Multimodal Remote Sensing Image Registration With Accuracy Estimation at Local and Global Scales

Mikhail L. Uss, Benoit Vozel, Vladimir V. Lukin, and Kacem Chehdi

Abstract—**This paper focuses on the potential accuracy of remote sensing (RS) image registration. We investigate how this accuracy can be estimated without ground truth available and used to improve registration quality of mono- and multimodal pair of images. At the local scale of image fragments, the Cramér–Rao lower bound (CRLB) on registration error is estimated for each local correspondence between coarsely registered pair of images. This CRLB is defined by local image texture and noise properties. Opposite to the standard approach, where registration accuracy is only evaluated at the output of the registration process, such valuable information is used by us as an additional input knowledge. It greatly helps in detecting and discarding outliers and refining the estimation of geometrical transformation model parameters. Based on these ideas, a new area-based registration method called registration with accuracy estimation (RAE) is proposed. In addition to its ability to automatically register very complex multimodal image pairs with high accuracy, the RAE method is able to provide registration accuracy at the global scale as a covariance matrix of estimation error of geometrical transformation model parameters or as pointwise registration standard deviation. This accuracy does not depend on any ground truth availability and characterizes each pair of registered images individually. Thus, the RAE method can identify image areas for which a predefined registration accuracy is guaranteed. This is essential for RS applications imposing strict constraints on registration accuracy such as change detection, image fusion, and disaster management. The RAE method is proved successful with reaching subpixel accuracy while registering eight complex mono-/multimodal and multitemporal image pairs including optical-to-optical, optical-to-radar, optical-to-digital elevation model (DEM) images, and DEM-to-radar cases. Other methods employed in comparisons fail to provide in a stable manner accurate results on the same test cases.**

Index Terms—**Area-based registration, Cramér–Rao lower bound (CRLB), digital elevation model (DEM) to radar image registration, multimodal/multitemporal registration, optical-to-DEM, optical-to-radar, polynomial model, registration accuracy, signal-dependent noise model, subpixel accuracy.**

Manuscript received January 20, 2016; revised May 24, 2016; accepted June 18, 2016. Date of publication July 29, 2016; date of current version September 27, 2016.

M. L. Uss and V. V. Lukin are with the Department of Transmitters, Receivers and Signal Processing, National Aerospace University, 61070 Kharkov, Ukraine (e-mail: uss@xai.edu.ua; lukin@ai.kharkov.com).

B. Vozel and K. Chehdi are with IETR UMR CNRS 6164, University of Rennes 1, 22305 Lannion, France (e-mail: benoit.vozel@univ-rennes1.fr; kacem.chehdi@univ-rennes1.fr).

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

Digital Object Identifier 10.1109/TGRS.2016.2587321

I. Introduction

IMAGE registration (coregistration) is typical for many applications of remote sensing (RS) where one needs to bring two (reference and template) or more images of the same area taken by different means and in different conditions to the same coordinate system [1], [2]. Such a transformation allows fusion of multiview, multitemporal, multichannel, and multimodal images [3]. Images can also be registered (superimposed) to sensed terrain digital elevation model (DEM) or digitized topographic map [4]. This provides preconditions for image joint analysis, extraction and dissemination of their informational content, terrain classification, and solving other typical RS tasks [5].

The main requirements to a good registration technique are the following. First, it should provide an appropriate, typically subpixel, registration accuracy [6]. It is important to have some characterization of registration accuracy obtained at the output of the registration process for a given set of processed images (data) in order to be able to control the registration outcomes and to assure that they are appropriate for solving a considered task [7], [8]. Second, a method should be universal and reliable enough, i.e., it has to be applicable for different types of images and various kinds of underlying (sensed) terrains with minimizing probability of registration failure [9]–[11]. Finally, it is desirable to have a fully automatic registration which, at the same time, has to be simple enough to be realized in reasonable time [9], [11].

The main criterion for assessing a registration method performance is its registration accuracy. This accuracy can be described directly or indirectly by a variety of measures proposed in the literature [12]. The traditional criteria for assessing registration accuracy are MSE or RMSE as well as number and coverage of found correspondences. In practice, it is required to provide accuracy as high as possible for a given pair (or a set) of images. With this obvious and straightforward requirement, quantifying precisely the accuracy of a registration method is, nevertheless, an open problem as it will be discussed in the next section.

Concerning the second requirement, a universal registration method should be able to superimpose different types of data modalities typical for RS applications: optical-to-optical, optical-to-radar, and radar-to-radar images as well as optical- and radar-to-DEM or topographic data [4], [10], [11], [13], [14]. For each registration scenario and each registered pair of images, it is desirable either to provide registration accuracy required for a particular application or to specify that such an accuracy cannot be reached.

As for the third requirement, we concentrate in the following on fully automatic techniques. This does not mean that no *a priori* information is used. Many modern methods employ data on image corner geographic coordinates which are often available for exploited RS systems [15]. Moreover, we pay attention to the computation efficiency of the considered and proposed techniques, keeping in mind that complexity of a method can restrict its applicability.

