Title: An Approach to Fine Coregistration Between Very High Resolution Multispectral Images Based on Registration Noise Distribution

This paper appears in: IEEE Transactions on Geoscience and Remote Sensing

Date of Publication: December, 2015

Author(s): Youkyung Han, Francesca Bovolo, Lorenzo Bruzzone

Volume: 53, Issue: 12

Page(s): 6650 - 6662

DOI: 10.1109/TGRS.2015.2445632

An Approach to Fine Coregistration Between Very High Resolution Multispectral Images Based on Registration Noise Distribution

Youkyung Han, *Member, IEEE*, Francesca Bovolo, *Senior Member, IEEE*, and Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*—Even after applying effective coregistration methods, multitemporal images are likely to show a residual misalignment, which is referred to as registration noise (RN). This is because coregistration methods from the literature cannot fully handle the local dissimilarities induced by differences in the acquisition conditions (e.g., the stability of the acquisition platform, the off-nadir angle of the sensor, the structure of the considered scene, etc.). This paper addresses the problem of reducing such a residual misalignment by proposing a fine automatic coregistration approach for very high resolution (VHR) multispectral images. The proposed method takes advantage of the properties of the residual misalignment itself. To this end, RN is first extracted in the change vector analysis (CVA) polar domain according to the behaviors of the specific multitemporal images considered. Then, a local analysis of RN pixels (i.e., those showing residual misalignment) is conducted for automatically extracting control points (CPs) and matching them according to their estimated displacement. Matched CPs are used for generating a deformation map by interpolation. Finally, one VHR image is warped to the coordinates of the other through a deformation map. Experiments carried out on simulated and real multitemporal VHR images confirm the effectiveness of the proposed approach.**

***Index Terms*—Change vector analysis (CVA), image coregistration, registration noise (RN), remote sensing, very high resolution (VHR) multispectral images.**

I. Introduction

With the launch of Earth observation satellites equipped with very high resolution (VHR) sensors (e.g., IKONOS, QuickBird, GeoEye-1, and WorldView-2), the availability of images with a ground field of view of less than 1 m has increased. As satellites can periodically acquire images of the Earth's surface, multitemporal images can play an important role in a wide range of applications. However, to utilize the multitemporal images, they should be precisely aligned to each other. Image coregistration, which is the process of spatially overlaying two or more images of the same scene [1], is thus a fundamental preprocessing requirement. Highly precise coregistration allows for better information extraction from multitemporal images.

Most of the coregistration procedures between multitemporal images consist of four steps. First, control points (CPs), which are the objects that correspond to distinctive and representative points of the investigated scene, are extracted from each image independently. Second, each CP from one image is matched with the corresponding CP of the other image using a similarity measure. Third, matched CPs are employed to estimate a transformation model. Fourth, one image (slave image) is warped to the other image (master image). In order to carry out the coregistration task, the most critical part is to extract CPs and match them. In the literature, two kinds of approaches can be found: area-based methods and feature-based methods [1]. Area-based methods use windows of predefined size or even the entire images for the correspondence extraction. They work well with images that have few salient objects because they do not perform extraction of CPs. The limitation of these methods is that they cannot work when the multitemporal images show significant distortions or large geometric differences. They are also sensitive to the intensity changes in images. Feature-based methods extract CPs based on the representative points, e.g., line intersections, starting and ending points of lines, or centroid pixels of close-boundary regions. The extracted CPs between images are matched by various feature descriptors or similarity measures along with spatial relationships among the CPs. The feature-based methods can be applied, although the images have significant distortions and geometric differences. However, they are less effective when the images have few salient features and tend to extract a large number of false-matched pairs of CPs (i.e., outliers) if there are many similar features within a given scene.

Both the area-based and feature-based methods have been variously and effectively employed in CP matching of low and medium spatial resolution images [2]–[5]. However, they resulted to be less effective when dealing with new-generation satellite sensors having a very high geometrical resolution, which may observe the same scene from different paths and angles, thus resulting in images with nonrigid geometric differences. Accordingly, recorded scenes may show more complex geometric relief displacements and large distortions with respect to the low- and medium-resolution acquisitions [6]. Moreover, complex height variations in the terrain cause severe relief

Manuscript received January 16, 2015; revised May 15, 2015; accepted June 5, 2015. This work was supported in part by the Italian Ministry of Education, University and Research through Programmi di Ricerca di Rilevante Interesse Nazionale—PRIN 2012 under the Project "Very high spatial and spectral resolution remote sensing: a novel integrated data analysis system."

Y. Han and F. Bovolo are with the Fondazione Bruno Kessler, Center for Information and Communication Technology, 38122 Trento, Italy (e-mail: bovolo@fbk.eu).

L. Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy.

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.2015.2445632

displacements and local distortions in the image [7]. Thus, the feature-based methods (which can correct local distortions) are more effective when dealing with VHR multitemporal images rather than the area-based methods [8]–[12]. In order to account for VHR image properties, methods that extract CPs using scale-invariant feature transform (SIFT) have been developed, which represent the feature-based matching method [13]. The SIFT technique has been adapted and improved to make it more appropriate for VHR images in order to increase the registration accuracy [14]–[17]. It is also applied to both the coregistration of large images [8], [18] and of multisensor images such as SAR and optical images [19]–[21]. However, these feature-based approaches have a limitation because CPs are matched based on the similarity between the properties of the local regions. It means that only regions showing similar geometric properties can be considered candidates for the CP matching. However, when the CPs are extracted from a region having local distortion, it may become difficult to find their exact corresponding location due to the distortion itself. Therefore, the local regions that are distorted or show dissimilar properties between images cannot be completely corrected by these approaches. Some studies introduced nonrigid transformation models to mitigate the local distortions, under the assumption that CPs are evenly distributed over the images, [22]–[25]. However, having CPs evenly distributed over the images is difficult; thus, these models cannot completely solve the problem.

Despite the aforementioned methods may achieve high overall coregistration accuracy, there is still a margin to improve it. The margin relies in the possible significant residual misregistration occurring at local level even after coregistration, which is also referred to as registration noise (RN) [26], [27]. The RN occurs due to several factors causing distortions in VHR images (e.g., the stability of the acquisition platform, the off-nadir angle of the sensor, the structure of the considered scene, etc.). The most critical component of this noise is related to those pixels that spatially correspond after coregistration but belong to different objects at the two dates (i.e., the border region of objects or high-frequency area in the images). Those pixels affect in a negative way the multitemporal information extraction processes. They reduce the effectiveness of change detection between multitemporal images, leading to false alarms [28]. They impact on the accuracy of data fusion and multitemporal segmentation [29]. RN properties are analyzed in [30] in the feature space of change vector analysis (CVA) in the polar domain. RN information has been employed together with a segmentation technique to improve change detection in VHR images in [31].

In this paper, we propose to exploit RN properties for reducing the residual local misalignment affecting multitemporal images after applying state-of-the-art coregistration methods. By reducing local misalignment, the method aims at increasing the accuracy of coregistration and thus of multitemporal information extraction methods. To this end, residual misregistration information is directly employed in the design of an approach for fine coregistration of VHR multitemporal images. After standard coregistration, RN extraction is conducted in the CVA polar domain [30] and pixels that are locally dominated by RN (i.e., those that still have to be aligned) are used as CPs. The matching is conducted at a local level after estimating the amount of displacement. A deformation map is generated by interpolation using matched CP pairs. Finally, the slave image is warped to the coordinates of the master image throughout a deformation map. The experiments carried out on simulated and real multitemporal VHR images confirm the effectiveness of the proposed approach.

The remainder of this paper is structured as follows. We recall the concepts and properties of RN derived in the CVA in the polar domain in Section II. Section III illustrates the proposed automatic fine coregistration technique based on the local analysis of RN pixels. Section IV describes simulated and real data sets made up of VHR optical images and presents the design of experiments. Section V shows the experimental results obtained on the data sets. Finally, conclusion and future work are given in Section VI.

II. Background on RN Estimation

In [30], RN pixels (i.e., pixels affected by residual misalignment) have been identified through CVA in the polar domain. In this domain, RN pixels can be identified by solving two independent 1-D problems (i.e., one along the magnitude variable and one along the direction one) by applying linear decision boundaries. In the following, we briefly recall the main concepts of CVA and RN detection in the polar domain. These concepts are fundamental for the proposed technique.

