Automatic Image Registration Through Image Segmentation and SIFT

Hernâni Gonçalves, Luís Corte-Real, *Member, IEEE*, and José A. Gonçalves

***Abstract*—Automatic image registration (AIR) is still a present challenge for the remote sensing community. Although a wide variety of AIR methods have been proposed in the last few years, there are several drawbacks which avoid their common use in practice. The recently proposed scale invariant feature transform (SIFT) approach has already revealed to be a powerful tool for the obtention of tie points in general image processing tasks, but it has a limited performance when directly applied to remote sensing images. In this paper, a new AIR method is proposed, based on the combination of image segmentation and SIFT, complemented by a robust procedure of outlier removal. This combination allows for an accurate obtention of tie points for a pair of remote sensing images, being a powerful scheme for AIR. Both synthetic and real data have been considered in this work for the evaluation of the proposed methodology, comprising medium and high spatial resolution images, and single-band, multispectral, and hyperspectral images. A set of measures which allow for an objective evaluation of the geometric correction process quality has been used. The proposed methodology allows for a fully automatic registration of pairs of remote sensing images, leading to a subpixel accuracy for the whole considered data set. Furthermore, it is able to account for differences in spectral content, rotation, scale, translation, different viewpoint, and change in illumination.**

***Index Terms*—Automatic image registration (AIR), image segmentation, optical images, scale invariant feature transform (SIFT).**

I. Introduction

IMAGE registration is still far from being a commonly automatized process, in particular regarding remote sensing applications. Although several methods have been proposed in the last few years [2], [8], [9], [11], [31], [32], geometric correction of satellite images is, in practice, mostly a manual work. The manual procedure is associated to inter- and intraoperator subjectivities, beyond being a time-consuming task.

Geometric correction of satellite images may involve several factors which should be considered (both radiometric and geometric transformations). Despite the fact that a local translation may be the main distortion for small segments of satellite images, rotation and scale effects may also be present, as well as distortions associated to the terrain relief and panoramic view. Furthermore, significant differences on the spectral content between the images to be registered also increase the difficulty in automating the registration process.

The main concept regarding automatic registration of satellite images is to obtain an accurate set of tie points and then apply the transformation function which is most suitable to the pair of images to be registered. A considerably large number of approaches may be found in the literature regarding the automatic obtention of tie points—being mainly area- or feature-based methods—by means of the image intensity values in their close neighborhoods, the feature spatial distribution, or the feature symbolic description [32].

Image segmentation comprises a wide variety of methods [3], [26], either for monochrome or color images (or to a single or multiple bands of satellite images). Most image segmentation methods can be classified according to their nature: histogram thresholding, feature space clustering, region-based approaches, edge detection approaches, fuzzy approaches, neural networks, physics-based approaches, and any combination of these [3]. Any of these generally intends to transform any image to a binary image: objects and background. The use of image segmentation as a step in image registration had been scarcely explored [5], [12], [17]. Therefore, further improvements under the scope of methodologies for automatic image registration (AIR) may be achieved, particularly combining image segmentation with other methods.

Dare and Dowman [5] proposed an improved model for automatic feature-based image registration, based on multiple feature extraction and feature matching algorithms. With the combination of extraction and matching algorithms, it was possible to identify common features in multisensor situations, from which tie points can be derived. The tie points were initially obtained from matching the centroids, followed by the matching of the pixels of the patch edges. This approach is quite sensitive to differences on the patch delineation, which may often occur when using images from different sensors or even with temporal differences.

Local descriptor is a widely used technique in several image- and video-based tasks [22]. One of the main advantages of local descriptors is that they are distinctive, are robust to occlusion, and do not require segmentation. The concept behind it is to detect image regions covariant to a class of transformations, which are then used as support regions to compute invariant descriptors, i.e., the detectors provide the

Manuscript received April 8, 2010; revised August 26, 2010 and December 30, 2010; accepted January 22, 2011. Date of publication March 17, 2011; date of current version June 24, 2011. The work of H. Gonçalves was supported by Fundação para a Ciência e a Tecnologia, Portugal.

H. Gonçalves is with the Departamento de Geociências, Ambiente e Ordenamento do Território, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal, and also with the Centro de Investigação em Ciências Geo-Espaciais, Universidade do Porto, 4169-007 Porto, Portugal (e-mail: hernani.goncalves@fc.up.pt).

L. Corte-Real is with the Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia, Universidade do Porto, 4169-007 Porto, Portugal, and also with the Institute for Systems and Computer Engineering (INESC) Porto, 4200-465 Porto, Portugal (e-mail: lreal@inescporto.pt).

J. A. Gonçalves is with the Departamento de Geociências, Ambiente e Ordenamento do Território, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal (e-mail: jagoncal@fc.up.pt).

Digital Object Identifier 10.1109/TGRS.2011.2109389

regions which are used to compute the descriptors [22]. A comprehensive review on the comparison of affine region detectors may be found in [21]. Mikolajczyk and Schmid [22] compared the performance of descriptors computed for local interest regions of gray-value images. There are three main classes of descriptors: distribution-based descriptors, spatial-frequency techniques, and differential descriptors. They compared the descriptor performance for affine transformations, scale changes, rotation, blur, jpeg compression, and illumination changes [22]. Mikolajczyk and Schmid have found, based on their experiments, that the scale invariant feature transform (SIFT) [19]—which is a distribution-based descriptor—was among those which obtained the best results for most of the tests, with lower performance for textured scenes or when edges are not reliable.

The SIFT approach allows for the extraction of distinctive invariant features from images, which can be used to perform reliable matching between images presenting a substantial range of affine distortion, change in 3-D viewpoint, addition of noise, and change in illumination [19]. A SIFT descriptor is a 3-D histogram of gradient location and orientation, where location is quantized into a 4×4 location grid and the gradient angle is quantized into eight orientations, resulting in a descriptor of dimension 128 [22]. Despite the several advantages of using the SIFT approach, it does not produce meaningful results when directly applied to remote sensing images, as supported by the recent approaches applying SIFT to remote sensing images [4], [16], [20], [23], [28], [30].

Li *et al.* [16] proposed an adaptation on the original method proposed by Lowe [19], where the feature descriptor is refined and the use of the Euclidean distance is replaced by a joint distance. The method proposed by Li *et al.* assumes that often remotely sensed images have no local distortions, and so, geometric distortions can be modeled by "shape-preserving mapping" model [32] (translation, rotation, and scaling only). However, when an image with nadir looking is to be registered with an image with a considerable viewing angle (such as most of Satellite Pour l'Observation de la Terre (SPOT) images and other higher spatial resolution satellite images), this assumption fails. It is also not adequate for situations where the terrain relief has significant variations across the considered scene.

Mukherjee *et al.* [23] proposed a method for detection of multiscale interest points for later registration of hyperspectral imagery. They proposed spectral descriptors for hyperspectral interest points that characterize each interest point based on the spectral information and its location and scale. Their method mainly differs from the Lowe's keypoint detection algorithm in the sense that principal component analysis (PCA) is applied to the hyperspectral imagery, and nonlinear function for combining difference of Gaussian (DoG) responses along spectral dimension is applied prior to local extrema detection [23]. They considered in their experiments four time-lapse images acquired by the Hyperion sensor, retaining a subset of the available spectral bands (such as uncalibrated or saturated data channels) and using random regions with around 200 scan lines. Nevertheless, this methodology is not appropriate for multi- or single-band images, which is still presently the main imagery source in remote sensing applications.

Sirmaçek and Ünsalan [28] have recently used SIFT keypoints and graph theory applied to IKONOS images, under the scope of urban-area and building detection. They state that the standard SIFT implementation is not sufficient for urban-area and building detection from satellite images alone, since the presence of many similar and nearby buildings in the satellite images is a quite frequent problem [28]. Moreover, as mentioned by the authors, their building-detection method may not detect buildings if the contrast between their rooftop and the background is low. Although the work by Sirmaçek and Ünsalan is about urban-area and building detection, and not AIR, this is another remote sensing application where the simple application of SIFT to remote sensing images is not sufficient.

