Title: Segmentation-based Fine Registration of Very High Resolution Multitemporal Images

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

Date of Publication: 14 February 2017

Author(s): Y. Han, F. Bovolo, L. Bruzzone

Volume: 55, Issue: 5

Page(s): 2884 - 2897

DOI: 0.1109/TGRS.2017.2655941

SEGMENTATION-BASED FINE REGISTRATION OF VERY HIGH RESOLUTION MULTITEMPORAL IMAGES

Youkyung Han[1], Francesca Bovolo[1], *Senior Member, IEEE*, and Lorenzo Bruzzone[2], *Fellow, IEEE*

***ABSTRACT*— In this paper a segmentation-based approach to fine registration of multispectral and multitemporal Very High Resolution (VHR) images is proposed. The proposed approach aims at estimating and correcting the residual local misalignment (also referred to as Registration Noise (RN)) that often affects multitemporal VHR images even after standard registration. The method extracts automatically a set of object representative points associated to regions with homogeneous spectral properties (i.e., objects in the scene). Such points result to be distributed all over the considered scene and account for the high spatial correlation of pixels in VHR images. Then, it estimates the amount and direction of residual local misalignment for each object representative point by exploiting residual local misalignment properties in a multiple displacement analysis framework. To this end a multiscale differential analysis of the multispectral difference image is employed for modelling the statistical distribution of pixels affected by residual misalignment (i.e., RN pixels) and detect them. The RN is used to perform a segmentation-based fine registration based on both temporal and spatial correlation. Accordingly, the method is particularly suitable to be used for images with a large number of border regions like VHR images of urban scenes. Experimental results obtained on both simulated and real multitemporal VHR images confirm the effectiveness of the proposed method.**

[1]Fondazione Bruno Kessler, Center for Information and Communication Technology, Trento, Italy
[2]Dept. of Information Engineering and Computer Science, University of Trento, Trento, Italy

***Index Terms*— Registration, Multitemporal images, Object representative points, Registration noise, Very high resolution images, Urban areas, Remote sensing.**

I. INTRODUCTION

Image registration is the process of spatially overlaying two or more images acquired over the same geographical area at different times [1]. In this field, significant advances have been presented in the literature since Very High Resolution (VHR) multitemporal images are available according to the launch of various satellites equipped with VHR sensors (e.g., QuickBird, GeoEye, WorldView). Image registration methods can be divided into two main categories: area-based and feature-based [1]. Area-based methods directly calculate the correlation between the images or a subset of them. The final result is given as the one that maximizes the correlation. Feature-based methods extract salient features from the images and perform salient features matching. In general, the feature-based methods are recommended over the area-based ones for multitemporal VHR images [2],[3]. The effectiveness of such kind of approaches depends on the accuracy of salient features extraction and matching. Thus several studies were conducted on this issue by considering VHR image characteristics [4]–[6]. They mainly focus on the identification of Control Points (CPs) from salient features such as corners, intersections, and edges. Scale-Invariant Feature Transform (SIFT) [7], Speeded-Up Robust Features (SURF) [8], Wavelet decomposition [9]–[11], and Harris points [12] are among the most effective approaches for CPs extraction. These approaches tend to select CPs along boundaries and edges of objects. As there is a large number of salient features in VHR images, the amount of CPs and their concentration become critical. In this situation, the matching process is likely to fail since many CPs have similar spectral properties in their neighboring pixels (which are used to measure similarity for

the matching) even if they do not correspond to each other. A huge number of CP pairs with imprecise matching generates significant local distortions when employed in establishing the transformation function between the input and the reference images. Therefore, despite the overall performance of such methods is good, the registration in regions with a high concentration of salient features (i.e., sharp borders) is likely to be poor. In other words, even after registration, multitemporal images are locally affected by residual misalignment. The impact of this phenomenon increases as the concentration of salient features does. This is the case of scenes acquired over urban areas where manmade objects (e.g., buildings and roads) create a high concentration of salient features. When dealing with VHR images, differences in acquisition conditions of the input and reference images (e.g., acquisition angle) make different sides of the same object to be sensed and the CPs matching less precise. This contributes to residual local misalignment after registration [13], [14].

Non-rigid transformation models for warping [3] may mitigate local and nonlinear distortions. However, for a good registration performance, they require CPs to be distributed over the entire scene, which is a complex task [15]–[18]. Many studies focused on extracting evenly distributed CPs in multitemporal images, but this is still a challenging issue [19],[20].

To overcome the aforementioned problems and improve the performance of registration, some approaches employed segmentation techniques [21]–[24]. Dare and Dowman [21] used spatial attributes of segments, such as area, perimeter, and length and width of the bounding rectangle for CPs matching. The spatial information helps to match the CPs extracted from images acquired by sensors that have different radiometric and geometric properties. Gonçalves et al. applied the segmentation concept for image registration with two objectives: i) to remove small regions that are not appropriate for CPs detection [22]; and ii) to extract spatial attributes of

objects for better registration performance [23]. Troglio et al. extracted ellipsoidal features using watershed segmentation for planetary image registration [24].

In this paper we propose an approach to fine registration which aims at reducing local residual misalignment that affects multitemporal VHR images after applying a standard manual/automatic registration method. As mentioned above, even though standard registration methods can provide high overall registration accuracy, they may cause significant residual misalignment locally [20]. This effect becomes more critical when VHR images are considered and when they are acquired over areas with sharp geometries (e.g., urban areas). Thus the main goal of the proposed approach is to minimize local residual misalignment under the assumption that standard registration has been applied to multitemporal images and that residual distortion/misalignment exists locally and may show different intensity and direction. The proposed approach: i) estimates local residual misalignment, and ii) uses it to further improve the alignment locally by means of a fine-registration step. This is achieved by first identifying a set of points distributed all over the scene that are representative for objects in VHR images (e.g., buildings, roads, fields). Object representative points are defined as centroids of segments computed on the reference image and they account for the spatial correlation of neighboring pixels in VHR images. Differently from standard CPs, object representative points tend to distribute all over the reference image and have a low probability to cluster to each other. At the same time, they are adaptive with respect to the image geometrical content and become denser in sharp geometry areas and sparser other areas. Thus the method effectively chases both sharper and smother geometries and becomes particularly suitable for urban scenes where standard methods tend to show higher local residual misalignment. The object representative points are used as points for applying the relative warping function between the input and the reference

images. The warping function is estimated according to the intensity and direction of residual local misalignment. The estimation procedure relies on a multiple displacement analysis of the local residual misalignment conducted according to multiscale differential analysis of the multispectral difference image proposed in [13],[25],[26]. This allows to model the statistical distribution of pixels affected by residual misalignment and to effectively detect them. This phenomenon is referred in the literature as Registration Noise (RN) [13],[25],[26]. Before warping it, unreliable object representative points are removed by evaluating their relative spatial relation and by identifying unreliable segments associated to shadows.

Summarizing, the proposed approach provides three main contributions: i) Fine registration is carried out for accurate and precise geometric alignment using multitemporal correlation throughout the RN concept. ii) The proposed mechanism accounts for the spatial correlation among neighboring pixels and the complex geometry of VHR images by the use of segments. Thus pixels belonging to the same object are treated in a homogenous way when warping is applied. iii) Object representative points for estimating/applying the warping function are distributed all over the scene allowing for an effective correction of local misalignments. Thus the fine registration produces accurate performance at local level all over the considered scene. Experiments carried out both on simulated and real datasets acquired from VHR sensors confirm the effectiveness of the proposed approach.

The remainder of the paper is organized as follows. Section II illustrates the proposed fine registration technique based on object representative points. Section III describes simulated and real datasets, introduces indices for quantitative performance analysis, and presents the experimental setup. Section IV illustrates the experimental results. Finally, section V draws the conclusion of this paper.