Let us consider two VHR multispectral images X_1 and X_2 acquired over the same geographical area at different times at t_1 and t_2, respectively. Let us assume that they have been already coregistered according to any state-of-the-art coregistration method and that the images do not show significant radiometric differences to successfully apply CVA (the reader is referred to [32] for further details on the advantages of having radiometrically corrected images). The multispectral difference image is computed based on the CVA technique by subtracting the spectral feature vectors associated with each corresponding pixel in the two considered images. The multispectral difference image X_Δ is made up of spectral change vectors (SCVs) defined as

$$X_\Delta = X_2 - X_1. \tag{1}$$

Under the assumption of working into a 2-D feature space, the change information contained in the SCVs can be described by the magnitude ρ and the direction ϑ of change vectors, i.e.,

$$\rho = \sqrt{(X_{\Delta,1})^2 + (X_{\Delta,2})^2}, \rho \in [0, \rho_{\max}] \tag{2}$$

$$\vartheta = \tan^{-1}\left(\frac{X_{\Delta,1}}{X_{\Delta,2}}\right), \vartheta \in [0, 2\pi) \tag{3}$$

where $X_{\Delta,b}$ represents the bth component (spectral band) of $X_\Delta (b = \{1, 2\})$, and $\rho_{\max}$ is the highest magnitude value of SCVs in the considered images. The change information for a generic pixel in spatial position (x, y) can be represented in the magnitude-direction domain by its components $\rho(x, y)$ and $\vartheta(x, y)$.

In [32], it has been shown that, in the polar representation, unchanged SCVs have a low magnitude and cluster around

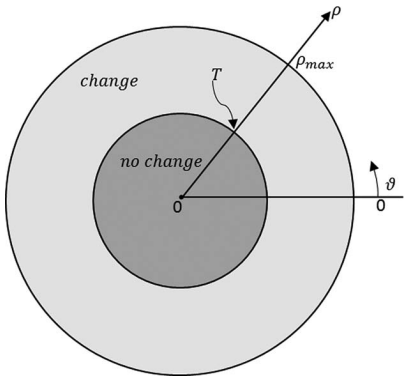


Fig. 1. Representation of the changed and unchanged decision regions in the CVA polar domain.

the origin, whereas changed SCVs have a high magnitude and cluster far from the origin. Along the magnitude variable SCVs $X_\Delta(x,y)$ can be labeled as changed or unchanged according to the following rule:

$$X_\Delta(x,y) \in \begin{cases} \text{change}, & \text{if } \rho(x,y) \geq T \\ \text{no change}, & \text{otherwise} \end{cases} \quad (4)$$

where T is a threshold value that can be set according to any of the methods proposed in the literature [33]–[35]. T divides the polar domain into a circle of unchanged samples centered at its origin and with a radius equal to the threshold T (dark gray area in Fig. 1), and an annulus of changed samples having inner radius equal to T (light gray area in Fig. 1). As changed SCVs show preferred directions according to the kind of change occurred on the ground, different kinds of changes can be isolated within the annulus along the direction variable.

The concepts above have been used in [30] to analyze the effects and properties of RN. On one hand, RN affects unchanged pixels by increasing the spread of the cluster in the circle of unchanged pixels with respect to the case of perfectly aligned images. On the other hand, dominant RN shows in the annulus of the changed pixels properties similar to those of changed pixels. The latter is the most critical RN component from the multitemporal analysis point of view as it is caused by the comparison of misaligned pixels that belong to different objects in X_1 and X_2 and results in critical errors. In [30], the RN multiscale properties were analyzed. Multiscale analysis was conducted by decomposing the multitemporal images according to a multiscale transformation (a stationary wavelet transform (SWT) with a recursively upsampled bicubic filter). CVA was applied to each pair (X_1^n, X_2^n) of images, where the superscript $n (n = 0, 1, \ldots, N)$ indicates the resolution level (note that $X_1^0 \equiv X_1$). The analysis of the distribution of the direction of SCVs at different resolution levels showed that the clusters of dominant RN in the annulus of changed pixels exhibit unstable statistical properties versus the scale of the images. Their impact on the statistical distribution of the direction variable decreases when the scale decreases. On the contrary, real changes tend to be stable with the scale. According to this observation, it is possible to identify dominant RN by comparing the probability density function of pixels in the annulus of changed samples at different resolutions. The conditional density of RN distribution $\hat{p}^{\mathrm{RN}}(\vartheta|\rho \geq T)$ has thus been derived as [30], [36]

$$\hat{p}^{\mathrm{RN}}(\vartheta|\rho \geq T) = C\left[P^0(\rho \geq T)\hat{p}^0(\vartheta|\rho \geq T) - P^N(\rho \geq T)\hat{p}^N(\vartheta|\rho \geq T)\right] \quad (5)$$

where $P^0(\rho \geq T)$ and $P^N(\rho \geq T)$ are the probabilities of SCVs having values in the magnitude domain higher than T at resolution level 0 (original image) and at lower level N, respectively; $\hat{p}^0(\vartheta|\rho \geq T)$ and $\hat{p}^N(\vartheta|\rho \geq T)$ are the marginal conditional densities of the direction variable of the same SCVs at full and Nth resolution, respectively; and C is a constant defined such that $\int_0^{2\pi} \hat{p}^{\mathrm{RN}}(\vartheta|\rho \geq T)d\vartheta = 1$. In (5), the marginal conditional density is estimated according to the Parzen window technique with a Gaussian kernel as [37]. It is worth noting that a high value of $\hat{p}^{\mathrm{RN}}(\vartheta|\rho \geq T)$ corresponds to a high probability that the pixel is contaminated by RN (i.e., high difference among distributions at different resolution levels).

III. Proposed Approach to Fine Image Coregistration

In this section, we propose a fine coregistration technique based on the RN estimated in the CVA polar domain. The proposed method is designed for correcting residual misalignment that affects multitemporal image pairs after the state-of-the-art coregistration techniques have been applied. Most of the research on coregistration between VHR multitemporal images has focused on increasing the overall registration accuracy. Such approaches consider residual misregistration between images as an inevitable error. Here, instead, we focus on residual misregistration as a source of information. The aim is to take advantage of RN to increase the registration accuracy. Under the assumption of multitemporal images being already coregistered by state-of-the-art methods, the proposed technique is applied according to three steps: 1) CP extraction based on RN distribution; 2) CP matching according to a local displacement analysis; and 3) generation of the deformation map and image warping. Fig. 2 represents the block scheme of the proposed coregistration procedure. In the following, a detailed description of each step is given.

A. CPs Extraction Based on RN Distribution

RN describes the effects of the residual misregistration between VHR multitemporal images (see Section II and [30] for further details) after coregistration has been applied. Residual local misalignments depend on the stability of the acquisition platform, the off-nadir angle of the sensor, the structure of the considered scene, seasonal change, and so on. They act as an obstacle when the multitemporal images are employed into

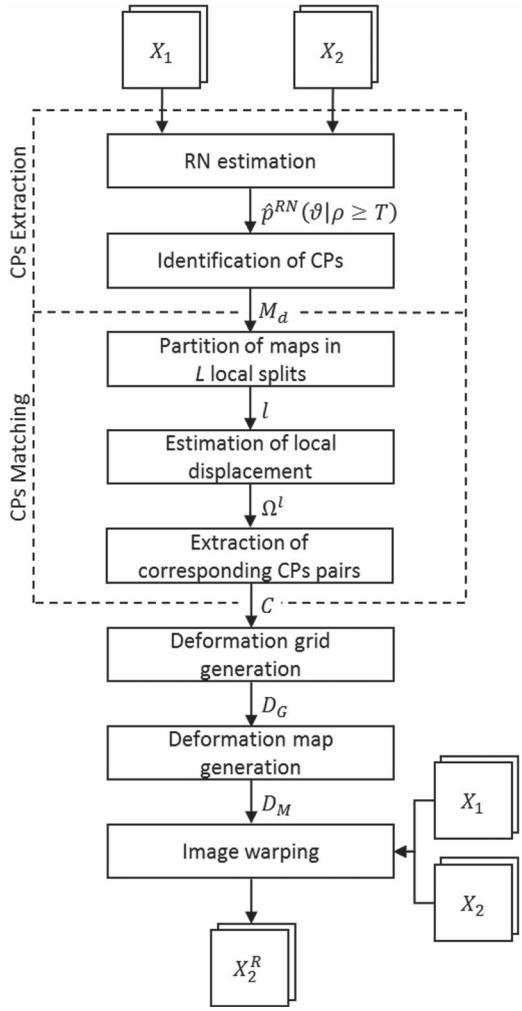


Fig. 2. Block scheme of the proposed coregistration procedure.

diverse applications. However, the coregistration result can be improved if the regions affected by misalignment are identified and geometrically corrected. This can be done by analyzing the distribution of RN estimated along the direction variable in CVA. Once RN pixels are extracted, they are employed as CPs to reduce the misalignment effects in the multitemporal images. The proposed strategy, carrying out the registration process by directly focusing on the critical pixels affected by misalignment, is able to improve the coregistration accuracy with respect to standard methods.