Cheng *et al.* [4] presented a method of robust affine invariant feature extraction for image matching called ED-MSER, which combines MSER (maximally stable extremal region), SIFT, and a filtering strategy. The core of the approach relies on a hierarchical filtering strategy for affine invariant feature detection based on entropy and spatial dispersion quality constraints. It consists in evaluating the information entropy and spatial dispersion quality of all features detected by MSER, removing the features with low information entropy and bad distribution, and just selecting the features with high information entropy and good distribution. Finally, SIFT is used as ED-MSER descriptor, as SIFT has been demonstrated to be superior to others in resisting common image deformations [4]. Although the presented experiments demonstrated that ED-MSER can get much higher repeatability and matching score compared to the standard MSER and other algorithms, its ability for multisensor image registration tasks will be limited.

In this paper, a robust and efficient method for AIR is proposed, which combines image segmentation and SIFT, complemented by an outlier removal stage. The reference and unregistered images may differ in translation, rotation, and scale and may present distortions associated to the terrain relief and significantly different spectral content. The methodology is described in Section II, the results of its application to four different pairs of images are illustrated in Section III, and the discussion is presented in Section IV.

II. Methodology

Let us consider $(X_{\mathrm{REF}}, Y_{\mathrm{REF}})$ as the coordinates of a point from the reference image and $(X_{\mathrm{NEW}}, Y_{\mathrm{NEW}})$ as (pixel, line) of the corresponding point in the new image to be registered. The relation between $(X_{\mathrm{REF}}, Y_{\mathrm{REF}})$ and $(X_{\mathrm{NEW}}, Y_{\mathrm{NEW}})$ may be written as

$$\begin{cases} X_{\mathrm{REF}} = f(X_{\mathrm{NEW}}, Y_{\mathrm{NEW}}) \\ Y_{\mathrm{REF}} = g(X_{\mathrm{NEW}}, Y_{\mathrm{NEW}}) \end{cases} \tag{1}$$

where f and g are the functions which better describe the relation between the coordinates of the two images. The type of function may depend on several factors, such as the sensor acquisition model and terrain distortion, among others. In the presence of a set of N conjugate points, the previous equations may be solved for the function coefficients, through

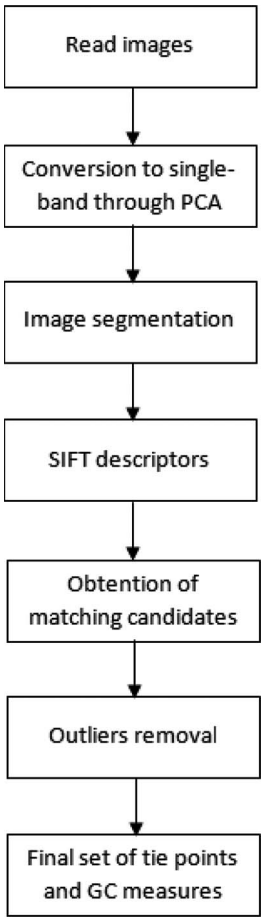


Fig. 1. Main steps of the proposed methodology for AIR.

the most appropriate method in each case (usually the least square method). The main difficulty relies on an automatic and accurate identification of the N conjugate points, which is a particular challenge in several remote sensing applications. The main steps of the proposed methodology for AIR are shown in Fig. 1 and include the following: conversion to single-band image, image segmentation, SIFT descriptors, obtention of a set of matching candidates, outlier removal, and final set of tie points and corresponding geometric correction measures. These steps will be separately described in the following.

A. Conversion to Single Band Through PCA

In the case that one or both images to be registered have more than one spectral band, a data reduction method should be individually applied to each image to facilitate the later segmentation stage. This step should account for minimizing the loss of important information for the later stages of the methodology, which is a consequence of applying data reduction methods. It should allow for considering one single band, which explains the majority of the image variability. The method of PCA was first derived and published by Hotelling [13]. Principal components are still a basic tool for image description used in numerous applications, allowing, in general, for the reduction of the data dimension. Under the scope of image processing, PCA allows for the reduction of the number of bands of an image through a linear combination of them. Additionally, it allows for considering one single band, which explains the majority of the image variability.

Let I be an image of size $m \times n \times k$ pixels, where k is the number of spectral bands. This image stack I is rearranged as a new image I_{vector} of size $mn \times k$ pixels, where each column of I_{vector} corresponds to the stacked pixels of the corresponding band of I.

Let us consider $v = (v_1, v_2, \ldots, v_k)^{\text{T}}$, which is a k-element vector formed by the values of a particular pixel position across the k spectral bands, i.e., each line of I_{vector}. Defining $m_v = E\{v\}$ as the expected value of v and $C_v = E\{(v - m_v)(v - m_v)^{\text{T}}\}$ as the respective covariance matrix, let A be a matrix whose rows are formed from the eigenvectors of C_v in such an order that the first row of A is the eigenvector corresponding to the largest eigenvalue and so on. The Hotelling transform

$$y_v = A(v - m_v) \tag{2}$$

is what is also known as the principal component transform. Through this transform, it is possible to concentrate on a single band (the first principal component hereby assigned as J) most of the variability explained by the original k bands. This is a widely known method of data reduction, and further details can be found in [6].

B. Image Segmentation

The segmentation of an image allows for its "simplification," since it significantly reduces the number of different pixel values. Although it is also associated with a loss of information on the image content, the decision of using an original or segmented image will depend on the context of the AIR method. Image segmentation is a process of partitioning an image into nonintersecting regions such that each region is homogeneous and the union of two adjacent regions is not homogeneous [26]. Let $P()$ be a homogeneity predicate defined on groups of connected pixels and J the first principal component (of size $m \times n$ pixels) of I, obtained as described in the previous section. Segmentation is a partitioning of image J into a set of l connected regions such that

$$\bigcup_{i=1}^{l} S_i = J \text{ with } S_i \cap S_j = \oslash, \qquad i \neq j \tag{3}$$

and the uniformity predicate $P(S_i) =$ true for all regions S_i and $P(S_i \cup S_j) =$ false when S_i is adjacent to S_j.

A large number of segmentation methods can be found in the literature, but there is no single method which can be considered good for all images, nor are all methods equally good for a particular type of image [26]. The existing image segmentation methods include gray-level thresholding, iterative pixel classification, surface-based segmentation, edge detection, and methods based on fuzzy set theory [26]. Thresholding-based methods can be classified according to global or local thresholding and also as either bilevel thresholding or multithresholding. For the aforementioned facts, we decided to

consider the nonparametric and unsupervised Otsu's thresholding method [25].

The Otsu's thresholding method may be recommended as the simplest and standard method for automatic threshold selection, which can be applied to various practical problems [25]. Although the Otsu's thresholding method is usually applied to images with a bimodal histogram, it may also provide a meaningful result for unimodal or multimodal histograms where a precise delineation of the objects present on the scene is not a requirement. Some examples are illustrated in [25], where the histogram shape is nearly unimodal and a meaningful segmentation is obtained. The key concept behind this method is to obtain an optimal threshold that maximizes a function of the threshold level. The optimal threshold is selected by a discriminant criterion, in order to maximize the separability of the resultant classes in gray levels. The procedure utilizes only the zeroth- and the first-order cumulative moments of the gray-level histogram. Further details may be found in [25].

According to this bilevel thresholding, the image J pixels are assigned as 0 or 1. Then, the connected components in the binary image are identified and assigned a number, and objects with size less than 0.1% of the image size are removed in order to reduce the computation effort, without compromising the method performance. The labeled image with the small regions removed is then stretched to a 16-b unsigned precision in order to improve the detector obtention at the next step (SIFT).

C. SIFT

One of the most powerful approaches for the obtention of local descriptors is the SIFT [19], [22]. The SIFT approach transforms image data into scale-invariant coordinates relative to local features and is based on four major stages: scale-space extrema detection, keypoint localization, orientation assignment, and keypoint descriptor [19].