II. PROPOSED SEGMENTATION-BASED FINE REGISTRATION APPROACH

The proposed segmentation-based fine registration approach works on multitemporal VHR images under the assumption that they have been already registered by any manual/automatic registration method. The goal is to mitigate local residual misalignments if any and thus to improve registration accuracy by using both spatial correlation between neighboring pixels available in VHR images and temporal correlation between multitemporal images. Let X_1 and X_2 be the already registered reference and input images collected at different times. The input image (X_2) is the one to be warped to the coordinates of the reference image (X_1) by the proposed fine registration approach. The proposed approach mainly consists of five steps: i) detection of segment-based object representative points; ii) residual local misalignment estimation by multiple displacement analysis; iii) shadow detection; iv) displacement of object representative points; v) input image warping. Figure 1 represents the block scheme of the proposed approach. A detailed description of each step is given in the following subsections.

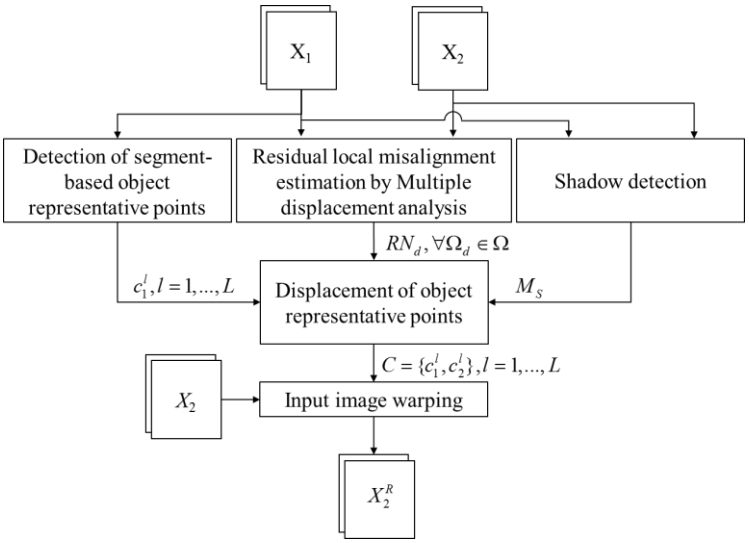


Figure 1. Block scheme of proposed segmentation-based fine registration approach.

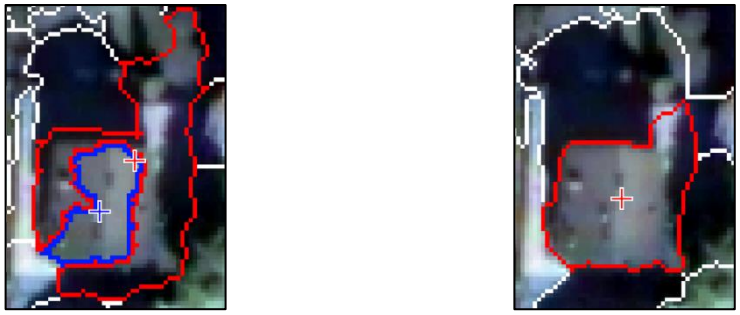
A. Detection of segment-based object representative points

In the first step a set of points is identified to be used for fine registration by estimation of a warping function. Such points should be: i) distributed over the entire scene to capture local behaviors effectively; and ii) representative of objects so that while warping pixels belonging to the same object will be treated homogenously. Representative points for objects (i.e., regions of spatially connected pixels with homogeneous spectral properties) are detected by image segmentation of the reference image X_1. Any segmentation method from the literature can be used. Here, without any loss of general validity, we selected the Simple Linear Iterative Clustering (SLIC) superpixel method since it is computationally efficient and extracted segments adhere well to object boundaries [32]. For sake of completeness we recall below the main steps of SLIC. According to [32], the method is an adaptation of *k*-means clustering to segmentation. First, the image is divided into L hexagonal regions such that their centers are sampled at the same distance with spacing $S = \sqrt{N/L}$ (N is the number of pixels in X_1). The L centers are used as initial seeds for clustering. Each pixel within a limited search region is associated to a cluster/hexagon center by considering the distance measure D that accounts for proximity both in the spatial and spectral domain [32],[33]:

$$D = \sqrt{d_c^2 + \left(\frac{d_s}{S}\right)^2 m^2} \tag{1}$$

where d_c is the color distance between the cluster center and each pixel within a search region of size $2S \times 2S$ computed in the LAB color space [34] (where L is the lightness of the color, and A and B are the colors along red/green and blue/yellow axes), whereas d_s is the spatial distance between the spatial position of two pixels. m is the compactness parameter that balances the relevance between spectral and spatial proximity while segmenting. According to [32], m can

assume values in the range [1; 40]. A pixel is assigned to the cluster that results in the smallest *D*. Once the assignment procedure is complete, cluster centers are updated by computing the mean value of pixels in the cluster. The process is iterated until the cluster changes between subsequent iterations are negligible. Unassigned isolated pixels are linked up to the nearest cluster, however if a large number of adjacent pixels remain unassigned a new cluster is created [32]. At the end, since objects are made of spatially connected pixels that show similar spectral behaviors, the shape of segments adheres the objects boundaries. Thus, the c_1^l centroid (i.e., cluster center) of the *l*th segment ($l = 1, \dots, L$) can be assimilated to the center of the object represented by the segment and assumed as the object representative point in the reference image X_1 [14],[21]. The number L of segments should be large enough to limit the number of large heterogeneous segments that embrace more objects (i.e., under segmentation). Centroids of heterogeneous segments cannot be considered as object representative since they represent more than one object. Accordingly, a certain degree of over segmentation is preferred. The performance improvement (see Sec. IV for a quantitative analysis) compensates for the increase of the computational burden. Moreover, segments should be compact so that object representative points are located inside the segment. Centroids outside the segments cannot be considered as representative for the object. Figure 2 shows an example of centroid location according to different values of the compactness parameter *m*. The centroid (cross mark) and its segment boundary are displayed with the same color. Higher compactness results in centroids inside the segment (i.e., red cross in Figure 2.b) and thus in a representative point for the corresponding object, whereas lower compactness results in centroids outside the segment (i.e., red cross in Figure 2.a) and thus less representative for the object. The impact of compactness on registration performance is illustrated in the experiments (see Sec. IV).



(a) (b)

Figure 2. Examples of centroids location and their boundaries according to different compactness parameters: (a) Lower compactness (*m* = 1) and (b) higher compactness (*m* = 40).

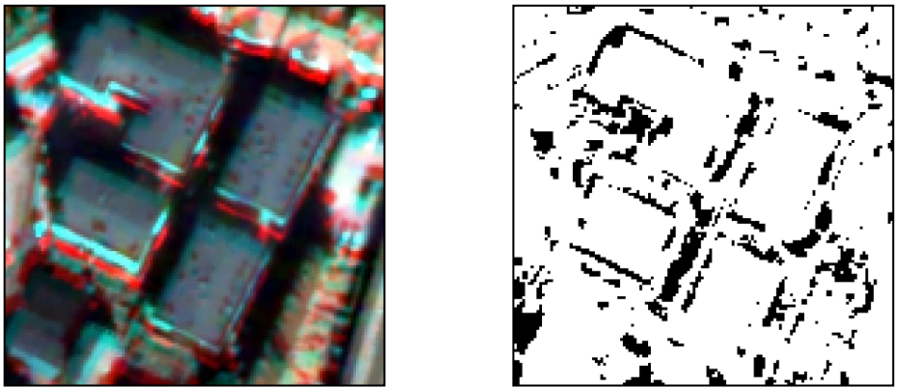
B. Residual local misalignment estimation by multiple displacement analysis

The second step aims at establishing the intensity and direction of local residual misalignment. To this end the concept of RN introduced in [26],[27] is employed. The estimated values are the ones of the correction to be applied to object representative points for fine registration. The analysis relies on change detection and multiple displacement concepts.

RN pixels are defined as pixels that have the same spatial position in multitemporal images, but do not correspond to the same area on the ground due to residual misalignment after registration. They are especially visible along the object borders. In order to illustrate RN behaviors in VHR multitemporal images, Figure 3.a shows a multitemporal false-color composite of images acquired over the same geographical area. Because of misalignment of 5 pixels in both horizontal and vertical directions, red/cyan structures appear parallel to building borders where non corresponding objects overlap. This effect can be modeled as a multitemporal noise component and has been extensively analyzed in [26]. Here we recall some of the main outcomes of that work and we refer the reader to [26] for the details. Since RN is a multitemporal noise component and misaligned samples tend to behave as changed pixels, multitemporal image comparison can be employed to identify it [26]. The detection is conducted in the Change Vector Analysis (CVA) [27],[29] feature space, i.e., the space of the multispectral difference image X_Δ.