The conditional density of RN is estimated according to (5). The distribution $\hat{p}^{\mathrm{RN}}(\vartheta|\rho \geq T)$ includes contributions from both dominant and nondominant RN. However, only dominant RN contributions are considered for CP extraction as they are the most critical from the multitemporal analysis point of view [30], [31]. The dominant RN pixels are those having high values of $\hat{p}^{\mathrm{RN}}(\vartheta|\rho \geq T)$. They can be separated from the nondominant RN pixels by thresholding the RN density. To this end, a threshold T_{RN} is applied to the conditional density of RN. All the SCVs falling in direction intervals where $\hat{p}^{\mathrm{RN}}(\vartheta|\rho \geq T)$ is higher than T_{RN} are labeled as CPs and employed for correcting the locally misaligned regions.

In a general feature-based matching process for VHR images, CPs are extracted on each image and matched themselves by directly using intensity values of their neighboring pixels or by generating description vectors to estimate similarity. On the contrary, here, CPs are extracted by taking advantage of multitemporal correlation: They are the pixels showing misalignment between images. In other words, they are not extracted on each image independently. There is no estimation on the position of CPs on each image. Therefore, it is not possible to detect CPs correspondence, but we can estimate the amount of displacement between the two images that characterizes each CP. To this end, we consider different possible displacements between the two images. Let $\Omega = \{\Omega_1, \ldots, \Omega_D\}$ be the set of considered D displacements. Let us assume that each displacement Ω_d only models rigid translations in x and y direction, i.e., $\Omega_d = \{\Delta x_d, \Delta y_d, d = 1, \ldots, D\}$ and that displacements are small (i.e., $\Delta x_d, \Delta y_d$ are less than few pixels). These assumptions are both reasonable since images have been initially coregistered. Thus, large differences in scale and rotation between images have been already corrected, and residual small-scale and rotation differences can be locally approximated as rigid translations [38]. Such rigid translations are expected to be small as state-of-the-art coregistration methods provide in general good performance. Once Ω has been defined, we create a set of possible displacements of the master image X_1 by taking the slave image X_2 and translating it according to the set of misalignment values Ω. Let $X_2^D = \{X_2^d,\ d = 1, \ldots, D\}$ be the set of slave images after translation by the D displacements in Ω. For each pair made up of the master image X_1 and the slave image $X_2^d (X_2^d \in X_2^D)$, we derive the conditional density $\hat{p}_d^{\mathrm{RN}}(\vartheta_d|\rho_d \geq T)$ and generate the RN map M_d. When the displacement is d, the RN map $M_d (d = 1, \ldots, D)$ is defined as

$$M_d(x,y) = \begin{cases} 1, & \text{if } \hat{p}_d^{\mathrm{RN}}(\vartheta_d(x,y)|\rho_d(x,y) \geq T) \geq T_{\mathrm{RN}} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where $\rho_d(x,y)$ and $\vartheta_d(x,y)$ are the magnitude and direction values of SCVs at coordinates (x,y) in a given scene, under the condition that the displacement is d.

B. CP Matching According to a Local Displacement Analysis

In order to match the CPs, we perform a local analysis of the RN maps M_d associated to the D displacements in Ω. Let us divide each RN map into L subimages of size $p \times q$. For each split $l (l = 1, \ldots, L)$, we estimate amount of misregistration AM_d^l as the number of misaligned pixels resulting for a given displacement d as

$$\mathrm{AM}_d^l = \sum_{\substack{1 \leq x \leq q \\ 1 \leq y \leq p}} M_d^l(x,y) \quad (7)$$

where $M_d^l(x,y)$ indicates whether the pixel with coordinates (x,y) in the lth split of the dth RN map has been labeled as misregistered according to (6).

The variable computed in (7) can be used to estimate the relative position between corresponding CPs (i.e., pixels showing misalignment in the original master–salve pair images X_1, X_2).

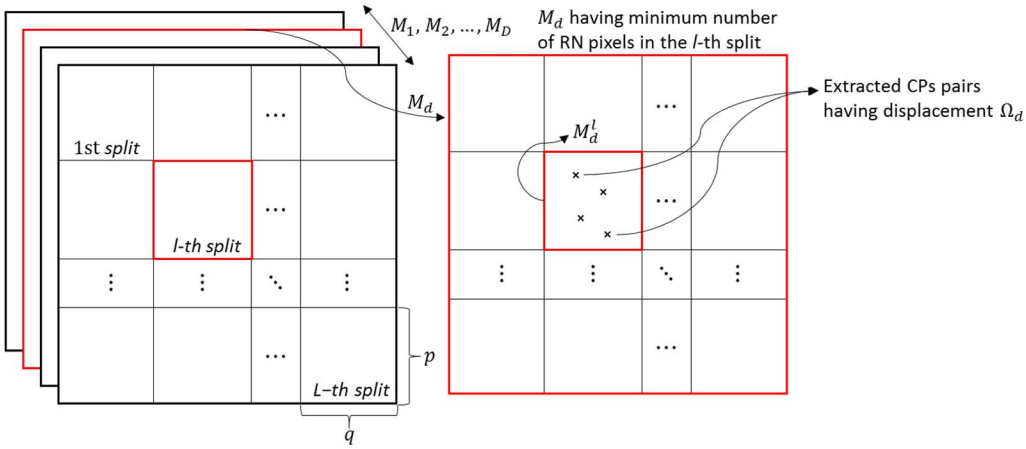


Fig. 3. CP generation from RN pixels by local analysis.

If there is a split for which few CPs are extracted compared with other partitions, this split is less misaligned and thus more precisely coregistered. This concept can be applied to estimate the amount of displacement in a generic local region. Thus, the displacement $\Omega_d \in \Omega$ associated to the minimum number of misaligned pixels AM_d^l is selected as the displacement for all the CPs in the lth split. The local displacement Ω^l associated with the lth split is computed as

$$\Omega^l = \arg\min_{\Omega_d \in \Omega} \left\{ \mathrm{AM}_d^l \right\}. \tag{8}$$

It is important to remark that, despite AM_d^l is estimated by considering all the pixels with nonzero values in M_d^l in the lth split, the final displacement value is associated to CPs only. The displacement value is explicitly derived for each split that contains CPs (i.e., RN pixels) in the original master–slave pair of images X_1, X_2. Let $C = \{c_1^m, c_2^m\},\ m = 1, \ldots, M$, be the set of CP pairs, and let $(x_{1,l}^m, y_{1,l}^m)$ be the spatial position of the mth generic CP c_1^m in the master image extracted in the lth split. The spatial position $(x_{2,l}^m, y_{2,l}^m)$ of the corresponding CP c_2^m in the slave image is defined as

$$\begin{cases} x_{2,l}^m = x_{1,l}^m - \Delta x_l \\ y_{2,l}^m = y_{1,l}^m - \Delta y_l \end{cases}, \quad \forall\, l = 1, \ldots, L, \quad \forall\, m = 1, \ldots, M \tag{9}$$

where $\Omega^l = \{\Delta x_l, \Delta y_l\}$ is the independently estimated local displacements in x-direction and y-direction in the lth split, and M is the total number of CPs (i.e., $M = \Sigma_{l=1}^{L} m_l$, where m_l is the number of CPs in the lth split). All the CPs belonging to the same split are associated to the same displacement. Accordingly, the size of the split assumes a relevant role. A small size results in higher registration accuracy since it guarantees a more precise estimation of local residual displacement. However, this increases the time for completing the registration process. On the other hand, a large split reduces the computational burden at the cost of more rough estimation of the displacement. Fig. 3 shows the concept for the generation of matched CPs from the RN map with a local analysis of the displacement.