Let $J(x, y)$ be an image and $L(x, y, \sigma)$ the scale space of J, which is defined as

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

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

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

The scale-space extrema detection begins with the detection of local maxima and minima of $D(x, y, \sigma)$, defined as the convolution of a difference of Gaussian with the image $J(x, y)$ [18]

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

The detection is performed by searching over all scales and image locations in order to identify potential interest points that are invariant to scale and orientation.

Once a set of keypoint candidates is obtained, the next step is to accurately localize them. This is performed by rejecting those keypoints, which have low contrast or are poorly localized along an edge, by a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. Unstable extrema with low contrast are detected by considering a threshold over the extremum of the Taylor expansion (up to the quadratic terms) of $D(x, y, \sigma)$.

The third stage of the SIFT approach is the orientation assignment to each keypoint, based on local image gradient directions. This allows for the representation of each keypoint relative to this orientation, achieving invariance to image rotation. It is performed through an orientation histogram formed from the gradient orientations of sample points within a region around the keypoint, having 36 bins covering the 360° range of orientations. Each sample added to the histogram is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a σ that is 1–5 times that of the scale of the keypoint. Then, a thresholding-based procedure refined by a parabola fitting is used to accurately determine the keypoint orientation through the orientation histogram.

The last stage of the SIFT approach is the keypoint descriptor. The previously described steps assigned the location, scale, and orientation of each keypoint. The motivation for the computation of a more complex descriptor is to obtain a highly distinctive keypoint and invariant as possible to variations such as change in illumination or 3-D viewpoint. Each resultant SIFT descriptor is a 128-element feature vector, whose detailed explanation can be found in [19].

D. Obtention of Matching Candidates

Under the scope of automatic registration of satellite images, since several distortion effects may be present in an acquired image (as already mentioned in the Introduction), it is desirable to have a reference image with as little distortions as possible (such as an orthoimage with no shadow effects and similar spectral content). Having that in mind, the SIFT descriptors of the reference image may be used as a reference database of keypoints, used for matching the keypoints derived from the image to be registered. In this paper, we have considered the nearest neighbor approach for keypoint matching as proposed in [19].

The nearest neighbor is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector. An effective measure for a matching validation is the ratio between the distance of the closest neighbor and the distance to the second closest neighbor, hereafter assigned as d_{ratio}. An efficient nearest neighbor indexing is performed through the best-bin-first algorithm [1]. Although, in [19], a distance ratio threshold of 0.8 was proposed, a sensitivity analysis of this parameter to different satellite images has been performed in this work.

E. Outlier Removal

Even after the identification of matching candidates after removal of incorrect initial matches as described in the previous section, remote sensing images still produce unreliable tie points which lead to a further incorrect geometric correction. Each SIFT keypoint specifies four parameters: 2-D location (x and y), scale, and orientation (θ). The proposed refinement

```
while (RMS_all ≥ 1) and (number of iterations ≤ 10) {
    Compute the bivariate histogram of the matching candidates;
    Identify the maximum absolute frequency of the histogram;
    Exclude matching candidates belonging to an histogram bin with absolute
    frequency ≤ 10% of the maximum absolute frequency of the histogram;
    Compute RMS_all of the remaining matching candidates;
    Update number of iterations;
}
```

Fig. 2. Pseudocode of the outlier removal stage described in Section II-E.

of excluding unreliable tie points is an iterative procedure based on the principle that the "correct set" of matching keypoints corresponds to a denser region of the 2-D representation of the horizontal (Δ_x) and vertical (Δ_y) distances between the matching candidates. A pseudocode of the algorithm is shown in Fig. 2. An example of a set of matching candidates is shown in Fig. 3(a), corresponding to the segmentation of the images shown in Fig. 4. Each iteration of the outlier removal procedure consists in the analysis of the bivariate histogram associated to the previously described scatter plot, retaining only those bins with an absolute frequency greater than 10% of the maximum absolute frequency. The number of bins is a sensitive point and is selected according to the Sturges rule [29]. This rule provides an objective indication of the number of bins as being $1 + 3.322 \log_{10} N$, where N is the number of observations. The histogram in Fig. 3(b) corresponds to the bivariate histogram of the data in Fig. 3(a). The procedure stops when the registration accuracy measure RMS_{all} (described in the next section) is below one or the maximum number of iterations is achieved. Based on our experiments, a maximum number of ten iterations are sufficient for the procedure to converge to a subset of valid matching keypoints. The brighter points in Fig. 3(a) correspond to the points considered valid after the first iteration of outlier removal.

F. Final Set of Tie Points and Geometric Correction Measures

The final set of tie points is composed by the initial matching candidates after the removal of the outliers detecting according to the procedure described in the previous section. The performance of the proposed methodology for AIR was evaluated through measures recently proposed [10], which allow for an objective and automatic evaluation of the image registration process quality, based on theoretical acceptance levels. The description of these measures may be found in [10] and include the following: N_{red} (number of redundant points), RMS_{all} (rmse considering all Control Points (CPs) together), RMS_{LOO} (rmse computation of the CP residuals, based on the leave-one-out method), p_{quad} (statistical evaluation of residual distribution across the quadrants), $BPP(1.0)$ (bad point proportion with norm higher than 1.0), S_{kew} (statistical evaluation regarding the presence of a preference axis on the residual scatter plot), S_{cat} (statistical evaluation of the goodness of CP distribution across the image), and ϕ (a weighted combination of the seven previously referred measures). Both rms measures are normalized to the pixel size.

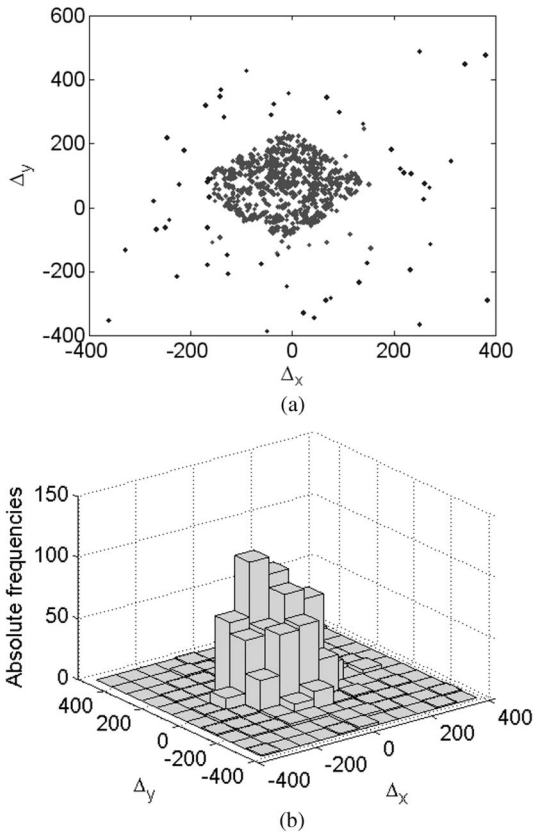


Fig. 3. (a) Scatter plot of a set of horizontal (Δ_x) and vertical (Δ_y) distances between the matching candidates obtained from segmentation of the pair of images in Fig. 4 considering $d_{\text{ratio}} = 0.8$. (b) Bivariate histogram of the values in (a) with 11 bins in each direction.

III. Results

The methodology previously described in Section II for AIR was applied to four different pairs of remote sensing images. For each pair, the most proper transformation function was applied according to the involved distortion. The performance of the method was evaluated through measures recently proposed in the literature [10], as also described in Section II. The Landsat and Hyperion images are courtesy of the U.S. Geological Survey. The images were processed on a computer with an Intel Core 2 6400 2.13-GHz processor and 2.0 GB of physical memory, using MATLAB Release 2009 b.

A. Landsat/Landsat: Simulated Distortion

This pair of images was obtained from a Landsat scene acquired on 2009-10-15 by Landsat 5, corresponding to the northwest part of the Iberian Peninsula, with a sun azimuth and elevation angles at the time of acquisition of 157° and 37°, respectively. A segment of 512 × 512 pixels from band 5 was selected as the reference image. A segment of similar size was selected from band 7 after a simulated rotation of 30° and a simulated scale factor of 0.7, with this being the image to be registered onto the reference image. Both images are shown in Fig. 4. The simulated distortion can be corrected through a first-order polynomial.