CVA computes X_Δ by applying pixel-by-pixel subtraction to X_1 and X_2:

$$X_\Delta = X_2 - X_1 \tag{2}$$



(a) (b)

Figure 3. (a) Multitemporal false-color composite of misaligned VHR images and (b) Registration noise map.

Without loss of generality, let us assume that X_Δ has two spectral bands only. In a 2D feature space, the information in X_Δ can be described in a polar coordinate system by computing the magnitude ρ and the direction ϑ:

$$\rho = \sqrt{\left(X_{\Delta,1}\right)^2 + \left(X_{\Delta,2}\right)^2}, \rho \in [0, \rho_{max}] \tag{3}$$

$$\vartheta = tan^{-1}\left(\frac{X_{\Delta,1}}{X_{\Delta,2}}\right), \vartheta \in [0,2\pi) \tag{4}$$

where $X_{\Delta,b}$ denotes the bth spectral band of X_Δ $(b = \{1,2\})$ and ρ_{max} denotes the highest magnitude value in X_Δ. The ambiguity introduced by the π-periodicy of the inverse tangent function is solved by considering the signs of the X_Δ components. Accordingly, the values of ϑ are properly distributed over to the range $[0,2\pi)$.

According to the literature [27],[29], this feature space can be divided into two main regions by defining a decision threshold T along the magnitude (T can be set by any of the methods available in the literature [28]–[30]). The first one is a circle that embraces samples with a

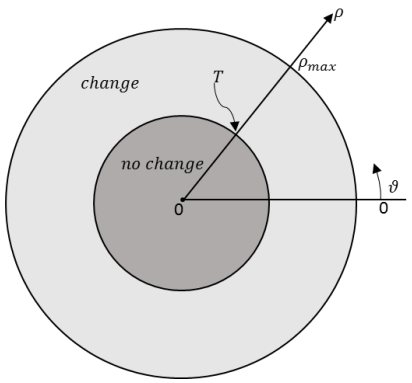


Figure 4. Representation of the changed and unchanged decision regions in the CVA polar domain.

magnitude lower than T (dark grey shaded area in Figure 4). Samples in this circle are the ones with similar spectral signatures in X_1 and X_2 and thus are labeled as unchanged. The second one is an annulus (light grey shaded area in Figure 4) that includes samples with magnitude greater than the threshold. Samples in the annulus are the ones that show different spectral signatures in X_1 and X_2. Such difference in the multitemporal signature may be generated by the presence of changes on the ground or by the comparison between misaligned pixels (i.e., RN pixel). In [26], it is demonstrated that, while smoothing the geometrical details and the sharp object borders in the multispectral difference image, RN pixels tend to reduce their magnitude. Therefore, the smoothed version of the multispectral difference image X_Δ shows a smaller number of samples that along the direction ϑ fall above the threshold T in the light grey shaded annulus. Accordingly, the identification of RN pixels is possible by a multiscale differential analysis of the direction distribution in the polar domain at full resolution and at low resolution N [26]. From the analytical point of view, we can write the conditional density of RN distribution $\hat{p}^{RN}(\vartheta|\rho \geq T)$ as [26]:

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

where $P^0(\rho \geq T)$ and $P^N(\rho \geq T)$ denote the probabilities of SCVs having values in the magnitude domain higher than T at the original image scale and at low resolution level N,

respectively, $\hat{p}^0(\vartheta|\rho \geq T)$ and $\hat{p}^N(\vartheta|\rho \geq T)$ denote the marginal conditional densities of the direction variable of the SCVs at full resolution and at resolution *N*, respectively. *C* denotes a constant defined to satisfy the condition $\int_0^{2\pi} \hat{p}^{RN}(\vartheta|\rho \geq T)d\vartheta = 1$. The statistical variables in (5) can be estimated in an unsupervised way according to [26],[29],[31]. If $\hat{p}^{RN}(\vartheta|\rho \geq T)$ is above a threshold value T_{RN}, the probability of having pixels contaminated by RN is high and thus misaligned samples are detected. Figure 3.b shows an example of RN map obtained for the image pair in Figure 3.a.

In order to use RN information for fine registration, the amount and direction of residual misalignment that affect RN pixels should be estimated. This is achieved by a multiple displacement analysis. Since multitemporal images are already preliminary registered, the largest differences in scale and rotation are mitigated. Therefore, local residual misalignments, including small residual scale and rotation displacements, can be modeled as small rigid shifts [35]. Accordingly, we create a set of possible multiple displacements between the reference image X_1 and the input image X_2 by shifting the latter according to a predefined set of misalignment values while considering the level of residual misalignment between them. Let us assume that $\Omega = \{\Omega_1, \dots, \Omega_D\}$ is the set of D displacements. Each displacement has two components $\Omega_d = \{\Delta x_d, \Delta y_d\}$. Let $X_2^D = \{X_2^d,\ d = 1, \dots, D\}$ be the set of input images after shifting X_2 by the D displacements in Ω. For each combination of the reference image and one of the D shifted input images, we derive the RN conditional density $\hat{p}_d^{RN}$ $(d = 1, \dots, D)$ as presented in (5) and generate the RN map RN_d. When the displacement is *d*, the *d*th RN map RN_d is defined as

$$RN_d(x,y) = \begin{cases} 1, & \text{if } \hat{p}_d^{RN}(x,y) \geq T_{RN} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where (x, y) is the spatial position of samples. The displacement Ω_d that locally minimizes the

amount of misalignment represents the estimation of the local residual displacement to be corrected and thus the displacement to be applied to object representative points to effectively minimize the local residual misalignment between the input image X_2 and the reference image X_1.

C. Shadow detection

Before giving the details of the object representative points displacement and image warping steps, the shadow issue should be managed. Shadows are very common features close to the objects in VHR urban scenes. The spectral signature of shadow pixels shows high interclass similarity and strong differences with adjacent non-shadow pixels. Thus shadow pixels are likely to generate segments and to be associated to an object representative point. However, shadows are not objects and often occlude objects. Thus, centroids in shadow segments cannot be considered as object representative. In addition, due to different acquisition conditions (e.g., Sun azimuth angle, season), multitemporal images may show shadows with different shapes [36]. When performing multitemporal images comparison, such differences may result in samples with high magnitude values that are likely to behave as RN pixels even if they are not. Their contribution to the displacement estimation impacts in a negative way on fine registration performance. Accordingly, shadow pixels should be identified and removed.

Shadow detection is performed by the invariant color models introduced in [37], where the author demonstrates that in the HIS color space, shadows are likely to have: i) low intensity values because the electromagnetic radiance emitted from the Sun is obstructed; and ii) higher hue values than those of adjacent regions from the same surface when hue values are normalized in [0,1]. Accordingly, the shadow index is defined as [37]:

$$SI_i = \frac{H_i + 1}{I_i + 1} \tag{7}$$

where H_i and I_i stand for the hue and intensity values of image X_i $(i = \{1,2\})$. SI_i assumes values greater than a threshold T_{SI} in the presence of shadows, thus:

$$M_{S_i}(x,y) = \begin{cases} 1, & \text{if } SI_i(x,y) \leq T_{SI} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