C. Generation of the Deformation Map and Image Warping

Depending on the kind of deformation to be corrected, a different transformation model should be employed. In the case of standard coregistration in VHR multitemporal images, analytical models (which are physical models related to the sensors or parametric models estimated by taking CP pairs) are usually applied [29]. However, they are not suitable for a fine registration that mitigates the local residual distortion. In this case, fine sampling of the deformation is necessary. Thus, we generate a deformation map D_M using CP pairs. The deformation map is represented by a displacement vector associated to every pixel of the master image [29]. The CP pairs extracted by the proposed technique are irregularly scattered in the image. Thus, interpolation is used to estimate the deformation in an appropriate way for the irregular CPs. We apply a natural neighbor interpolation to generate a deformation grid D_G, which is a 2-D vector of regularly sampled displacements in the x-direction and y-direction [39]. The natural neighbor method determines the value of a regular point through a weighted average of the values of the neighboring CPs. These neighbors are determined by Voronoi regions generated from the original point set [39]. From the interpolation method, we generate the deformation grid with regular sampling. The grid has the same size as the split dimension. Finally, cubic spline interpolation method is applied to the deformation grid to generate the deformation map. The last step consists in the warping of the slave image to the master one according to the obtained deformation map D_M. Bilinear interpolation is employed for estimating pixel values at each position. The warped slave image X_2^R is generated as

$$X_2^R = D_M\left(D_G(X_2)\right). \tag{10}$$

IV. Data Set Description and Design of Experiments

A. Data Set Description

In order to evaluate the effectiveness of the proposed fine coregistration method and its usefulness in the data set acquired by different sensors, we used multitemporal images acquired over the city of Trento (Italy) from the QuickBird and WorldView satellite multispectral sensors. The QuickBird images have a

Fig. 4. Data sets acquired over the city of Trento, Italy. QuickBird data set: (a) Master image (July 2006). (b) Slave image (October 2005). WorldView data set: (c) Master image (August 2010). (d) Slave image (May 2011). All the images are shown in a true-color composition (i.e., 3, 2, and 1 bands are assigned to RGB channels in the QuickBird data set, and 5, 3, and 2 bands are assigned to RGB channels in the WorldView data set).

panchromatic band with 0.6-m spatial resolution and four multispectral bands [blue (450–520 nm), green (520–600 nm), red (630–690 nm), and NIR (760–900 nm)] with 2.4-m spatial resolution. The WorldView images have a panchromatic band with a spatial resolution of 0.5 m and eight multispectral bands. Compared with QuickBird, there are four additional spectral bands: coastal (400–450 nm), yellow (585–625 nm), red edge (705–745 nm), and NIR 2 (860–1040 nm) bands with 2.0 m spatial resolution. From these images, we generated one simulated and two real data sets. The proposed method was first applied to the simulated data to investigate its properties and effectiveness in a controlled environment. Then, it was applied to the real multitemporal data to demonstrate its practical application performance.

The master image (X_1) of the simulated data set is the QuickBird image acquired in July 2006 with 14.1° off-nadir angle. It is made up of 1000×1000 pixels [see Fig. 4(a)]. The slave image (X_2) is constructed from the master one by including a deliberate distortion in order to evaluate the effectiveness of the proposed coregistration method. Different deformations both in vertical and horizontal directions with a sinusoidal transform were applied to model the nonlinear distortions. Coregistration results obtained for deformations within displacements from 0 to 5 pixels in both horizontal and vertical directions were similar. Here, we report the result obtained with a distortion of sinusoidal deformation in the negative horizontal direction with 5-pixel amplitude and 100-pixel period, and in positive vertical direction with 3-pixel amplitude and 150-pixel period, respectively. The distorted slave image is resampled to the same size of the master one by a bilinear interpolation.

The first real data set has as a master image, i.e., the 2006 QuickBird scene (X_1). The slave image (X_2) is the QuickBird image collected in October 2005 with 9.8° off-nadir angle [see Fig. 4(b)]. The size of the slave image is 1200×1200 pixels, so that it fully covers the area of the master image. The second real data set is made up of two WorldView multispectral images acquired over the same area of the QuickBird images. The master image (X_1) was acquired in August 2010 with 19.3° off-nadir angle and has a size of 1200×1200 pixels [see Fig. 4(c)]. The slave image (X_2) was taken in May 2011 with 12.9° off-nadir angle and has a size of 1400×1400 pixels in order to cover the master image [see Fig. 4(d)]. The test site includes various objects such as small buildings and roads, and land covers such as agricultural fields, bare soil, etc. A visual analysis pointed out that the upper part of the scene is relatively homogeneous since it mainly shows agricultural fields. A smaller number of RN pixels and thus CPs are expected here. The bottom part of the scene shows high complexity and heterogeneity since it is associated to an urban area. Here, a larger amount of RN pixels and thus of CPs is expected.

Before applying the proposed fine coregistration approach, multitemporal images were pansharpened by the widely used

Gram–Schmidt method [40] and were radiometrically corrected. The use of radiometrically corrected images improves the accuracy of the CVA and RN detection. Indeed, it reduces the impact of false alarms. Thus, detected CPs show a higher reliability. Any standard radiometric correction technique can be applied. However, for the sake of simplicity and following [32], here, we subtract the mean value of each spectral band. Then, the multitemporal images were coregistered by a state-of-the-art technique. The latter operation is not applied to the simulated data set since the amount of deformation already small. In the real data set, coregistration is applied by two SIFT-based matching methods [13], [41]. Detailed explanation of the SIFT methods and how they are used in our experiments is presented in the following.

B. Reference Coregistration Technique: SIFT

The SIFT method is a widely used matching algorithm that extracts CP pairs for estimating a transformation model in order to fulfill coregistration [13]. It consists of three steps, i.e., feature extraction, feature description, and feature matching. In the feature extraction step, the Laplacian with difference-of-Gaussian filters is approximated in order to extract the features. Let $X(x, y)$ be an image, then the Laplacian image with Gaussian filter $L(x, y, \sigma)$ is defined as

$$L(x, y, \sigma) = G(x, y, \sigma) * X(x, y) \tag{11}$$

where $*$ denotes the convolution operation in x and y, and $G(x, y, \sigma)$ is a Gaussian filter. Then, local extrema of difference images made by the difference-of-Gaussian (DOG) function are extracted as candidate features. The DOG function $D(x, y, \sigma)$ with the image $X(x, y)$ is defined as

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * X(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{12}$$

where k denotes a constant multiplicative factor with two nearby scales in the scale space constructed by the DOG function. A precise model is fitted to determine the features exact position and scale. The main orientation of each feature is identified based on its local image properties, and the features are then described based on a region of pixels in its neighborhood that are rotated to each orientation to achieve a rotation-invariant descriptor with a 128-element vector. The feature-matching step adopts a minimum Euclidean distance for the descriptor vector for each feature of the master image to find the nearest neighbor in the slave image to its corresponding feature. To establish correct matching, the ratio of the distances to the closest and second closest neighbors should be less than a predefined threshold. More detailed explanation of the SIFT method is introduced in [13].

One of the problems when the SIFT method is applied to VHR images is that all CPs in a corresponding image are candidates for the extraction of matching point regardless of their relative position. To mitigate this effect, we carried out the alignment of the images in the real data sets by the SIFT method with restricted search space for the matching of CPs (we call this process "modified SIFT method" from now on) [41]. To apply the modified SIFT-based alignment, the threshold of the minimum Euclidean distance between the description vectors is set to 0.7, and the ratio of the closest and second closest distances is set to 0.7, referring to [13]. By considering the geometric accuracy of the VHR data, the radius of the circular buffer is set by multiplying the scale of each feature of the master image by 50 pixels (around 25–30 m). A projective transformation is used to align the images [42], [43], which is well known to be appropriate for the geometric correction of VHR images. Outliers, which are extracted from geometrically different position between images, are removed by applying the random sample consensus (RANSAC), which is one of the representative estimation methods of the transformation model [44]. The slave image is then aligned to the coordinates of the master image by the estimated projective transformation. After applying both standard and modified SIFT coregistration, the root mean square errors (RMSEs) of real data sets were less than 1 pixel, except for the case in which the standard SIFT method was applied to the QuickBird data set. In this case, a reasonable warped image could not be generated due to the large number of false-matched points detected by the standard SIFT method. In detail, the RMSE of the QuickBird data set after the modified SIFT method was 0.83 pixels, and the RMSEs of WorldView data set were 0.82 and 0.65 pixels after applying the standard and the modified SIFT methods, respectively.

