engineering from Xidian University, Xi'an, China, in 2011, where he is currently working toward the M.S. degree.

His research interests include image registration and machine learning.

Licheng Jiao (SM'89) received the B.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1982 and the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

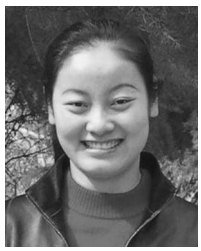
Since 1992, he has been a Professor with the School of Electronic Engineering, Xidian University, Xi'an, where he is currently the Director of the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China. He is in charge of about 40 important scientific research projects and has published more than 20 monographs and a hundred papers in international journals and conferences. His research interests include image processing, natural computation, machine learning, and intelligent information processing.

Dr. Jiao is a member of the IEEE Xi'an Section Execution Committee, the Chairman of the Awards and Recognition Committee, the Vice Board Chairperson of the Chinese Association of Artificial Intelligence, a Councilor of the Chinese Institute of Electronics, a committee member of the Chinese Committee of Neural Networks, and an expert of the Academic Degrees Committee of the State Council.



Dayong Tian received the B.S. degree in electronic information science and technology from Xidian University, Xi'an, China, in 2011, where he is currently working toward the M.S. degree.

His research interests include image segmentation and shape registration.



Shuang Wang (M'07) received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 2000 and 2003, respectively.

She is currently a Professor with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, Xidian University. Her research interests include synthetic aperture radar image processing and pattern recognition.