The multitemporal shadow map M_S associated with both X_1 and X_2 detects a shadow if a shadow exists in at least one of the two images, i.e.,

$$M_S(x,y) = M_{S_1}(x,y) M_{S_2}(x,y) \tag{9}$$

D. Displacement of object representative points

In this step, for each object representative point c_1^l $(l = 1, \dots, L)$ in the reference image, the amount of displacement is established. This information is used to locate a corresponding object representative point c_2^l $(l = 1, \dots, L)$ in the input image for each c_1^l. This is different with respect to standard methods where the corresponding points in X_1.and X_2 are identified first, and the displacement is estimated from matching them. To this end the output of the multiple displacement analysis after removing shadow segments is employed. Let ARN_d^l be the variable that codes the number of RN pixels for the generic segment l $(l = 1, \dots, L)$ with displacement d. The ARN_d^l can be defined as

$$ARN_d^l = \sum_{\substack{(x,y)\in l \\ l=1,\dots,L}} RN_d^l(x,y) \tag{10}$$

where $RN_d^l(x,y)$ indicates whether the pixel of coordinates (x,y) in the lth segment of the dth RN map has been labeled as RN pixel according to (6). The ARN_d^l is used to determine the displacement to be applied to the object representative point c_1^l to correct for local residual

misalignment. Let us assume that for a given segment less RN pixels are detected compared with other possible displacements when imposing displacement Ω_d. This segment will be less misaligned and thus more precisely registered when the two images are shifted by the amount of the displacement Ω_d. This concept allows to estimate the amount of displacement between X_1 and X_2 for each object representative point in that segment. In other words, the displacement $\Omega_d \in \Omega$ associated to the minimum number of misaligned pixels ARN_d^l is selected as the displacement for c_1^l in the *l*th segment that generates the best local alignment between X_1 and X_2. Such local displacement Ω^l of the *l*-th segment is computed as

$$\Omega^l = arg \min_{\Omega_d \in \Omega} \{ARN_d^l\} \tag{11}$$

Thus the proposed method identifies the local relative displacement between X_1 and X_2. It uses the object representative points in X_1 as points to apply the displacement and thus to estimate the warping function (see Sec. II.E). This is possible as the RN concept provides an estimation of displacement on those points where the geometry change is sharper and local residual displacement more critical. Once displacement is known, fine registration is performed accounting for the homogeneity information in the segments. This mechanism is intrinsically different from the one employed by standard segmentation-based matching approaches, which identify the centroids of segments in both the reference and input images and use them as CP pairs for estimating the displacement [21],[23].

E. Input image warping

Let (x_1^l, y_1^l) be the spatial coordinates of c_1^l. The spatial position (x_2^l, y_2^l) of the corresponding point c_2^l in the input image can be determined by applying the estimated displacement Ω^l as

$$\begin{cases} x_2^l = x_1^l - \Delta x_l \\ y_2^l = y_1^l - \Delta y_l \end{cases}, \quad \forall l = 1, \dots, L \tag{12}$$

where $\Omega^l = \{\Delta x_l, \Delta y_l\}$ denotes the estimated local displacements in x and y directions for the lth segment estimated in the previous step (11). Once the process is over, for each c_1^l $(l = 1, \dots, L)$ a set of object representative points pairs $\mathrm{C} = \{c_1^l, c_2^l\}$ $(l = 1, \dots, L)$ is obtained.

The set of point pairs C is employed to establish the warping function that maps the entire input image X_2 to the reference image X_1. Before that, possible critical pairs (i.e., the ones that can cause poor matching) are removed to guarantee the reliability of the transformation model. This is done by considering the spatial relation between c_1^l and c_2^l throughout an affine transformation [3]. An affine transformation is first estimated using the least squares method with all the object representative pairs in C. The pair having the largest Root Mean-Square Error (RMSE) is removed, and the transformation is estimated again on the remaining pairs. The process is repeated until RMSE of all the remaining pairs is less than a predefined threshold. Let us assume that there are M remaining object representative pairs $\mathrm{C}^m = \{c_1^m, c_2^m\}$ $(m = 1, \dots, M)$ after outlier removal step. As the object representative points are associated to segment centroids, they can be still considered as being distributed all over the entire image, even though some of them have been removed. This condition allows to apply a non-rigid transformation model [20]. We employ the piecewise linear function, which is known to be appropriate for mitigating local distortion between VHR images [15],[38]. The piecewise linear function divides X_1 into triangular regions by the Delaunay's triangulation algorithm while using object representative points as vertexes of the triangles. When the object representative points in X_1 are triangulated, the corresponding object representative points in the input image X_2 are triangulated accordingly. Then, each triangulated region in X_2 is mapped to the corresponding region of X_1 through the

affine transformation. Since the non-rigid model considers each local triangular region independently, local distortions are mitigated. Regions associated with segments involved in the removed object representative points are warped by the piecewise linear functions constructed from object representative points located near them. Let $M_{PL}(.)$ be the piecewise linear function constructed from the C^m $(m = 1, \dots, M)$ pairs. The warped input image X_2^R is computed as

$$X_2^R = M_{PL}(X_2) \tag{13}$$

The estimated warping function is able to chase the geometric properties of the scene and thus at a given extent also the differences of relief displacements over buildings generated by the off-nadir viewing angles. Since segments are smaller and more numerous where geometric changes are sharper, the warping function will be sharper as well. On the contrary, it will be smoother in regions where segments are larger.

III. DATASET DESCRIPTION AND DESIGN OF EXPERIMENTS

A. Dataset description

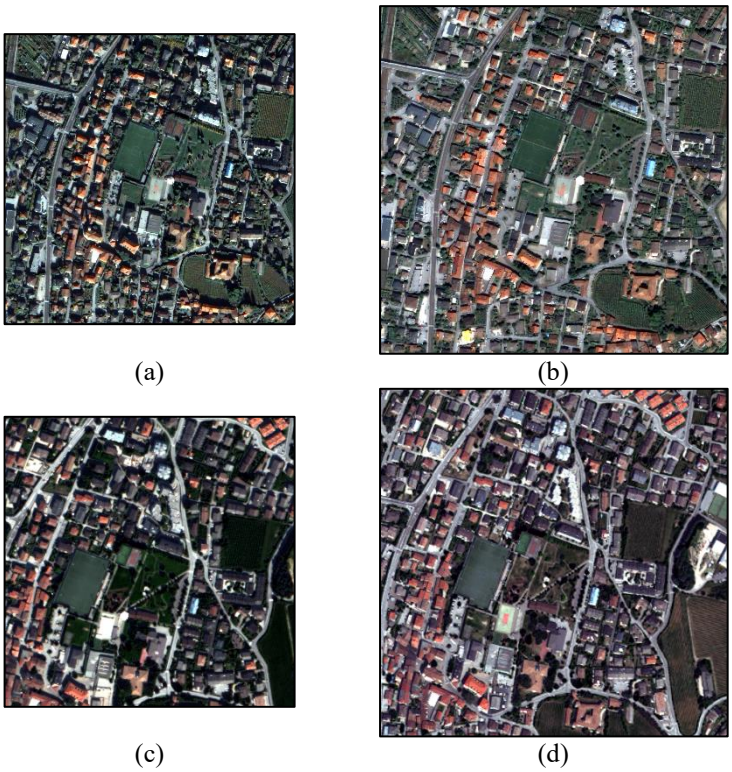
To evaluate the effectiveness of the proposed fine-registration approach, we employed multitemporal images acquired over the city of Trento (Italy). QuickBird and WorldView satellite multispectral full scenes were used to construct the datasets. The QuickBird images consist of a panchromatic band having 0.6m spatial resolution and four multispectral bands [blue (450-520nm), green (520-600nm), red (630-690nm), and near-infrared (NIR) (760-900nm)] having 2.4m spatial resolution. The multitemporal images were acquired in October 2005 (X_1) and July 2006 (X_2). The WorldView images consist of a panchromatic band having 0.5m spatial resolution and eight multispectral bands [coastal (400-450nm), blue (450-510nm), green (510-580nm), yellow (585-625nm), red (630-690nm), red edge (705-745nm), NIR 1 (770-895nm),

and NIR 2 (860-1040nm)] having 2.0m spatial resolution. The multitemporal images were acquired in August 2010 (X_1) and May 2011 (X_2). Both QuickBird and WorldView data cover an area including the city of Trento and its surroundings and show significant differences in terms of shadows and objects view angles because of the differences in the acquisition seasons and off-nadir angles (9° and 14°, and 18° and 12.9°, respectively). The proposed approach was first applied to a simulated dataset to examine its properties and effectiveness in a controlled environment. Then, it was applied to the real multitemporal datasets to demonstrate its practical application performance.