The obtained measures as a function of d_{ratio} are shown in Fig. 5, and the final set of keypoints (location, scale, and

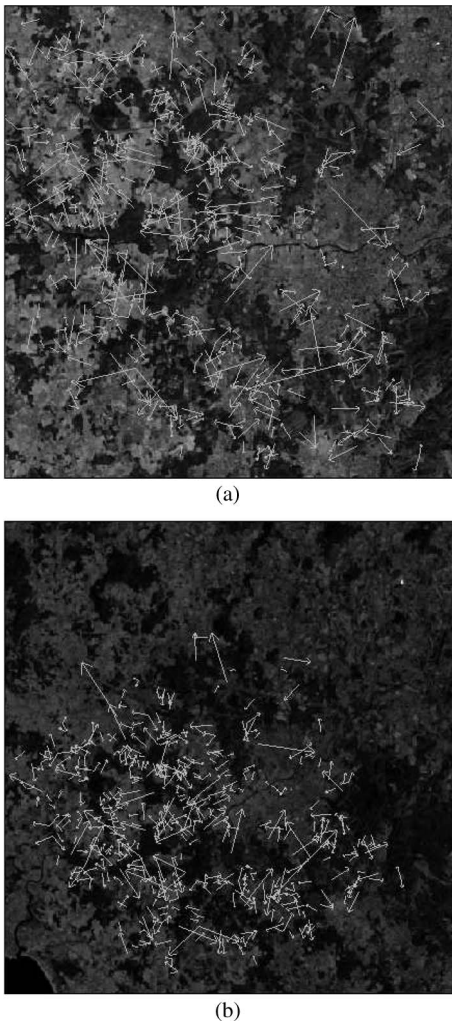


Fig. 4. (a) Reference image extracted from band 5 of a Landsat 5 scene (512 × 512 pixels). (b) Image extracted from band 7 of a Landsat 5 scene rotated by 30° and with a scale factor of 0.7 (512 × 512 pixels). The final set of keypoints (location, scale, and orientation representation) used for registration is superimposed on both images for $d_{\mathrm{ratio}} = 0.7$.

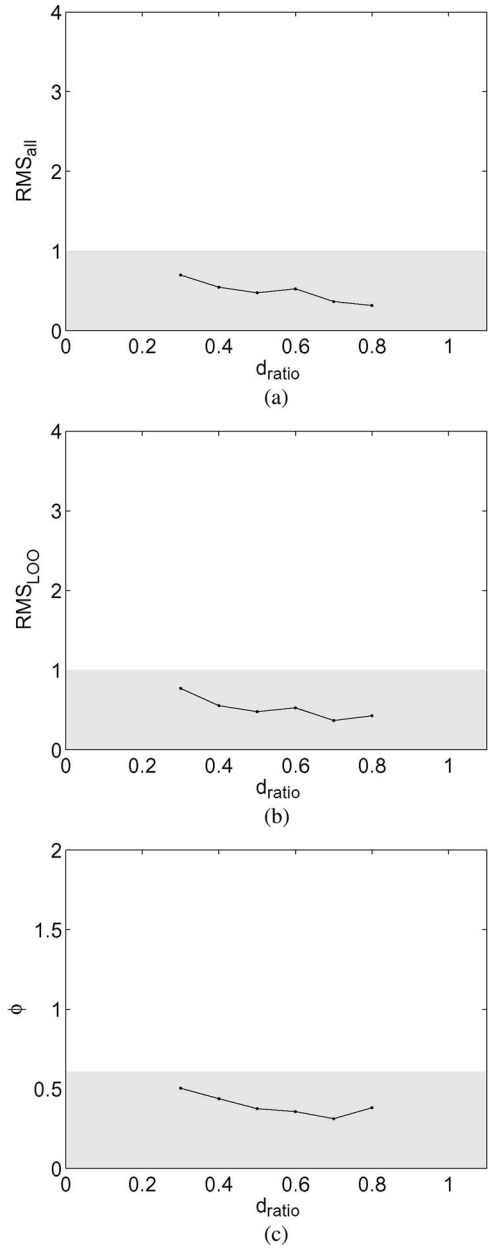


Fig. 5. Measures (a) RMS_{all}, (b) RMS_{LOO}, and (c) ϕ as a function of d_{ratio} regarding the registration of the pair of images described in Section III-A.

orientation representation) for $d_{\mathrm{ratio}} = 0.7$ (the value which led to the lowest value of ϕ) superimposed on both images (as white arrows) is shown in Fig. 4. The low rms values, which are achieved even for low values of d_{ratio}, are noticeable. In particular, for $d_{\mathrm{ratio}} = 0.4$, 124 tie points have been obtained, associated to an $RMS_{\mathrm{all}} = 0.54$ (and an $RMS_{\mathrm{LOO}} = 0.55$) and $\phi = 0.438$ (below the acceptance level of 0.605). For $d_{\mathrm{ratio}} = 0.3$, although a lower number of 21 tie points have been obtained, a subpixel accuracy ($RMS_{\mathrm{all}} = 0.70$ and $RMS_{\mathrm{LOO}} = 0.77$) has also been achieved. These results are not so far from those obtained with $d_{\mathrm{ratio}} = 0.7$ ($RMS_{\mathrm{all}} = 0.36$, $RMS_{\mathrm{LOO}} = 0.37$, and $\phi = 0.313$), which led to the obtention of 527 tie points. Therefore, the small number of tie points obtained with $d_{\mathrm{ratio}} = 0.3$ is quite accurate and sufficient to perform an accurate registration of this pair of images at the subpixel level, with different spectral content and a significant affine (rotation, scale, and translation) distortion. The registration accuracy for $d_{\mathrm{ratio}} = 0.7$ can be visually assessed through Fig. 6.

B. Hyperion/Hyperion

This second pair of images was obtained from two EO1/Hyperion scenes covering the Chesapeake Bay acquired at different times: 2002-09-06 and 2004-10-06. From the 242 available spectral bands, 26 saturated bands and 44 uncalibrated bands were excluded from further processing. Hyperion scenes are, in general, quite narrow, which led to the only available segment of size 256 × 256 pixels common to both scenes with a significant land content. The segment from the 2002 scene was considered as the reference image. As described in Section II, PCA was applied to each 256 × 256 segment containing the remaining 172 spectral bands. The proportions of variance explained by the first principal component were 90.2% and 89.8% for the 2002 and 2004 segments, respectively.

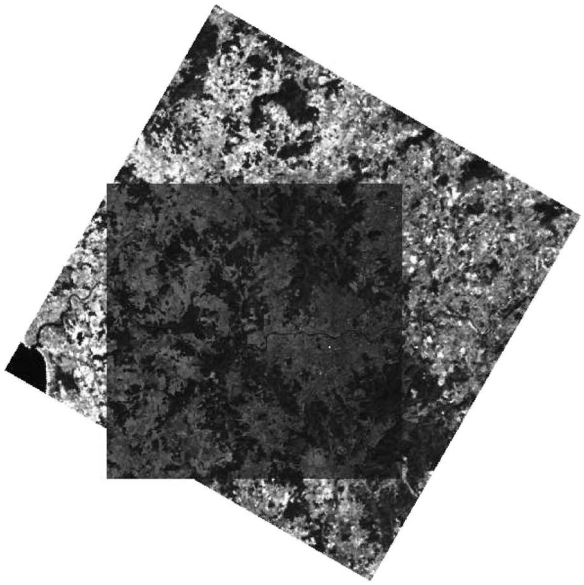


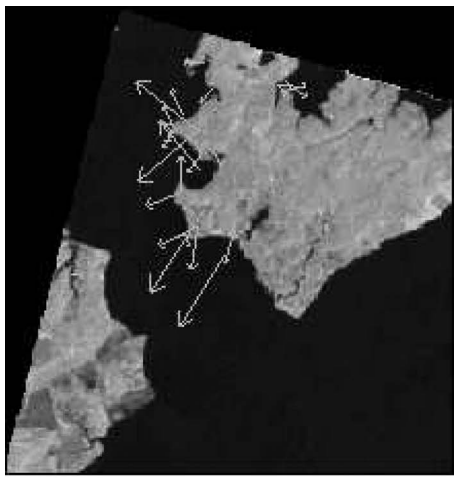
Fig. 6. Image in Fig. 4(a) superimposed on image in Fig. 4(b) after contrast stretched to allow for a better visualization of the registration accuracy, with respect to $d_{\text{ratio}} = 0.7$.