C. Performance Evaluation Indexes

The master and slave images of the simulated data set are derived from the same QuickBird acquisition. Thus, if all spatial distortions introduced in the slave image are corrected by the coregistration approach, the two images become the same. Therefore, the coregistration accuracy can be evaluated by estimating the correlation coefficient (CC) and the normalized mutual information (NMI), which are representative of the similarity between the two inputs. The CC value of images X_1 and X_2 is calculated as

$$\mathrm{CC}(X_1, X_2) = \frac{\sigma_{X_1 X_2}}{\sqrt{\sigma_{X_1} \sigma_{X_2}}} \tag{13}$$

where $\sigma_{X_1 X_2}$ denotes covariance between the two images, and σ_{X_1} and σ_{X_2} are the standard deviations of the two images, respectively. The CC can range from -1 to 1, with 1 indicating perfect correlation (i.e., the two inputs are exactly the same images).

The NMI index measures the statistical correlation of images. The NMI value is defined as follows:

$$\mathrm{NMI}(X_1, X_2) = \frac{H(X_1) + H(X_2)}{H(X_1, X_2)}. \tag{14}$$

Here, $H(X_1)$ and $H(X_2)$ are the entropy of images X_1 and X_2, respectively, and $H(X_1, X_2)$ is the joint entropy of the two images. Larger values correspond to a greater similarity between two images.

TABLE I
RELIABLE RANGES OF PARAMETERS USED IN THE PROPOSED APPROACH

Parameters	Values
Spectral bands	Data set dependent
Multiscale level N	3, 4
Threshold for magnitude T	Automatic selection
Threshold for RN density T_{RN}	$[10^{-4}, 10^{-3}]$
Displacements Ω_d (pixels)	[-5, +5]
Sampling interval for Ω_d (pixels)	[0.5, 1.0]

The real data sets consist of multitemporal images acquired at different times and under different conditions so that they have different radiometric properties. Therefore, the similarity-based indexes cannot be employed for evaluating the coregistration results. For the numerical assessment of performance on the real data sets, the RMSE and its standard deviation (STD) are estimated over checkpoints extracted by experienced photo interpreters. Let $(\Delta x_i^c, \Delta y_i^c)$ be the residual difference on a checkpoint pair, then the RMSE and its STD are calculated as

$$\text{RMSE} = \frac{1}{M}\sum_{i=1}^{M}\sqrt{(\Delta x_i^c)^2 + (\Delta y_i^c)^2} \tag{15}$$

$$\text{STD} = \sqrt{\frac{\sum_{i=1}^{M}\left(\sqrt{(\Delta x_i^c)^2 + (\Delta y_i^c)^2} - \text{RMSE}\right)^2}{M-1}} \tag{16}$$

where M is the total number of checkpoints.

V. EXPERIMENTAL RESULTS

In order to assess the effectiveness of the proposed method, we applied it to the preprocessed simulated and real data sets. To this end, some parameters need to be fixed. The proposed approach demonstrated to be robust within a reasonable range of variations of such parameters. Table I provides suggestions resulting from a wide range of experiments conducted on the available data. In the succeeding section, an analysis of the impact of the split size on the fine registration performance is provided. The rest of the parameters are fixed as follows. RN has been identified according to [30] by selecting two spectral bands. After some experiments, the red and NIR bands demonstrated to be the most effective ones for the considered data set (however, any other pair can be selected without any loss of general validity). Multiscale decomposition was conducted by Daubechies-4 SWT [45], but also in this case, any other methods can be applied. Three levels $(N = 3)$ were computed. The decision threshold T on SCVs magnitude was automatically selected by applying Bayesian thresholding [46]. Note that, when very large images are considered, more sophisticated approaches such as the split-based one [47] can be used. T_{RN} was set to 10^{-4}. To estimate the local displacement for each CP, a set of displacement values Ω_d was defined by translating the slave image in the x-direction and y-direction from -5 to $+5$ pixels with 0.5-pixel interval using a bilinear interpolation. The range of displacements to be investigated is fixed according to the maximum expected residual misalignment after the application of state-of-the-art coregistration techniques.

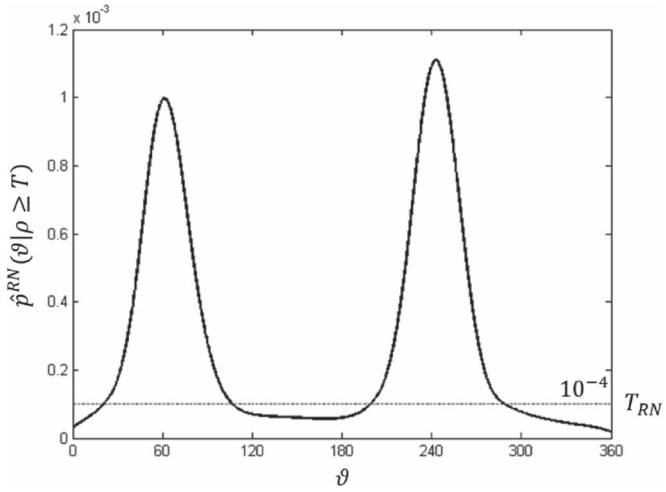


Fig. 5. Estimated conditional density of RN for the simulated data set.

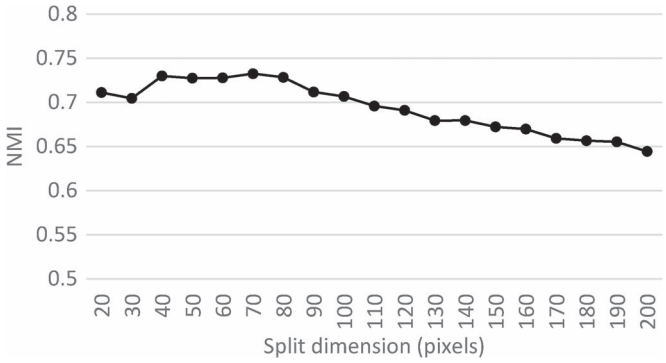


Fig. 6. NMI values versus the split dimension (simulated data set).

A. Results: Simulated Data Set

Fig. 5 shows the conditional density of RN derived according to (5) for the simulated data set. The pixels having $\hat{p}^{\text{RN}}(\vartheta|\rho \geq T)$ larger than the threshold T_{RN} were extracted as candidate CPs. RN maps M_d were derived for each displacement d. Each M_d was divided into splits of size $p \times q$. Then, the amount of extracted CPs at each split l, i.e., AM_d^l, versus the amount of displacement was computed. To observe the performance versus the split dimension, we considered different regular sizes (i.e., $p = q$) of splits and calculated the NMI values between the two images after coregistration. Fig. 6 shows the NMI values versus the split size ranging between 20×20 and 200×200 pixels. As one can see, the NMI values are relatively stable when varying the split size. However, as expected, smaller sizes of split guarantee slightly better results than larger ones.

Based on these results, we fixed the optimal split size to 50×50 pixels. Fig. 7(a) shows the extracted CPs as yellow cross on the master image. As expected, CPs are mainly extracted from areas that contain boundaries (in the considered scene, they are mainly associated to buildings and roads in the urban area). Few CPs were extracted from more homogeneous areas (in this case, mainly the agricultural fields). From the extracted CPs, we generated the deformation grids in the x-direction and y-direction by applying the natural neighbor interpolation with