The simulated data set includes the QuickBird image taken in October 2005 (X_1). A sub-set of 1000×1000 pixels was considered that includes urban area only (Figure 5.a). The input image (X_2) is constructed from the reference image by including a deliberate nonlinear distortion. Several simulated datasets were generated by creating multiple X_2 images having different deformations both in vertical and horizontal directions with a sinusoidal transform. The distorted input images are resampled to the same size of the reference one by a bilinear interpolation. The experimental analysis for the different distortions showed similar results. Here we report the results obtained with a distortion of sinusoidal deformation in negative horizontal direction with 4-pixel amplitude and 150-degree period, and in positive vertical direction with 3-pixel amplitude and 200-degree period, only.

The real datasets were generated from the QuickBird and WorldView full scenes by defining two sub-scenes: i) One that covers a pure urban area, and ii) one that includes a sub-urban region showing both buildings and orchards. The pure urban QuickBird pair has the same reference image (X_1) as the simulated one, whereas the input image (X_2) is taken from the July 2006 acquisition (Figure 5.b). The size of the input image is 1200×1200 pixels, and it totally covers

the area of the reference one. The sub-urban sub-scene (Figure 6) is taken from the same pair of full scenes, but in a spatially disjoint region with respect to the urban one. The size of the reference (Figure 6.a) and input images (Figure 6.b) is 1000×1000 and 1200×1200 pixels, respectively. The WorldView sub-scenes represent the same urban and sub-urban areas covered by the QuickBird ones in order to validate the effectiveness of the proposed method in similar conditions, but with different sensors. The reference image (X_1) of the WorldView urban sub-scene is a 1000×1000 pixels subset of the August 2010 image (Figure 5.c), whereas the input image (X_2) is a 1200×1200 pixels subset of the May 2011 one and covers the entire reference image (Figure 5.d). The sub-urban reference image is 1200×1200 pixels (Figure 6.c), and the input one is 1400×1400 pixels (Figure 6.d).



(a) (b) (c) (d)

Figure 5. Pure urban sub-scene, city of Trento, Italy. QuickBird pair: (a) Reference image (October 2005), (b) Input image (July 2006). WorldView pair: (c) Reference image (August 2010), (d) Input image (May 2011).

In the preprocessing, all images were pansharpened by the Gram-Schmidt method [39]. In the simulated dataset, neither registration nor radiometric corrections were applied since the reference and input images were generated from the same QuickBird scene. Images in real datasets should be both registered and radiometrically corrected before applying the proposed approach instead. For minimizing the radiometric differences, a rough radiometric correction was applied by subtracting the mean value to each spectral band [27].

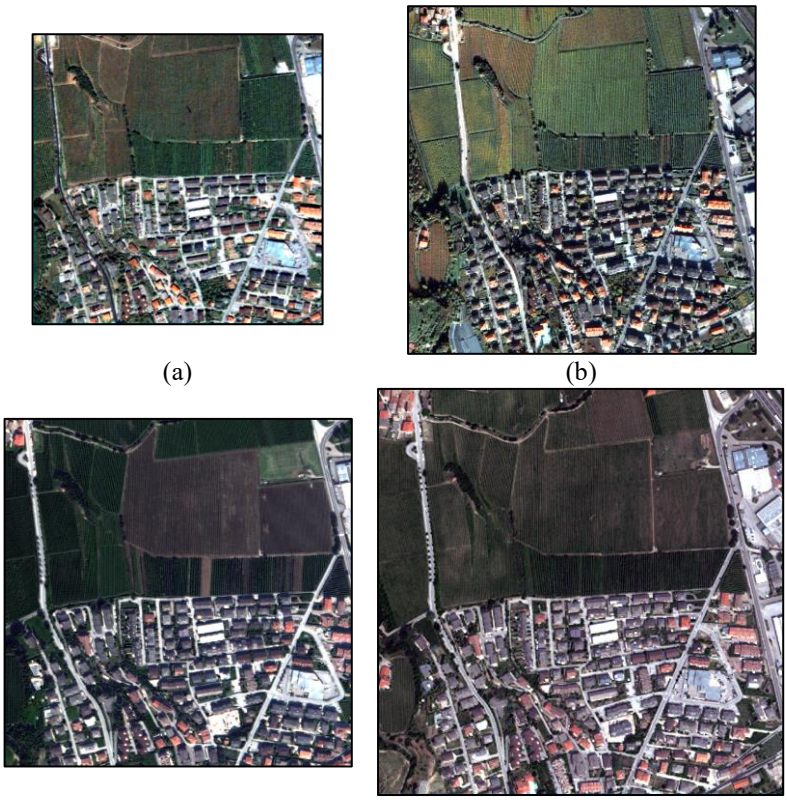


Figure 6. Sub-urban sub-scene, south of the city of Trento, Italy. QuickBird pair: (a) Reference image (October 2005), (b) Input image (July 2006). WorldView pair: (c) Reference image (August 2010), (d) Input image (May 2011).

B. Performance evaluation indexes

As discussed, distributed object representative points are important to construct the reliable non-rigid transformation model [15]–[17]. For evaluating object representative points (or CPs)

quality, we use a Distribution Quality (DQ) index [41]. It assesses the distribution quality of points by generating triangulation. DQ considers the area and shape of the triangles formed by the object representative points. The area descriptor D_A and shape descriptor D_S can be defined as

$$D_A = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{A_i}{\bar{A}} - 1\right)^2}{n-1}}, \qquad \bar{A} = \frac{\sum_{i=1}^{n} A_i}{n} \tag{14}$$

$$D_S = \sqrt{\frac{\sum_{i=1}^{n}(S_i - 1)^2}{n-1}}, \qquad S_i = \frac{3\mathrm{Max}(J_i)}{\pi} \tag{15}$$

where n denotes the number of triangles, A_i denotes the area of the triangle i, and $\mathrm{Max}(J_i)$ denotes the radian value of the largest internal angle of the triangle i. The DQ is defined as

$$DQ = D_A D_S = \frac{\sqrt{\sum_{i=1}^{n}\left(\frac{A_i}{\bar{A}} - 1\right)^2 \sum_{i=1}^{n}(S_i - 1)^2}}{n-1} \tag{16}$$

The smaller the value, the better the distribution of object representative points (or CPs).

Concerning the simulated dataset, where the reference and the input images are generated from the same QuickBird acquisition, the two images are expected to be identical when small simulated distortions are corrected. Therefore, the correlation coefficient (ρ) can be employed as the representative similarity measure to evaluate the registration accuracy. The correlation coefficient ρ between images X_1 and X_2 is calculated as:

$$\rho(X_1, X_2) = \frac{\sigma_{X_1 X_2}}{\sqrt{\sigma_{X_1}\sigma_{X_2}}} \tag{17}$$

where $\sigma_{X_1 X_2}$ denotes the covariance between the two images, and σ_{X_1} and σ_{X_2} denote the standard deviations of the two images. ρ has range from -1 to 1, with 1(-1) indicating perfect positive (negative) correlation.

The assertion above does not hold for real datasets where multitemporal images obtained at different dates show different radiometric properties. Therefore, the similarity-based indexes like correlation coefficient cannot be used for evaluating the registration performance. For the quantitative evaluation of performance on the real datasets, the RMSE and its standard deviation (STD) are calculated over checkpoints extracted by experienced image interpreters. Let $(\Delta x_i^c, \Delta y_i^c)$ be the residual difference in *x*- and *y*- directions on a checkpoint pair, the RMSE and its STD are estimated as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}((\Delta x_i^c)^2 + (\Delta y_i^c)^2)}{M}} \tag{18}$$

and

$$STD = \sqrt{\frac{\sum_{i=1}^{M}\left(\sqrt{(\Delta x_i^c)^2 + (\Delta y_i^c)^2} - RMSE\right)^2}{M-1}} \tag{19}$$

where *M* is the number of checkpoints.