The first principal component from both segments is shown in Fig. 7. The distortion present in this pair of images can be corrected through a first-order polynomial.

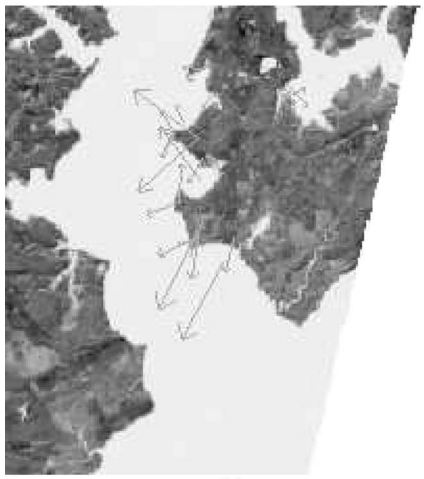
The results obtained for this pair of images are shown in Fig. 8, and the final set of keypoints (location, scale, and orientation representation) for $d_{\text{ratio}} = 0.8$ (the value which led to the lowest value of ϕ) superimposed on both images is shown in Fig. 7. A subpixel accuracy regarding both RMS_{all} and RMS_{LOO} and an overall validation of the registration accuracy have been achieved for values of d_{ratio} ranging from 0.2 to 1.0 (only the range 0.1–1.0 was considered). This reinforces the robustness of the proposed methodology for AIR, since, ranging from merely 5 tie points ($d_{\text{ratio}} = 0.2$) to 50 tie points ($d_{\text{ratio}} = 1.0$), a subpixel accuracy has been achieved. Furthermore, for $d_{\text{ratio}} = 0.8$ (associated to the lowest value of ϕ), the 28 obtained tie points were associated to an $RMS_{\text{all}} = 0.35$, $RMS_{\text{LOO}} = 0.41$, and $\phi = 0.302$. This is another example where not only it was possible to obtain a robust set of tie points for low values of d_{ratio} but also a quite accurate registration was achieved.

C. Hyperion/Landsat

The third considered pair of images is composed by the Hyperion scene from 2002 described in Section III-B and by the bands 1–5 and 7 from a Landsat 5 scene also covering the Chesapeake Bay acquired on 2009-06-29. As already mentioned in Section III-B, EO1/Hyperion scenes are quite narrow, and so, a segment of 256 × 256 pixels was the maximum region of intersection between these two images. The proportion of variance explained by the first principal component of the Landsat scene was 88.6%. The first principal component from both segments is shown in Fig. 9. The distortion present in this pair of images can also be corrected through a first-order polynomial.



(a)



(b)

Fig. 7. (a) Segment (reference) with 256 × 256 pixels from the first principal component of the Hyperion scene from 2002 covering the Chesapeake Bay. (b) Segment with 256 × 256 pixels from the first principal component of the Hyperion scene from 2004 covering the Chesapeake Bay. The final set of keypoints (location, scale, and orientation representation) used for registration is superimposed on both images for $d_{\text{ratio}} = 0.8$. Further details are given in Section III-B.

For this multisensor registration example, it was still possible to achieve a registration accuracy at the subpixel level for a wide range of d_{ratio} values (Fig. 10). The final set of keypoints (location, scale, and orientation representation) for $d_{\text{ratio}} = 0.7$ (the value which led to the lowest value of ϕ) superimposed on both images is shown in Fig. 9. Regarding RMS_{LOO}—which is a more fair evaluation of the rms error—it was also possible to achieve a subpixel accuracy for most of the considered d_{ratio} values. Furthermore, since $BPP(1.0)$ was around 0.2 for most values of d_{ratio}, excluding only those detected bad points would drastically reduce the obtained rms-based measures, which would be possible since a sufficient number of tie points were obtained. The high values obtained for S_{cat} (most near one) are justified by the large part of sea, where it is not possible to identify any matching keypoint. An overall validation of the registration accuracy has been achieved for most of the values of d_{ratio}, despite the fact that this pair of images is a multisensor example. The number of obtained tie points was between 9 for $d_{\text{ratio}} = 0.3$ and 33 for $d_{\text{ratio}} = 1.0$.

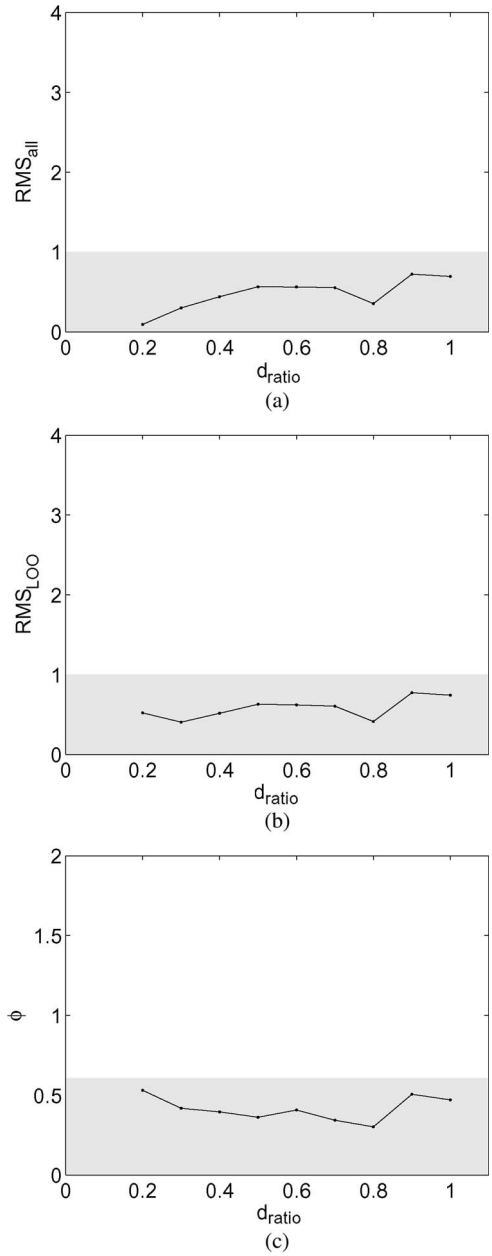


Fig. 8. Measures (a) RMS_{all}, (b) RMS_{LOO}, and (c) ϕ as a function of d_{ratio} regarding the registration of the pair of images described in Section III-B.