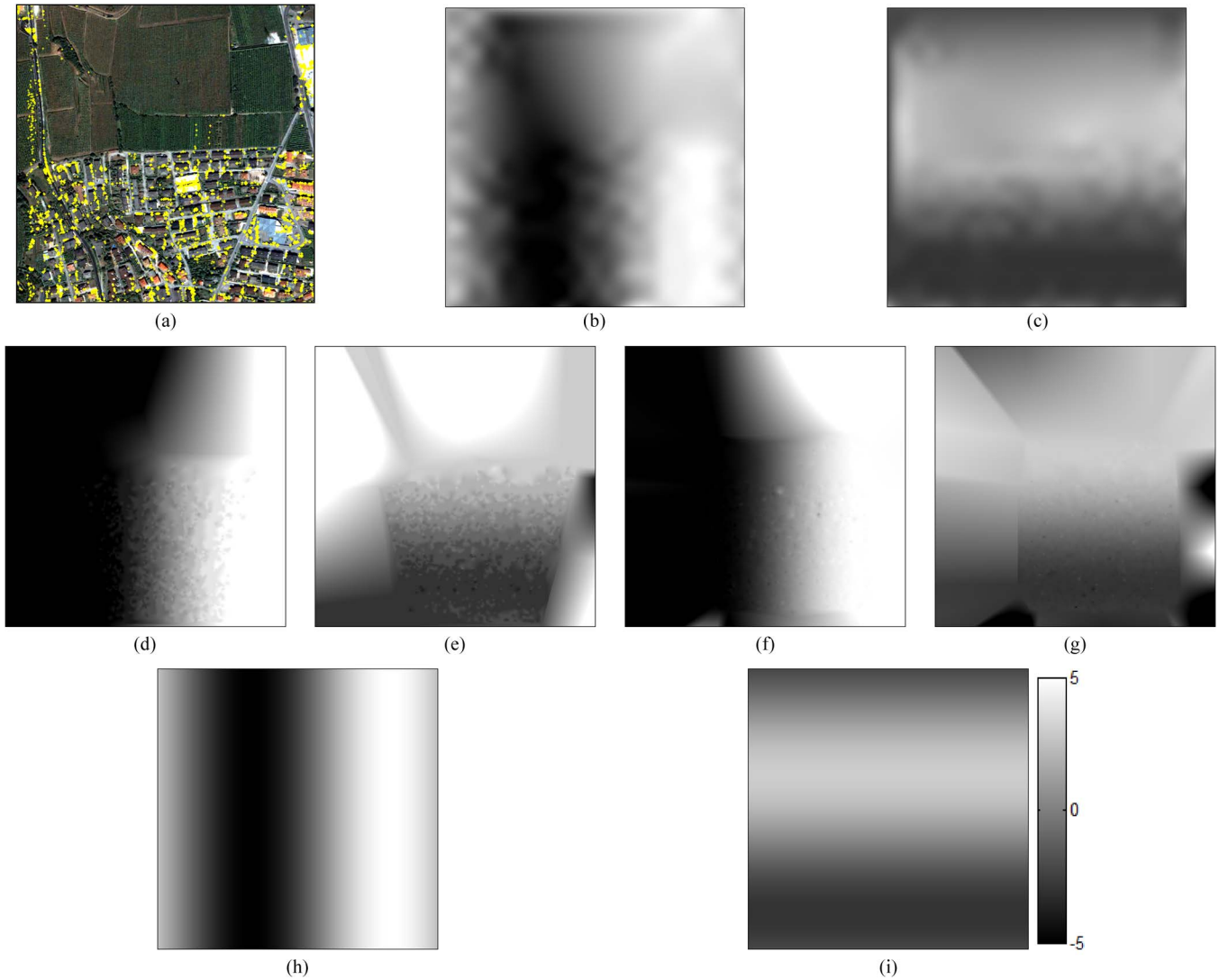
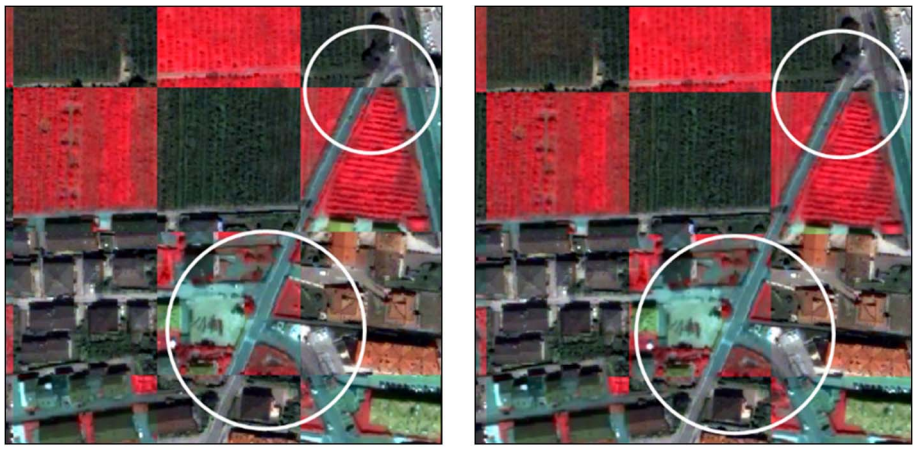


Fig. 7. Comparison of deformation maps in the simulated data set. (a) Extracted CPs from the proposed method and generated deformation map from these CPs in (b) x-direction and (c) y-directions. Deformation maps generated from standard SIFT method in (d) x-direction and in (e) y-direction. Deformation maps generated from the modified SIFT method in (f) x-direction and in (g) y-direction. Reference deformation maps in (h) x-direction and in (i) y-direction.

the same sampling interval of the split dimension. The deformation map was generated by applying the spline interpolation to the grids [see Fig. 7(b) and (c)]. For comparison, deformation maps were generated using CPs extracted by both the standard [see Fig. 7(d) and (e)] and modified [see Fig. 7(f) and (g)] SIFT methods. The proposed approach estimated a deformation map very similar to the reference deformation map [see Fig. 7(h) and (i)], i.e., the one applied to simulate the multitemporal data set. On the contrary, the maps generated from SIFT-based methods were not able to model the sinusoidal distortions. It is worth noting that the proposed method effectively generated the deformation map through the whole scene regardless of the distribution of extracted CPs.

Finally, the simulated image was warped to the coordinates of the master image by using the deformation map obtained by bilinear interpolation. For the visual assessment of the coregistration result, we generated a chessboard image in which the block of the master and warped images are repeatedly interlaced [see Fig. 8(b)]. The chessboard image of the simulated data set (before coregistration) is also shown for visual comparison [see Fig. 8(a)]. For the visual relief with the chessboard image, the blocks of the master image appear in true color, whereas the blocks of the slave image appear in false color (NIR, red, and green bands were assigned to the red, green, and blue channels, respectively). From the boundary between neighboring blocks, we can analyze the quality of the registration result achieved by the proposed approach. As one can see from the circled regions in Fig. 8(b), the chessboard image generated from the proposed method precisely aligned the line and shape information along the boundary of the blocks, whereas the result of the original images was not well aligned [see Fig. 8(a)].

Table II shows CC and NMI values. The standard and modified SIFT methods were not able to increase the CC and NMI values compared with the case of original simulated data set (no registration), whereas the proposed method could considerably increase these values. In terms of CC value, the proposed approach corrected almost all distortion within the simulated image because the CC value is close to one (0.9955). The



(a) (b)

Fig. 8. Chessboard images generated (a) before applying the proposed technique and (b) after applying the proposed technique (simulated data set).

TABLE II
CC AND NMI QUALITY INDEXES FOR THE SIMULATED DATA SET

Registration method	CC	NMI
No registration	0.9236	0.6053
Standard SIFT approach	0.9294	0.5977
Modified SIFT approach	0.9241	0.5983
Proposed approach	0.9955	0.7907

NMI value achieved by the proposed approach is 0.7907, which improved of around 0.2 points compared with the standard methods. From the visual and quantitative evaluation, it is worth noting that the two VHR images are precisely coregistered by the proposed coregistration technique, although the data geometrically have local distortions (i.e., caused by sinusoid displacements in the simulated data set).