C. Experimental setup

First of all the simulated dataset was employed in the experimental analysis to test the sensitivity of the proposed method to the segmentation step when varying the segmentation parameters. After that, the effectiveness of the proposed approach was assessed by comparing its registration performance with the ones computed:

i) Before applying registration.

ii) After applying state of the art (SoA) manual registration. To this end, 15 CP pairs were manually selected by photointerpretation for each dataset, and used for estimating the affine transformation for warping the input image to the reference one.

iii) After applying the SoA automatic SIFT-based registration [7]. In this case, the affine transformation was estimated by applying the RANdom SAmple Consensus (RANSAC) on CP pairs detected by the SIFT method [40]. The input image was warped to the coordinates of the reference one according to the estimated transformation.

iv) After applying a fine-registration approach to automatically pre-registered image pairs based on regular blocks [31]. This experiment aims at assessing the effectiveness of using the object adaptive spatial information carried out by segments. This is achieved by using regular blocks (instead of segments) that do not account for the spatial correlation and spectral homogeneity of objects in the images. One hand, blocks may include more than one object leading to poor estimation of the warping function. On the other, one object may divide over more blocks and different parts of the same object may be associated to displacements with different direction and intensity values. This may lead to a large local distortion of the object. Since blocks are not object representative, the matching is estimated on RN pixels which are usually clustered together on object boundaries. This induces geometric distortions when they are used to construct a transformation model for warping. Moreover, blocks include shadows which affects the registration process (see Sec. II.C).

In addition, for the urban sub-scene an analysis was conducted on the impact of performing or non-performing the shadow removal step when the proposed segment-based fine-registration approach is employed.

IV. EXPERIMENTAL RESULTS

The proposed method requires fixing some parameters. Their values have been selected based on

empirical experiments and our previous work [31]. In detail, CVA for RN identification was conducted on the red and NIR bands of the VHR images [26]. For the multiscale decomposition necessary for estimating the RN distribution, three levels (N=3) were computed by a Daubechies-4 non-decimated Stationary Wavelet Transform (SWT) [42]. The threshold T for the magnitude variable was automatically selected by applying the Bayesian decision rule for minimum error according to [43] and T_{RN} was set to 10^{-4}. The set of displacement values Ω_d to estimate residual misalignment was determined by translating the input image in both x and y directions from -5 to +5 pixels with 0.5-pixel interval. The distorted input images are resampled to the same size of the reference one by a bilinear interpolation.

A. Results: Simulated dataset

SLIC segmentation technique was applied to the reference image X_1 [32]. There are two main parameters to be selected for the segmentation: L (number of initial centroids) and m (compactness parameter). The parameter L is related to the size of segments directly. The parameter m describes the weight of spatial proximity versus that of color similarity when generating objects. In order to assess the impact of the segmentation step on the proposed fine-registration approach, we carried out experiments with various values of L and m parameters. Note that the average size of buildings in Trento is around $25\times20m^2$ [44]. Thus the optimal number of segments is expected to be around 700 (one for each building in the considered scene). Figure 7 provides the behavior of estimated correlation coefficient values by varying L and m parameters, and thus the number of segments. As one can see, the less the number of segments the worse the registration performance. This is because more than one object fits into a single segment. When the segment number becomes larger and thus closer or greater than the average building size, the correlation coefficient value becomes higher. Concerning the value of the

compactness parameter *m*, higher values generally showed slightly better and stable results, meaning that segments having compact shape work better. However, as the size of segments becomes small enough, the registration performance showed to be less dependent on the compactness. According to these experimental results, the value of *L* and *m* was set to 800 and 40, respectively, and all the experiments in the following were carried out with these values. This leads to 807 segments for the simulated dataset. It is worth to recall that the number of initial centroids and final segments might be different due to the SLIC procedure that merges unassigned isolated pixels to the nearest cluster and creates a new cluster if a large number of adjacent pixels remain unassigned. Figure 8 presents an example of the object representative points (yellow crosses) and boundaries of segments. As one can see, object representative points are located inside the segments, as segments tend to have a compact shape. They are also well-distributed and do not cluster to each other.

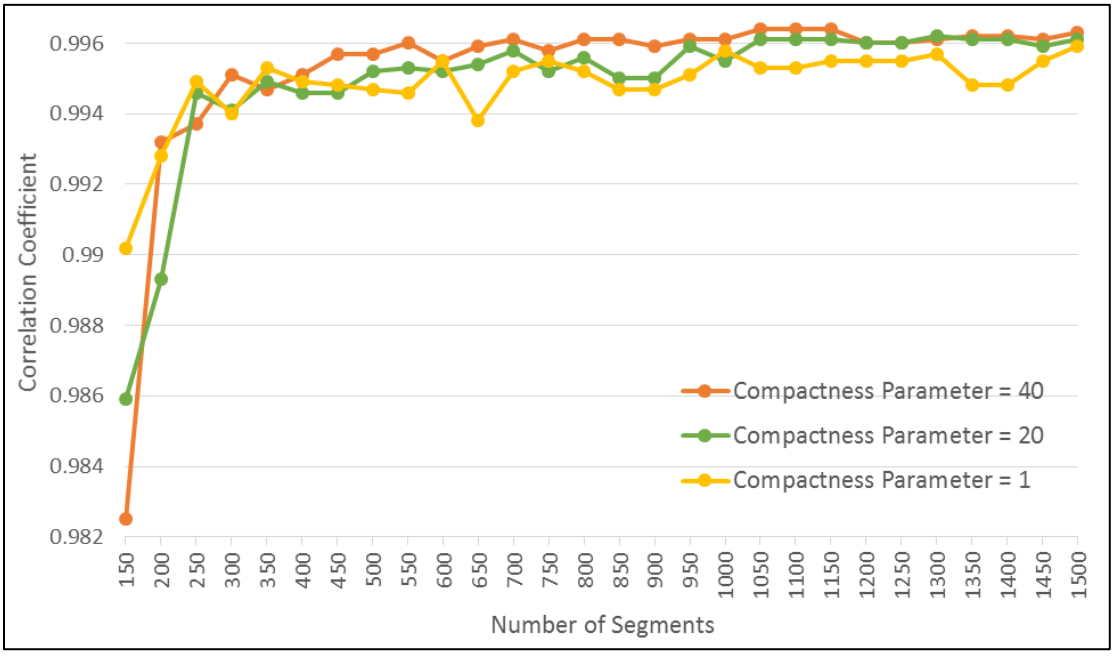


Figure 7. Correlation coefficient behaviors by varying segmentation parameters (simulated dataset).

Figure 8. Segmentation result (white boundary) and object representative points (yellow crosses) (simulated dataset).

Once fine registration is complete, visual assessment can be performed by observing the chessboard image where the blocks of the reference and the warped input images are repeatedly interlaced (Figure 9.b). For the visual contrast on the chessboard image, the blocks of the reference image are displayed as true-color composition, whereas the blocks of the warped input image are displayed as false-color composition (i.e., NIR, red, and green bands of the warped input image were assigned respectively to the red, green, and blue channels). The chessboard image of the simulated dataset before registration was also generated for visual comparison (Figure 9.a). From the quality of alignment on the boundaries among adjacent blocks, we can appreciate better the level of registration accuracy. White circles in Figure 9 highlight where the quality of the registration performance can be better observed. As one can see, lines and shapes were precisely aligned in the chessboard image generated by the proposed method (see Figure 9.b), whereas they were not aligned properly in the original dataset (see Figure 9.a).

The impact of the distribution of object representative points can be judged by DQ values (Table I). The block-based fine registration achieved a DQ of 2.29, which is even worse than the value

achieved by automatic detection approach. This is because RN pixels tend to cluster along boundaries of objects. The object representative points detected by the proposed method sharply reduced the DQ value to 0.27, which is even better than the one achieved by the manually selected CPs (i.e., 0.56). This confirms that better registration performance can be obtained by applying the non-rigid transformation (i.e., piecewise linear function in this paper). In terms of registration performance (Table I), the automatic SoA method was unable to increase the ρ value compared with the original non registered simulated data and the manual approach slightly improved it only, whereas both the fine-registration methods could considerably increase it. The proposed approach showed the highest ρ value as 0.997 (which is close to one) meaning that local distortions introduced in the simulated image have been corrected. From the visual and numerical evaluations, the two VHR images were precisely registered by the proposed approach under the condition that only local geometric distortions affect the data. This shows that the proposed approach is able to improve the local alignment even if registration is good over the entire scene.

(a) (b)

Figure 9. Chessboard images generated (a) before applying the proposed technique and (b) after applying the proposed technique (simulated dataset).