D. Orthophoto/ALOS

The last considered pair of images was a high spatial resolution data set, composed by the red, green, blue, and near-infrared bands of an orthophotograph and by an ALOS-PRISM scene, covering a region from the North of Portugal. A segment of 512 × 512 from these scenes was selected. The points with minimum and maximum height across the considered region are 45 and 69 m, respectively, obtained from the Shuttle Radar Topography Mission DEM [7]. Therefore, a first-order polynomial is sufficient to accurately register this pair of images. The proportion of variance explained by the first principal component of the ALOS segment was 88.6%. The first principal component from the orthophoto and the ALOS segment are shown in Fig. 11. This is an example where

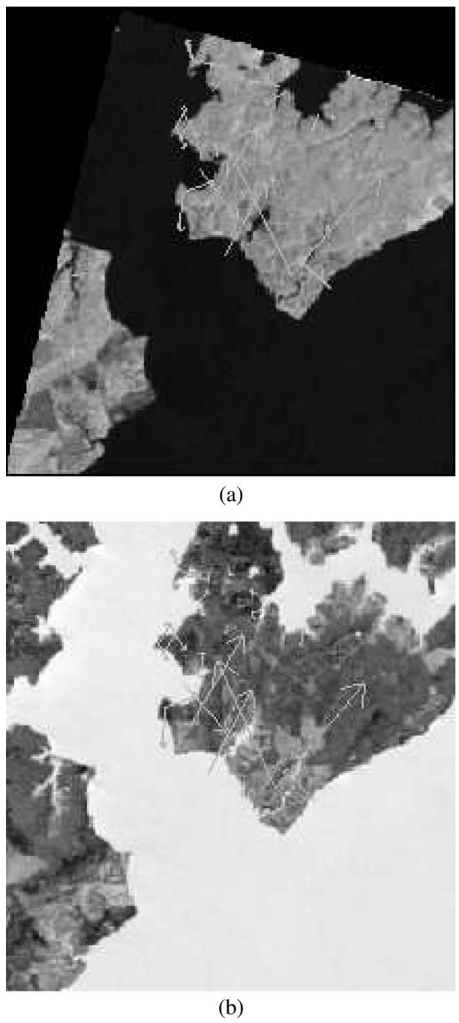


Fig. 9. (a) Segment (reference) with 256 × 256 pixels from the first principal component of the Hyperion scene from 2002 covering the Chesapeake Bay. (b) Segment with 256 × 256 pixels from the first principal component of the Landsat 5 scene from 2009 covering the Chesapeake Bay. The final set of keypoints (location, scale, and orientation representation) used for registration is superimposed on both images for $d_{\mathrm{ratio}} = 0.7$. Further details are given in Section III-C.

the histogram shapes of both images are quite similar to a unimodal distribution. Nevertheless, considering d_{ratio} as 0.9, it was possible to obtain a subpixel accuracy ($RMS_{\mathrm{all}} = 0.54$ and $RMS_{\mathrm{LOO}} = 0.66$) and an overall measure quite below the acceptance level ($\phi = 0.373$) with 22 tie points.

E. Comparison With Other AIR Methods

The proposed methodology for AIR was compared with three methods for the four previously described pairs of images. The first was based on the correlation coefficient (CC) [2], [15], [32], where each image was divided into tiles of size 128 × 128 pixels allowing for the obtention of a set of tie points based on the center of the tile and the identified peak of the similarity surface. The similarity measure mutual information (MI) was similarly considered as the second method of comparison, whereas SIFT was the third considered method [19]. The obtained results are presented in Table I. It is worth to mention that, for the three methods used for comparison, they were

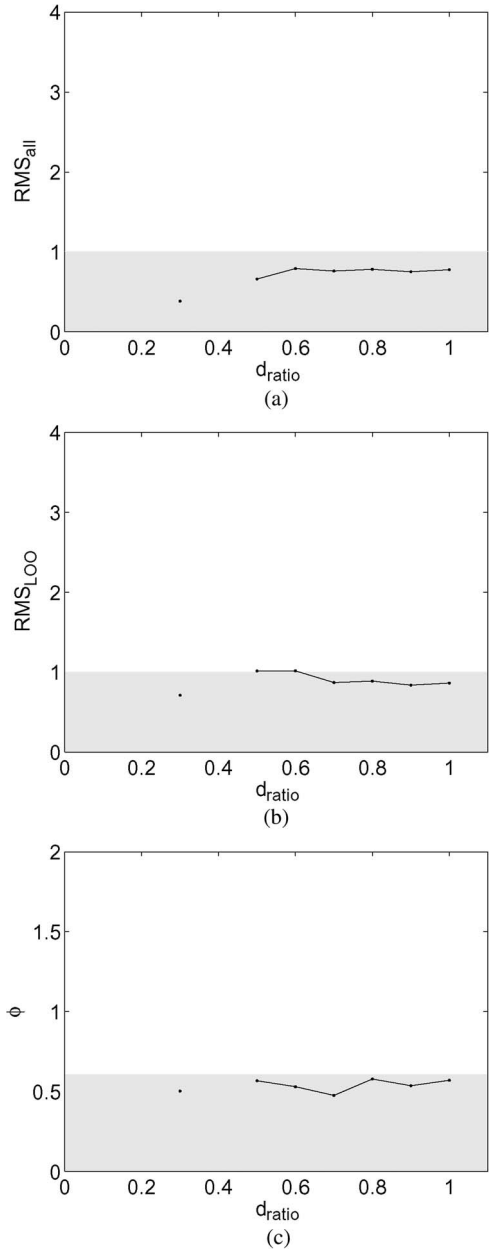


Fig. 10. Measures (a) $RMS_{\rm all}$, (b) $RMS_{\rm LOO}$, and (c) ϕ as a function of $d_{\rm ratio}$ regarding the registration of the pair of images described in Section III-C.

applied to the first principal component of the images whenever applicable, which is a part of the proposed methodology.

Although SIFT has achieved results comparable to our proposed method for Section III-A using $d_{\rm ratio} = 0.4$, it would lead to $RMS_{\rm all} = 34.000$, $RMS_{\rm LOO} = 34.193$, and $\phi = 8.994$ with $d_{\rm ratio} = 0.8$ (as proposed in [19]), which are clearly unacceptable results. Therefore, the use of the method as proposed in [19] may become quite risky, since a wrong choice of the $d_{\rm ratio}$ parameter may lead to dangerous results. On the contrary, the worst results obtained by our proposed methodology were $RMS_{\rm all} = 0.697$, $RMS_{\rm LOO} = 0.770$, and $\phi = 0.504$ for $d_{\rm ratio} = 0.3$. One of the main reasons behind this may be related to our "outlier removal" stage, which robustly eliminates those erroneous tie points. For the remaining pairs

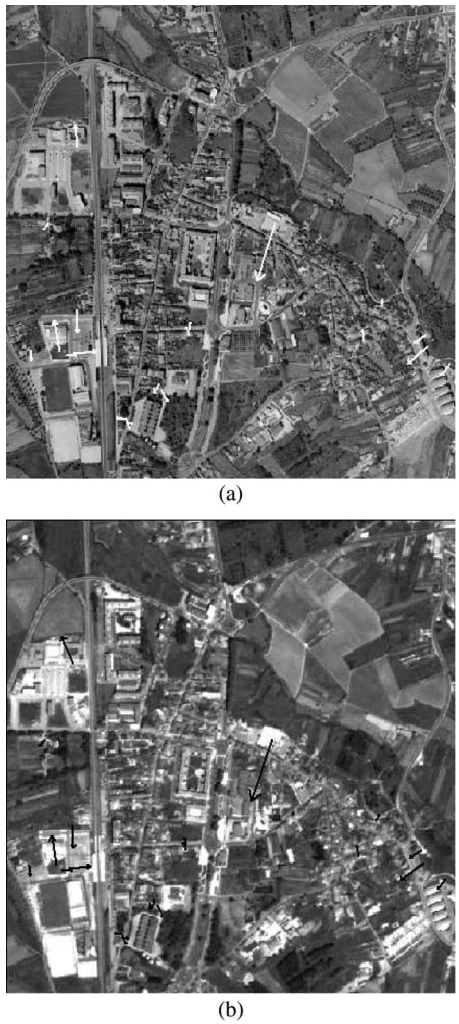


Fig. 11. (a) Segment (reference) with 512 × 512 pixels from the first principal component of the orthophotograph. (b) Segment with 512 × 512 pixels from an ALOS-PRISM scene. The final set of keypoints (location, scale, and orientation representation) used for registration is superimposed on both images for $d_{\rm ratio} = 0.9$. Further details are given in Section III-D.

of images, our method clearly outperformed SIFT, in particular for the medium spatial resolution pairs in Sections III-B and C. This may reveal a limitation of SIFT when dealing with medium spatial resolution images, since that for the high spatial resolution pair in Section III-D, SIFT led to better results, although worst than those obtained with our methodology. With the exception of the pair in Section III-C—where our proposed method clearly outperformed CC, MI, and SIFT—with the proposed methodology, it is possible to obtain more and accurate tie points than using merely SIFT. The advantage of using the proposed methodology against CC and MI is quite obvious, despite the fact that, with CC, it was possible to achieve comparable results for the image pair in Section III-D.