B. Results: Real Data Set

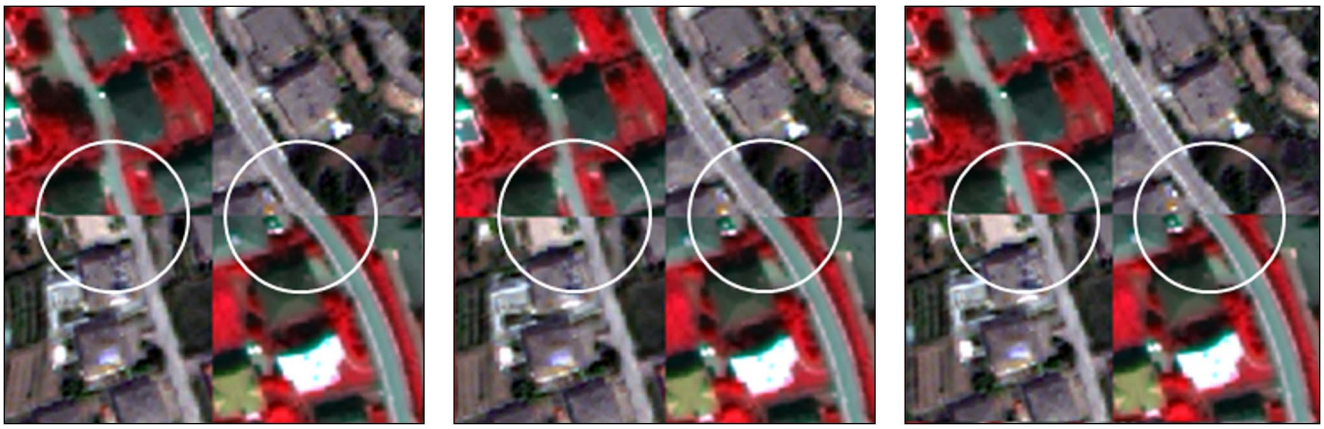
To further validate the proposed technique, experiments were carried out on the two pairs of real multitemporal VHR images described in Section IV. After preprocessing, the proposed method was applied with the same parameters used for the simulated data set. The split size was set to 50×50 pixels . For the comparison, we generated magnified chessboard images of the QuickBird images obtained by the proposed method when the split sizes were set to 50×50 , 100×100, and 150×150 pixels, respectively (see Fig. 9). As expected, boundaries of blocks in the chessboard image generated with 50×50 pixel size aligned better than those with other sizes (see, e.g., the circled region in Fig. 9). The chessboard images of the entire scene generated from the master and coregistered slave images of both data sets are shown in Fig. 10. As one can see along the boundaries between blocks, the master and warped slave images resulted precisely coregistered.

For the numerical assessment, the RMSE and its STD on 20 check points manually extracted from the master and warped images after the registration process have been computed. A comparison is also performed with the coregistration results achieved by the projective transformation estimated using CPs extracted from the standard and modified SIFT methods. The quantitative results are reported in Table III. All the coregistration methods improved the whole registration accuracy, except for the case where the standard SIFT approach was applied to the QuickBird data set. This is because a too large number of false-match points was detected to warp the image.

An analysis of the numerical results points out the effectiveness of the proposed automatic fine coregistration technique, which allows one to accurately coregister the images, obtaining better accuracy than those reached by both SIFT methods. In greater detail, for the QuickBird data set, the proposed fine coregistration sharply reduced the RMSE from 2.7 to 1.2 pixels and the STD value from 1.5 to 0.6 pixels when the modified SIFT approach was used for coregistration. For the WorldView data set, the standard SIFT method achieved an RMSE value around 3 pixels and a STD value of 1.8. The fine coregistration reduced of about three times both values, i.e., the RMSE resulted to be equal only to 1.1 pixels and the STD value became 0.5. Moreover, when the modified SIFT coregistered method was used, the proposed fine coregistration approach decreased the RMSE of about 2 pixels, from 2.9 to less than 1 pixel, and the STD value of about 5 times, from 2 to 0.4. It is worth noting that the proposed method can achieve a more uniform accuracy on the image due to the local analysis carried out according to the CPs extracted by using the property of the RN pixels. This is confirmed by the STD values of the RMSE, which are on both data sets, smaller for the proposed method than for the SIFT-based ones.

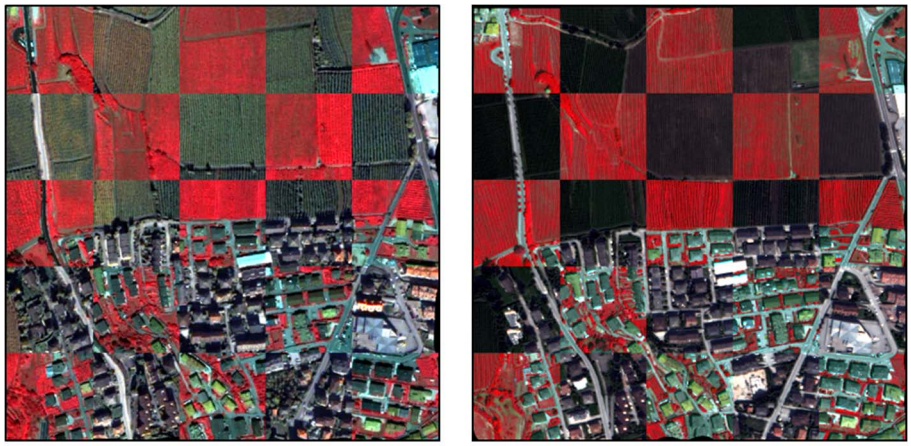
VI. DISCUSSION AND CONCLUSION

In this paper, we proposed a novel approach to fine coregistration between VHR multitemporal images that aims at mitigating the residual misalignments (also referred to as RN) affecting multitemporal images after applying state-of-the-art coregistration methods. RN is estimated according to a CVA-based approach and RN pixels are used as CPs. CP pairs for the warping of the slave image on the master one are derived by a local analysis of RN pixels. The deformation map for



(a) (b) (c)

Fig. 9. Comparison of magnified chessboard image from QuickBird real data set according to split size: (a) 50×50, (b) 100×100, and (c) 150×150 pixels.



(a) (b)

Fig. 10. Chessboard image of real data set generated from the proposed registration technique. (a) QuickBird and (b) WorldView data sets.

TABLE III
REGISTRATION RESULTS OBTAINED ON THE REAL DATA SETS

Co-Registration method		QuickBird dataset		WorldView dataset	
		RMSE (pixels)	STD (pixels)	RMSE (pixels)	STD (pixels)
No registration		25.512	1.659	37.139	7.806
Standard	SIFT method	-	-	3.009	1.818
	modified SIFT method	2.694	1.545	2.862	1.972
Proposed approach	after SIFT method	-	-	1.097	0.486
	after modified SIFT method	1.221	0.598	0.873	0.439

coregistration is generated by the CP pairs. The experimental results obtained both on simulated and real multitemporal VHR data sets confirmed that the proposed method can effectively improve the results achieved by state-of-the-art methods such as the SIFT-based ones.

A simulated data set showing nonlinear small geometric distortions was employed to confirm the capability to mitigate the distortions of the proposed approach. In such a condition, the proposed method showed significant improvement in the coregistration accuracy (i.e., the estimated deformation map was very similar to the one applied for generating the simulated data set and both CC and NMI values increased). On the contrary, the reference state-of-the-art methods could not mitigate the local and nonlinear distortions. This is demonstrated by the CC and NMI values that did not improve. The proposed method is expected to improve the coregistration accuracy of any state-of-the-art method (even those employing nonrigid transformations as piecewise linear functions, thin plate spline, or weighted mean transformation). This is because it conducts a local analysis on the output of the state-of-the-art coregistration techniques. Similar considerations hold for real data sets as well.

The use of the proposed fine coregistration approach implies a few additional minutes as it is applied after the use of state-of-the art coregistration methods. However, the additional computational time is compensated by the sharply higher alignment between images. This is a significant improvement since a

better alignment reduces the errors in the multitemporal image analysis applications.

The proposed method requires fixing a certain number of parameters. However, the method has low sensitivity to the parameter values within a reasonable range. Among the others, the most critical parameter to fix is related to the definition of the split size for the local analysis and displacement estimation. In this paper, we considered only fixed-shaped square splits. Despite that the optimal size can be selected, regular splits include heterogeneous objects. Thus, they may not completely capture the structure of each single object in the scene and the specific impact of local distortions on them. In order to mitigate this issue, future activity will focus on the use of adaptive local neighboring systems being able to capture the nature of each analyzed scene. This can reduce the error on the local displacement estimation and thus on the final coregistration. Another future work is to extend the proposed method to the coregistration of multisensor VHR image pairs.

ACKNOWLEDGMENT

The authors would like to thank Digital Globe Foundation for providing and cross-checking the WorldView images used in the experiments within the "MS-TS—Analysis of MultiSensor VHR image Time Series" project.

REFERENCES

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

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

[3] 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.

[4] J. G. Liu and H. Yan, "Phase correlation pixel-to-pixel image co-registration based on optical flow and median shift propagation," *Int. J. Remote Sens.*, vol. 29, no. 20, pp. 5943–5956, Oct. 2008.

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

[6] V. Arévalo and J. González, "An experimental evaluation of non-rigid registration techniques on QuickBird satellite imagery," *Int. J. Remote Sens.*, vol. 29, no. 2, pp. 513–527, Jan. 2008.