TABLE I. DQ AND ρ QUALITY INDEXES (SIMULATED DATASET).

Registration method	DQ	ρ
No registration	-	0.810
SoA manual	0.56	0.838
SoA automatic	1.12	0.810
Block-based fine registration	2.29	0.994
Proposed approach	0.27	0.997

B. Results: Real datasets

To further validate the proposed approach, experiments were performed on the four pairs of real VHR multitemporal images. After radiometric and geometric preprocessing, the proposed method was applied. The parameters are the same as the ones established for the simulated dataset. The threshold T_S for the shadow map generation was set to 200 and 215 for QuickBird and WorldView pairs, respectively. The chessboard images generated from the reference and warped input images are shown in Figure 10 (urban sub-scene) and Figure 11 (sub-urban sub-scene), respectively. By observing the boundaries of adjacent blocks, one can see that the two images look precisely registered.

For the quantitative assessment, the RMSE and its STD have been calculated on 20 manually checkpoints extracted from each image (it is worth nothing that these CPs are independent from the ones employed for the manual registration). The DQ value is also computed to evaluate the distribution quality of object representative points (and CPs). Table II and Table III show the registration performance calculated over checkpoints for urban and sub-urban sub-scenes, respectively.

On the urban sub-scene, SoA manual and automatic registration resulted in RMSE values of 2.76 and 3.70, and 1.52 and 2.25 pixels for the QuickBird and the WoldView datasets, respectively. Both fine-registration approaches could improve this performance. The registration accuracies of the block-based approach provided 2.08 and 1.36 RMSE values for the QuickBird and the WorldView datasets, respectively. Whereas, the proposed method resulted in subpixel RMSE

values, i.e., 0.70 and 0.89, 0.75 and 0.84 pixels with manual and automatic pre-registration for QuickBird and WorldView datasets, respectively. The results point out the effectiveness of the proposed approach, which allows to accurately register the images obtaining the highest accuracy both when manual and automatic pre-registration is applied. In the urban sub-scene, a high occurrence of shadows due to buildings can be observed. Thus, it was used to evaluate the impact of the shadow removal step on the proposed fine-registration performance. This is achieved by performing segment-based fine-registration without applying shadow removal. As one can see in Table II, registration becomes less accurate. RMSE increased from 0.89 to 1.40, and from 0.84 to 0.94 pixels for the QuickBird and WorldView datasets, respectively. Figure 12 shows a detail of the chessboard images obtained for the QuickBird urban sub-scene by neglecting the shadow removal step (Figure 12.a) and by considering it (Figure 12.b). In the case when neglecting the shadow removal step, it is possible to observe that the two buildings in the upper part of the images show some distortions along the top borders close to the shadow area and are aligned with slightly less accuracy.

Similar conclusions hold for the sub-urban sub-scene. The proposed method achieved an RMSE of 0.82 and 1.15, and of 0.80 and 0.85 pixels with manual and automatic pre-registration for the QuickBird and WorldView pairs, respectively. The RMSE reduction when compared to the results of the block-based fine registration is of 0.83 and 0.50, and 0.30 and 0.25 pixels, respectively. The improvement on sub-urban sub-scene is slightly less with respect to the one of the pure urban sub-scene. This is due to the fact that object representative points detected in orchard part are less representative than those in the urban one. In order to independently check the impact of the proposed fine approach on building and orchard areas of the sub-urban sub-scene, we estimated RMSE values from the CPs extracted on building and orchard areas,

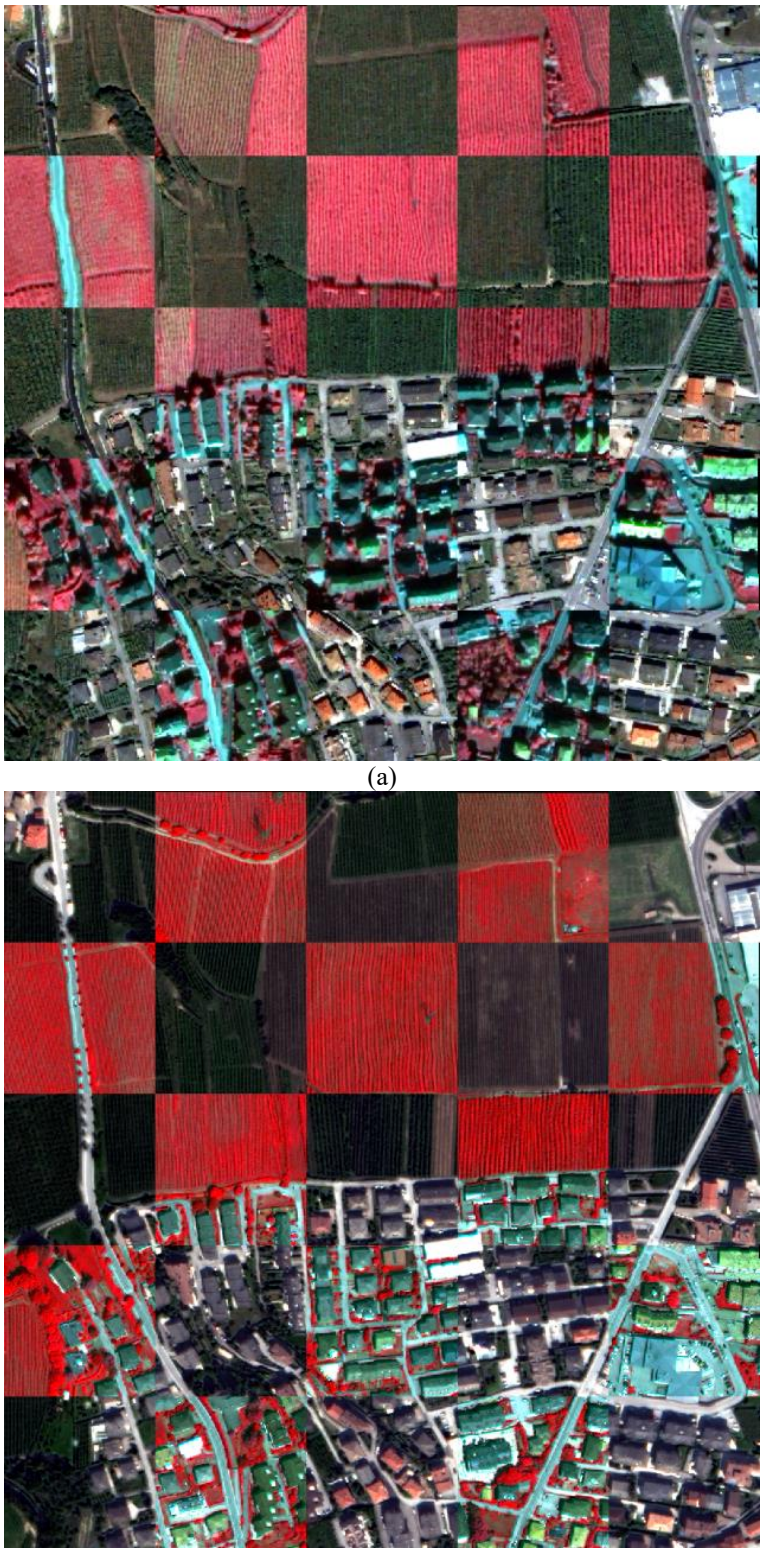
respectively. For the QuickBird images, the proposed method significantly improved the registration in the part of the sub-scene with buildings by reducing the RMSE from 3.94 and 1.72 for the SoA automatic and block-based fine-registration approaches, to 1.5 for the proposed one. It achieved performance slightly higher than the block-based fine registration approach in the orchard part, instead. The RMSE was reduced from 1.71 for the SoA automatic approach to 0.80 and 0.99 for the block-based fine-registration approach and for the proposed one, respectively. Similar results were shown in the WorldView sub-scene. In the part of the sub-scene with buildings, the RMSE was reduced from 3.10 and 1.10 for the SoA automatic and block-based fine-registration approaches, to 0.86 for the proposed approach. Whereas, in the orchard part, the RMSEs for the SoA automatic and block-based fine-registration approaches were reduced from 1.97 and 0.74, respectively, to 0.68 for the proposed one.