Regarding computational efficiency, the results of the proposed methodology presented in Table I were associated to processing times of 98, 28, 22, and 101 s, for the image pairs in Sections III-A–D, respectively. Although these are quite acceptable processing times, comparable to those obtained with the other three methods, there is still room for further improvements, since the implemented MATLAB code is not yet optimized in terms of computational efficiency.

TABLE I
MEASURES N_{red}, RMS_{all}, RMS_{LOO}, p_{quad}, $BPP(1.0)$, S_{kew}, S_{cat}, AND ϕ, REGARDING THE COMPARISON OF THE GEOMETRIC CORRECTION (OF THE PAIRS OF IMAGES IN SECTIONS III-A–D) PERFORMED BY THE PROPOSED METHOD, WITH CC [2], [15], [32], MI [2], [15], [32], AND SIFT [19], USING A FIRST-ORDER POLYNOMIAL. ABBREVIATIONS ARE EXPLAINED IN THE TEXT (SECTION II-F). [a] THE METHOD WAS NOT ABLE TO REGISTER THIS PAIR OF IMAGES. [b] INADEQUATE NUMBER OF POINTS TO COMPUTE p_{quad} [10]

Pair	AIR method	N_{red}	RMS_{all}	RMS_{LOO}	p_{quad}	$BPP(1.0)$	S_{kew}	S_{cat}	ϕ
III-A	Proposed ($d_{ratio} = 0.7$)	524	0.365	0.367	0.344	0.009	0.077	1.000	0.313
	CC				[a]				
	MI				[a]				
	SIFT [19] ($d_{ratio} = 0.4$)	278	0.235	0.237	0.187	0.004	0.050	1.000	0.257
III-B	Proposed ($d_{ratio} = 0.8$)	25	0.354	0.413	0.097	0.000	0.145	1.000	0.302
	CC	1	6.602	26.408	[b]	1.000	0.800	0.063	6.166
	MI	1	3.200	12.802	[b]	1.000	1.000	0.063	3.279
	SIFT [19] ($d_{ratio} = 0.9$)	62	337.287	335.661	0.996	1.000	0.672	0.409	87.830
III-C	Proposed ($d_{ratio} = 0.7$)	21	0.760	0.870	0.199	0.208	0.260	1.000	0.475
	CC	1	10.012	40.050	[b]	1.000	0.600	0.063	9.061
	MI	1	2.066	8.261	[b]	1.000	0.800	0.063	2.277
	SIFT [19] ($d_{ratio} = 0.7$)	2	81.048	283.547	[b]	1.000	0.900	0.969	62.326
III-D	Proposed ($d_{ratio} = 0.9$)	19	0.544	0.659	0.177	0.000	0.172	0.994	0.373
	CC	13	0.442	0.526	[b]	0.063	0.665	0.081	0.279
	MI	13	11.515	14.217	[b]	1.000	0.068	0.081	4.035
	SIFT [19] ($d_{ratio} = 0.6$)	12	1.115	1.329	[b]	0.333	0.421	0.995	0.688

F. Applicability of a Different Segmentation Method

Under the scope of an AIR methodology, it is strictly necessary that all stages are automatic, which includes the "image segmentation" phase. However, fully automatic image segmentation is still a present subject of research, in particular for natural (including remote sensing) images [24]. Nevertheless, in order to provide some sensitivity analysis of the proposed AIR methodology on the segmentation method, the k-means clustering technique was also considered as an alternative to the Otsu's thresholding method [6]. Despite the fact that it involves some parameters, the set of parameters "number of clusters," "type of distance," and "number of replicates" was equally defined for all the pairs of images as "2," "squared euclidean," and "4," respectively.

A similar performance with the k-means clustering technique for the four pairs of images was found. Regarding the data sets in Sections III-A–D, the values obtained for ϕ were 0.303, 0.347, 0.420, and 0.279, respectively. In the same order, these values were obtained with the values of d_{ratio} equal to 0.8, 0.3, 1.0, and 0.8, for which RMS_{LOO} values of 0.384, 0.183, 0.750, and 0.720 were obtained, respectively. These preliminary results indicate that the application of different segmentation methods than Otsu's thresholding may become a valid alternative.

IV. DISCUSSION

Although several methods have been proposed in the recent years [2], [8], [11], [31], [32], geometric correction of satellite images is, in practice, mostly a manual work. In this paper, a fully automatic algorithm for image registration has been proposed, which comprises PCA, image segmentation, SIFT, and a robust outlier removal procedure. Although this methodology was presented as an image-to-image registration, it can be used to match a new image to an available data set of vectorial information stored in a geographical information system, by converting the vectorial information to a raster format.

The application of PCA allowed for a proper reduction of each image dimension (for multi- or hyperspectral images), since the first principal component explained in all situations more than 88% of the total variance, without compromising the registration accuracy. The use of the remaining principal components does not provide meaningful information for later segmentation and may therefore be discarded in the later stages of the proposed methodology. Based on our experiments, only the second principal component provides some useful information, still leading to worst results than the first principal component.

Other methods of data reduction could have been used for this purpose, such as independent component analysis or the projection pursuit, among the linear projections, or other nonlinear projection methods such as the curvilinear component analysis [14]. However, PCA is still the mostly used method for reducing the number of spectral bands and has provided good results under the scope of the proposed methodology. Although testing other data reduction methods was not under the scope of this work, this is an aspect which will certainly deserve further research in the future.

One of the major strengths of the proposed methodology is the fact that it does not require an accurate segmentation of the objects present on a scene. In several situations, due to

differences in the sensor characteristics, in the spectral content or even (temporal) changes in the terrain, the objects may significantly differ from one image to the other, leading to a difficult accurate registration with the traditional segmentation-based AIR methods [5].

With respect to the image segmentation stage, since the object function $\sigma_B^2(k)$, or equivalently, the criterion measure $\eta(k)$, both described in [25], is always smooth and unimodal, it can be used as a method of evaluating the goodness of the segmentation result in further research. Moreover, a different segmentation approach was also tested according to a unimodal —for which specific segmentation methods exist [27]—or bimodal shape of the histogram. However, there was no significant improvement on the obtained results with this increasing of the method complexity, as the Otsu's thresholding method led to a sufficient segmentation for later processing. Nevertheless, the complexity of the proposed algorithm may be increased in the future, by considering more segmentation methods which may be found to be more adequate to certain types of applications. Although semiautomatic image segmentation methods are available in the literature [24], the segmentation stage has to be fully automatic. A more complex algorithm may require an intelligent system, capable of automatically deciding which method (or combination of methods) should be used for each case, such as decision trees among other possible alternatives [6].

The obtained results for the four pairs of images indicate that a wide range of geometric distortions can be corrected with the proposed methodology, namely, when registering large scenes with significant geometric distortions such as the terrain elevation, with particular importance for some products such as level 1A SPOT scenes.

For the second and third pairs of images (Hyperion/Hyperion and Hyperion/Landsat, multitemporal and multisensor cases), the proposed methodology clearly outperformed SIFT. The reason behind it may be the fact that SIFT is not adequate for remote sensing images—in particular for multitemporal and multisensor medium spatial resolution images—due to the difference of intensity mapping between the images [16]. The proposed stage of image segmentation is responsible in some way for overcoming this problem. For these two pairs of images, the proposed methodology also outperformed CC- and MI-based methods.

The low contrast between building rooftop and background is a frequent concern [28], as well as closely spaced buildings, which leads to undesired segmentation results. This is an inherent difficulty when dealing with single-band (panchromatic) images. Moreover, robust automatic extraction of objects from high spatial resolution images is still a present challenge [28]. Nevertheless, even using the Otsu segmentation method, it was possible to achieve an accurate registration with the pair orthophoto/ALOS.