[7] 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.

[8] C. Huo, C. Pan, L. Huo, and Z. Zhou, "Multilevel SIFT matching for large-size VHR image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 171–175, Mar. 2012.

[9] D. Capel and A. Zisserman, "Computer vision applied to super resolution," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 75–86, May 2003.

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

[11] Z. Xiong and Y. Zhang, "A novel interest-point-matching algorithm for high-resolution satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4189–4200, Dec. 2009.

[12] T. Long, W. Jiao, G. He, and W. Wang, "Automatic line segment registration using Gaussian mixture model and expectation-maximization algorithm," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 7, no. 5, pp. 1688–1699, May 2014.

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

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

[15] 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.

[16] Y. Han, Y. Byun, J. Choi, D. Han, and Y. Kim, "Automatic registration of high-resolution images using local properties of features," *Photogramm. Eng. Remote Sens.*, vol. 78, no. 3, pp. 211–221, Mar. 2012.

[17] Y. Huachao, Z. Shubi, and W. Yongbo, "Robust and precise registration of oblique images based on scale-invariant feature transformation algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 783–787, Jul. 2013.

[18] L. Wang, Z. Niu, C. Wu, R. Xie, and H. Huang, "A robust multisource image automatic registration system based on the SIFT descriptor," *Int. J. Remote Sens.*, vol. 33, no. 12, pp. 3850–3869, Jun. 2012.

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

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

[21] 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. 4328–4338, Jul. 2014.

[22] G. Hong and Y. Zhang, "Wavelet-based image registration technique for high-resolution remote sensing images," *Comput. Geosci.*, vol. 34, no. 12, pp. 1708–1720, Dec. 2008.

[23] V. Arévalo and J. González, "Improving piecewise linear registration of high-resolution satellite images through mesh optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3792–3803, Nov. 2008.

[24] Y. Han, J. Choi, Y. Byun, and Y. Kim, "Parameter optimization for the extraction of matching points between high-resolution multisensory images in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5612–5621, Sep. 2014.

[25] 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. 4, pp. 83–95, Apr. 2014.

[26] P. Gong, E. F. Ledrew, and J. R. Miller, "Registration-noise reduction in difference images for change detection," *Int. J. Remote Sens.*, vol. 13, no. 4, pp. 773–779, Mar. 1992.

[27] L. Bruzzone and R. Cossu, "Adaptive approach to reducing registration noise effects," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2455–2465, Nov. 2003.

[28] J. Townshend, C. Justice, C. Gurney, and J. McManus, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 1054–1060, Sep. 1992.

[29] J. Inglada and A. Giros, "On the possibility of automatic multisensor image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2104–2120, Oct. 2004.

[30] F. Bovolo, L. Bruzzone, and S. Marchesi, "Analysis and adaptive estimation of the registration noise distribution in multitemporal VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2658–2671, Aug. 2009.

[31] S. Marchesi, F. Bovolo, and L. Bruzzone, "A context-sensitive technique robust to registration noise for change detection in VHR multispectral images," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1877–1889, Jul. 2010.

[32] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.

[33] L. Bruzzone and D. Fernández-Prieto, "A minimum-cost thresholding technique for unsupervised change detection," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3539–3544, 2000.

[34] L. Bruzzone and D. Fernández-Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1170–1182, May 2000.

[35] T. Celik and K. K. Ma, "Unsupervised change detection for satellite images using dual-tree complex wavelet transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1199–1210, Mar. 2010.

[36] S. Marchesi and L. Bruzzone, "A registration-noise driven technique for the alignment of VHR remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 25–30, 2010, pp. 1023–1026.

[37] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.

[38] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it?-A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[39] R. Sibson, "A brief description of natural neighbour interpolation," in *Interpreting Multivariate Data*. Chichester, U.K.: Wiley, 1981, pp. 21–36.
[40] Y. Zhang and G. Hong, "An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images," *Inf. Fusion*, vol. 6, no. 3, pp. 225–234, Sep. 2005.
[41] Y. Han, B. Kim, Y. Kim, and W. Lee, "Automatic cloud detection for high spatial resolution multi-temporal images," *Remote Sens. Lett.*, vol. 5, no. 7, pp. 601–608, Jul. 2014.
[42] R. E. Kennedy and W. B. Cohen, "Automated designation of tie-points for image-to-image coregistration," *Int. J. Remote Sens.*, vol. 24, no. 17, pp. 3467–3490, Sep. 2003.
[43] 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.
[44] M. Fischler and R. 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.
[45] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[46] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1026–1038, Aug. 2002.
[47] F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1658–1670, Jun. 2007.

Youkyung Han (S'12–M'14) received the B.S. degree in civil, urban, and geosystem engineering and the M.S. and Ph.D. degrees in civil and environmental engineering from Seoul National University, Seoul, Korea, in 2007, 2009, and 2013, respectively.

He is currently a Postdoctoral Researcher with the Remote Sensing for Digital Earth Unit, Fondazione Bruno Kessler, Trento, Italy. His main research interests include image processing of very high resolution remote sensing data, such as image-to-image registration, segmentation, classification, object extraction, and change detection, and the analysis of multispectral/hyperspectral and multitemporal/hypertemporal images.

Dr. Han has served as a Reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *Photogrammetric Engineering and Remote Sensing*, and *Remote Sensing*. He received the International Postdoctoral Fellowship Program of the National Research Foundation of Korea.

Francesca Bovolo (S'05–M'07–SM'13) received the "Laurea" (B.S.) and "Laurea Specialistica" (M.S.) degrees in telecommunication engineering (*summa cum laude*) and the Ph.D. degree in communication and information technologies from the University of Trento, Trento, Italy, in 2001, 2003, and 2006, respectively.

She has been a Research Fellow with the University of Trento until June 2013. She is the Founder and the Head of the Remote Sensing for Digital Earth Unit, Fondazione Bruno Kessler (FBK), Trento, and a member of the Remote Sensing Laboratory (RSLab), University of Trento. Her main research activity is in the area of remote-sensing image processing. Her research interests include multitemporal remote-sensing image analysis; change detection in multispectral, hyperspectral, and SAR images and very high resolution images; content-based time series retrieval; and domain adaptation. She conducts research on these topics within the context of several national and international projects. She is one of the Coinvestigators of the Radar for Icy Moon Exploration (RIME) instrument of the ESA JUpiter Icy moons Explorer (JUICE).

Dr. Bovolo has been a member of the program and scientific committees of several international conferences and workshops, and a referee for several international journals. She served as the Technical Chair for the 2011 Sixth International Workshop on the Analysis of Multi-temporal Remote-Sensing Images (MultiTemp 2011). She served as a Cochair for the SPIE International Conference on "Signal and Image Processing for Remote Sensing" since 2014. She serves as the Publication Chair for the 2015 International Geoscience and Remote Sensing Symposium. She has been serving as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since January 2011. She has been a Guest Editor for the Special Issue on "Analysis of Multitemporal Remote Sensing Data" of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. She ranked first place in the Student Prize Paper Competition of the 2006 IEEE International Geoscience and Remote Sensing Symposium (Denver, August 2006).

Lorenzo Bruzzone (S'95–M'98–SM'03–F'10) received the Laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is currently a Full Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, radar, pattern recognition, and electrical communications. He is the Founder and the Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. He promotes and supervises research for many national and international projects. Among others, he is the Principal Investigator of the Radar for Icy Moon Exploration (RIME) instrument in the framework of the ESA JUpiter ICy moons Explorer (JUICE) mission. He is the author or coauthor of 161 scientific publications in referred international journals (111 in IEEE journals), more than 220 papers in conference proceedings, and 17 book chapters. He is the Editor or Coeditor of 15 books/conference proceedings and one scientific book. His papers are highly cited, as proven form the total number of citations (more than 12400) and the value of the h-index (57) (source: Google Scholar). His current research interests include remote sensing, radar and SAR, signal processing, and pattern recognition.

Dr. Bruzzone was invited as a Keynote Speaker in 24 international conferences and workshops. He has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society since 2009 and has been appointed as a Distinguished Speaker since 2012. He is the Cofounder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. He has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing since 2003. He has been the Founding Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE since 2013. He served as a Guest Coeditor for different special issues of international journals. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the *Canadian Journal of Remote Sensing*. He received the First Place Award in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium. Since that time, he has received many international and national honors and awards.