In addition, it is worth nothing that for both sub-scenes, the proposed approach achieved a more uniform accuracy over the entire scene. This is confirmed by the smaller STD values (Table II and Table III).

(a)

(b)

Figure 10. Chessboard images generated with multitemporal datasets after applying the proposed fine-registration approach: (a) QuickBird and (b) WorldView images (urban sub-scene).



(a) (b)

Figure 11. Chessboard images generated with multitemporal datasets after applying the proposed fine-registration approach: (a) QuickBird and (b) WorldView images (sub-urban sub-scene).

TABLE II. REGISTRATION RESULTS (URBAN SUB-SCENE).

Registration method		QuickBird dataset			WorldView dataset		
		DQ	RMSE (pixels)	STD (pixels)	DQ	RMSE (pixels)	STD (pixels)
No registration		-	16.87	2.13	-	21.95	8.24
SoA	manual	0.75	2.76	1.61	1.01	1.52	0.61
	automatic	2.00	3.70	2.21	1.54	2.25	1.51
Block-based fine registration		2.08	1.52	0.60	2.13	1.36	0.69
Proposed segmentation-based fine registration	SoA manual pre-registration	0.41	0.70	0.32	0.36	0.75	0.38
	SoA automatic pre-registration	0.40	0.89	0.37	0.35	0.84	0.32
	SoA automatic pre-registration without shadow removal	0.28	1.40	0.71	0.27	0.94	0.48

(a) (b)

Figure 12. A detail of the chessboard images generated with multitemporal urban sub-scene after applying the proposed fine-registration approach: (a) without the shadow removal step; and (b) with the shadow removal step (QuickBird).

TABLE III. REGISTRATION RESULTS (SUB-URBAN SUB-SCENE).

Registration method		QuickBird dataset			WorldView dataset		
		DQ	RMSE (pixels)	STD (pixels)	DQ	RMSE (pixels)	STD (pixels)
No registration		-	25.51	1.66	-	37.14	7.81
SoA	manual	1.20	2.14	1.02	1.20	1.78	0.85
	automatic	0.81	3.77	1.51	1.75	3.01	1.82
Block-based fine registration		2.17	1.65	0.82	2.32	1.10	0.49
Proposed segmentation-based fine registration	SoA manual pre-registration	0.19	0.82	0.28	0.31	0.80	0.34
	SoA automatic pre-registration	0.19	1.15	0.53	0.34	0.85	0.45

V. CONCLUSION

In this paper we proposed an approach to fine registration of multitemporal images that aims at

correcting local residual misalignments after standard registration. The method defines a set of object representative points that accounts for the spatial correlation and the spectral homogeneity of pixels in the reference image. This choice results in distributed object representative points and allows for an accurate chasing of the local behaviors of residual misalignment. The displacement of object representative points is established by a multiple displacement analysis of residual local misalignment. The quality of object representative points is improved by employing shadow information. Both qualitative and quantitative experimental results demonstrated that the proposed approach is able to improve the registration accuracy with respect to the literature registration approaches if residual misalignment exists. Despite the approach is general and can be applied to any kind of multitemporal images, it becomes particularly suitable for those scenes that show a large number of objects (and thus borders) where standard methods tend to generate larger local distortions. This is the case of VHR multitemporal images acquired over urban areas, but also of medium resolution images that contain sharp edges. On the opposite, scenes with a small number of objects (and thus borders) suffer less of local misalignments after registration and thus the registration improvement by the proposed method is smaller. In our experiments, both kinds of scenes have been considered and the proposed method guaranteed a registration accuracy smaller than one pixel with a RMSE standard deviation smaller than those obtained by state-of-the-art methods. This confirms that the registration performance is better over the entire image at both average and local level.

It is worth noting that the proposed approach may become less effective when scenes show very tall elements captured with large off-nadir angle. In such situations, as it happens for standard registration approaches, additional ancillary data (i.e., precise DSM) are required to improve the results.

As a future work, we plan to design an approach to mitigate the impact of heterogeneous segments and to improve the robustness of the proposed method for the use with multisensor images.

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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “SURF: Speeded up robust features,” *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp.346–359, Jun. 2008.

[9] S. Mallat and W. L. Hwang, “Singularity detection and processing with wavelets,” *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp.617–643, Mar. 1992.

[10] I. Zavorin and J. Le Moigne, “Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery,” *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 770–782, 2005.

[11] J. M. Murphy, J. L. Moigne, and D. J. Harding, “Automatic image registration of multimodal remotely sensed data with global shearlet features,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1685–1704, Mar. 2016.

[12] C. Harris and M. Stephens, “A combined corner and edge detector,” *Proc. Alvey Vis. Conf.*, pp.147–152, Sep. 1988.

[13] 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.

[14] Y. Han, F. Bovolo, and L. Bruzzone, “Fine co-registration of VHR images for multitemporal urban area analysis,” *in Int. Workshop Anal. Multi-Temporal Remote Sens. Images, (Multitemp)*, July 22–24, 2015.

[15] 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.

[16] 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.

[17] 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.

[18] J. Le Moigne, W. J. Campbell, and R. F. Cromp, “An automated parallel image registration technique based on the correlation of wavelet features,” *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1849–1864, Aug. 2002.

[19] 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.

[20] 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.

[21] P. Dare and I. Dowman, “An improved model for automatic feature-based registration of SAR and SPOT images,” *ISPRS J. Photogramm. Remote Sens.*, vol. 56, no. 1, pp. 13–28, Jun. 2001.

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

[23] H. Gonçalves, J. 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.

[24] G. Troglio, J. Le Moigne, J. A. Benediktsson, G. Moser, and S. B. Serpico, “Automatic extraction of ellipsoidal features for planetary image registration,” *IEEE Geosci. Remote*

Sens. Lett., vol. 9, no. 1, pp. 95–99, Jan. 2012.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] Y. Han, F. Bovolo, L. Bruzzone, "An approach to fine coregistration between very-high-resolution multispectral images based on registration noise distribution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6650–6662, Dec. 2015.

[32] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*,

vol. 34, no. 11, pp. 2274–2281, Nov. 2012.

[33] T. Mei, L. An and Q. Li, “Supervised segmentation of remote sensing image using reference descriptor,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 938–942, May 2015.

[34] G. H. Joblove and D. Greenberg, “Color spaces for computer graphics,” *in Proc. ACM SIGGRAPH*, Atlanta, Georgia, 1978, pp. 20–25.

[35] 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.

[36] L. Bruzzone and F. Bovolo, “A novel framework for the design of change-detection systems for very-high-resolution remote sensing images,” *Proc. IEEE*, vol. 101, no. 3, pp. 609–630, Mar. 2013.

[37] V. J. D. Tsai, “A comparative study on shadow compensation of color aerial images in invariant color models,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp.1661–1671, Jun. 2006.

[38] 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.

[39] 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.

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

[41] Q. Zhu, B. Wu, and Z. Xu, “Seed point selection method for triangle constrained image matching propagation,” *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 2, pp. 207–211, Apr., 2006.

[42] S. G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, no. 7, pp. 674–693, Jul. 1989.

[43] 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.

[44] C. Marin, F. Bovolo, and L. Bruzzone, “Building change detection in multitemporal very high resolution SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2664–2682, May 2015.

Youkyung Han (S’12–M’15) is an assistant professor of the School of Convergence & Fusion System Engineering at Kyungpook National University (KNU), Korea. Before joining KNU, he held a postdoc position with the Remote Sensing for Digital Earth unit at Fondazione Bruno Kessler (FBK), Italy (2014–2015), and the Lyles School of Civil Engineering at Purdue University, USA (2016). He received his B.S. degree in civil, urban and geo-system engineering from Seoul National University, Korea, in 2007, and his M.S. and Ph.D. degrees in civil and environmental engineering from Seoul National University, Korea, in 2009

and 2013, respectively. His major research interests include image processing of very high resolution remote sensing data such as image-to-image registration, segmentation, object extraction, and change detection. He recently has further interests on the analysis of multispectral/hyperspectral and multi-temporal images.