For most of the considered situations, extremely high values of measure S_{cat} have been obtained, particularly for those with strong rotation and scale effects. The reason behind this is that there is a considerable part of one image which is not covered by the other, and therefore, it will not be possible to identify any tie point on those areas. This comment also applies to regions without any distinctive objects such as sea, or even parts of one image with no data $(\mathrm{DN} = 0)$.

The outlier removal stage is crucial to obtain an accurate registration. If it was omitted, then the criterion of bad point detection included in $BPP(r)$ would not be able to detect erroneous matching candidates, since all tie points would be used to estimate an inadequate set of the transformation function parameters and therefore would wrongly model the geometric distortion. At the outlier removal stage, an alternative of using the scale and orientation keypoint parameters instead of the 2-D location has been tested. However, there was no improvement on the obtained results using this approach instead of the analysis of the horizontal and vertical tie point displacements.

To our knowledge, a sensitivity analysis on the d_{ratio} parameter has been performed in this work for the first time, regarding SIFT-based image registration methods. It allowed for a gain of sensitivity of its use with respect to different types of satellite images, as well as showing some misregistration results when using an inadequate value for parameter d_{ratio}. The proposed methodology for AIR revealed to be robust to a variation of this parameter, not providing any registration when the choice for the parameter is not adequate. This is a highly desirable characteristic when dealing with AIR methods.

An important aspect of the SIFT approach is that it generates large numbers of features that densely cover the image over the full range of scales and locations. For instance, a typical image of size 500 $\times$ 500 pixels will give rise to about 2000 stable features, depending on both image content and choices for various parameters [19]. Therefore, in particular cases of registering large scenes, the proposed methodology should be applied to individual tiles of the image, combining the set of final tie points for the large scene registration. Furthermore, for more complex distortions, including significant variations of the terrain height across the scene, an iterative application of the methodology (excluding the PCA stage) may be required to account for more pronounced geometric distortions.

In the future, it is expected to explore whether joining the combination of the matching candidates and outlier removal stages on an iterative basis may lead to an even more accurate registration, taking into account the most proper transformation function for each situation (also considering the sensor orbit model). It is justified by the fact that, after excluding those matching candidates considered outliers, recomputing the transformation function parameters may lead to an improvement of the obtained results. This may become a complex task when dealing with more complex transformation functions and therefore deserves a deep further research on this topic.

A fully AIR method has been proposed, which combines PCA, image segmentation, SIFT, and a robust outlier removal procedure. The combination of these techniques provides a robust and accurate scheme for AIR. It allows for the registration of a pair of images with different pixel size, translation, and rotation effects, and to some extent with different spectral content, able to lead to a subpixel accuracy. Furthermore, it has shown robustness against an automatic choice of the involved parameters, which is a highly desirable characteristic of this class of methods.

REFERENCES

[1] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbor search in high-dimensional spaces," in *Proc. Conf. Comput. Vis. Pattern Recog.*, Washington, DC, 1997, pp. 1000–1006.

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

[3] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: Advances and prospects," *Pattern Recog.*, vol. 34, no. 12, pp. 2259–2281, Dec. 2001.

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

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

[6] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley-Interscience, 2000.

[7] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, and D. Alsdorf, "The shuttle radar topography mission," *Rev. Geophys.*, vol. 45, p. RG2004, 1992.

[8] L. M. G. Fonseca and B. S. Manjunath, "Registration techniques for multisensor remotely sensed imagery," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1049–1056, Sep. 1996.

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

[10] H. Gonçalves, J. A. Gonçalves, and L. Corte-Real, "Measures for an objective evaluation of the geometric correction process quality," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 292–296, Apr. 2009.

[11] H. Gonçalves, J. A. Gonçalves, and L. Corte-Real, "Automatic image registration based on correlation and Hough transform," in *Proc. Image Signal Process. Remote Sens. XIV*, vol. 7109, *SPIE*, L. Bruzzone, C. Notarnicola, and F. Posa, Eds., 2008, p. 710 90J.

[12] A. Goshtasby, G. C. Stockman, and C. V. Page, "A region-based approach to digital image registration with subpixel accuracy," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-24, no. 3, pp. 390–399, May 1986.

[13] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 7, pp. 498–520, Oct. 1933.

[14] L. Journaux, I. Foucherot, and P. Gouton, "Reduction of the number of spectral bands in Landsat images: A comparison of linear and nonlinear methods," *Opt. Eng.*, vol. 45, no. 6, p. 067002, Jun. 2006.

[15] 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.

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

[17] Z. P. Liang, H. Pan, R. L. Magin, N. Ahuja, and T. S. Huang, "Automated image registration by maximization of a region similarity metric," *Int. J. Imaging Syst. Technol.*, vol. 8, no. 6, pp. 513–518, 1997.

[18] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. Computer Vision*, Corfu, Greece, 2008, pp. 1150–1157.

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

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

[21] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1/2, pp. 43–72, Nov. 2004.

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

[23] A. Mukherjee, M. Velez-Reyes, and B. Roysam, "Interest points for hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 748–760, Mar. 2009.

[24] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Interactive image segmentation by maximal similarity based region merging," *Pattern Recognit.*, vol. 43, no. 2, pp. 445–456, Feb. 2010.

[25] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. SMCB-9, no. 1, pp. 62–66, Jan. 1979.

[26] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, Sep. 1993.

[27] P. L. Rosin, "Unimodal thresholding," *Pattern Recognit.*, vol. 34, no. 11, pp. 2083–2096, Nov. 2001.

[28] B. Sirmaçek and C. Ünsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1156–1167, Apr. 2009.

[29] H. A. Sturges, "The choice of a class interval," *J. Amer. Stat. Assoc.*, vol. 21, no. 153, pp. 65–66, Mar. 1926.

[30] M. F. Vural, Y. Yardimci, and A. Temizel, "Registration of multispectral satellite images with orientation-restricted SIFT," in *IEEE IGARSS*, 2009, vol. 3, pp. III-243–III-246.

[31] 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.

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

Hernâni Gonçalves was born in Porto, Portugal, in 1979. He received the Licenciatura degree in technology applied mathematics from the Faculdade de Ciências, Universidade do Porto, Porto, in 2002 and the M.Sc. degree in computational methods in science and engineering from the Faculdade de Engenharia, Universidade do Porto, in 2004. He is currently working toward the Ph.D. degree in surveying engineering in the Faculdade de Ciências.

He was a Researcher under the scope of a project in biomedical engineering, in the analysis of biomedical signals. Since 2006, he has been a Researcher with the Centro de Investigação em Ciências Geo-Espaciais, Faculdade de Ciências. His research interests include image processing and biomedical signal analysis.

Luís Corte-Real (M'91) was born in Vila do Conde, Portugal, in 1958. He received the B.S. degree in electrical engineering from the Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, in 1981, the M.Sc. degree in electrical and computer engineering from the Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisbon, Portugal, in 1986, and the Ph.D. degree from the Faculdade de Engenharia, Universidade do Porto, in 1994.

In 1984, he was with the Universidade do Porto as a Lecturer of telecommunications. He is currently an Associate Professor with the Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia, Universidade do Porto. He has been a Researcher with the Institute for Systems and Computer Engineering of Porto (INESC Porto), since 1985. His research interests include image/video coding and processing.

José A. Gonçalves was born in Porto, Portugal, in 1964. He received the B.S. degree in surveying engineering from the Faculdade de Ciências, Universidade do Porto, Porto, in 1988, the M.Sc. degree in geographical information systems from the Department of Photogrammetry and Surveying, University College London, London, U.K., in 1993, and the Ph.D. degree from the University College London in 2001.

In 1989, he was with the Universidade de Trás-os-Montes e Alto Douro, Vila Real. In 1997, he was with the Universidade do Porto as a Lecturer on the surveying engineering field. He is currently an Assistant Professor with the Departamento de Geociências, Ambiente e Ordenamento do Território, Faculdade de Ciências, Universidade do Porto. He is also a Researcher at the Centro Interdisciplinar de Investigação Marinha e Ambiental. His research interests include photogrammetry, remote sensing and geo-information.