Therefore, accurate, universal, and fully automatic registration with a controllable accuracy is the focus of our research. It is worth noting here that image registration methods have been under design for several decades [16], and therefore, a lot of techniques have been already developed. Good surveys can be found in [1] and [17]. Keeping this in mind, we feel that it is necessary to give a brief review of existing techniques that can be considered as the closest analogs of the method that we propose in this paper (see Section II). In turn, distinctive features of our method are the following. First, our method analyzes and controls potential registration accuracy at local and global scales with application to both linear and nonlinear geometrical transformations (Section III). Due to this, one is provided with pixelwise evaluated registration accuracy for a given pair of processed images without any ground truth available. Some aspects of processing acceleration are considered in Section III as well. Second, our method is fully automatic and universal. This is demonstrated in Section IV, where eight different cases of registration for various possible types of images (data) are studied and quantified. Comparisons to other techniques are also presented there. Third, our method is accurate and reliable. Its ability to provide high registration accuracy (subpixel or close to subpixel) is demonstrated alongside with the ability to deal with complex registration cases where many other existing techniques fail to perform properly (see Section IV). The conclusion in Section V summarizes the obtained results and discusses possible further directions of research.

II. Brief Review of Existing Approaches

In this section, we briefly review the problem of quantifying registration accuracy achieved by the area-based or feature-based method. The difference between them is in the direct use of image intensities in the registration process for area-based methods (also called intensity-based methods) or differently an "informational extract" from these intensities, called features and possessing some degree of invariance to image formation conditions (illumination, geometrical distortions, and modality) for the feature-based method. We also recall existing approaches for multimodal image registration and reported accuracy of such methods.

This paper considers the widespread and modern registration approach [1] for which registration at the global image scale is fully based on correspondences between pairs of registered image fragments detected at the local scale using either feature-based or area-based methods. Following this approach, to quantify the registration accuracy at the global scale, one should be able to determine the registration accuracy of each eligible correspondence.

Concerning feature-based methods, solutions exist for simple point-like features only [18]. For complex features like scale-invariant feature transform (SIFT), Harris corners, or object contours, no solutions have been published yet to the best of our knowledge.

Using area-based methods, local correspondences are mainly found by optimizing a suitable similarity measure [19]. The potential registration accuracy achievable using the sum of squared differences measure was derived in [7] and [8] in the form of Cramér–Rao lower bound (CRLB) on estimation error of geometrical transform parameters. As it has been shown in [20], this solution does not reflect correctly the dependence of potential registration accuracy on signal-to-noise ratio (SNR) and correlation between the reference and template images. Thus, it cannot be considered for application in multitemporal and multimodal cases. Characterization of registration accuracy further complicates for more sophisticated similarity measures like normalized correlation coefficient (NCC), mutual information (MI), or phase correlation.

Without knowing the registration accuracy at the local scale, it is impossible to obtain the registration accuracy at the global scale, i.e., the estimation accuracy of the parameters of the geometrical transformation model between the reference and template images. For simple point-like features, the solutions for registration accuracy at the global scale were derived in [8] and [21]. The problem with these approaches is that they do not take into account outliers among correspondences, a situation rarely met in practice.

The maximum likelihood estimation sample consensus registration method initially developed by Torr and Zisserman in [22] and recently upgraded by Ma *et al.* and Zhao *et al.* in [23] and [24] comes furthermost in the direction of an adequate use of feature registration accuracy. It solves the registration problem by optimizing the complete likelihood function via the expectation maximization (EM) approach and by including registration accuracy of features (e.g., SIFT) in the score function. However, in this approach, feature registration accuracy does not come directly from the feature property analysis. Instead, it follows as an *a posteriori* estimate at the registration process output. This estimate is the same for all features, and thus, it does not characterize specifically individual features that may exhibit significantly different properties. Let us note that the registration accuracy at the global scale was not derived in [22]–[24]. Similarly, the registration error of point correspondences is characterized as a variance of residual registration noise in [25]. This variance is used at the outlier rejection stage and takes the same value for all correspondences.

The aforementioned methods illustrate a common approach that consists in characterizing registration accuracy *a posteriori* at the output of the registration process, mainly caused by inability to quantify registration accuracy at the local scale. For instance, if the registration accuracy of an individual feature cannot be directly derived from the analysis of registered image properties, one has to analyze feature position deviation from a ground truth. For research purposes, the ground truth can be obtained from collected ground control points (e.g., GPS based). Here, we assume a more realistic situation with no ground truth available. In this situation, the registration accuracy of the features is typically inferred by analyzing their deviation from the estimated geometrical model (e.g., utilizing leave-one-out approach). We use the term *a posteriori* to underline the dependence of the analysis carried out on the registration output. While such a reasonable analysis gives an idea of a registration method performance, it has two serious drawbacks. First, the registration accuracy of a method evaluated for one pair of images does not characterize its performance for other pairs. Indeed, registration accuracy depends on the similarity of the registered images that may vary significantly from one pair to another one (e.g., due to registration problem type: either monomodal, multitemporal, or multimodal) or within registered images as a result of cloud cover influence or different properties of land cover types like urban, rural, forest, water surface, etc. Second, this *a posteriori* registration accuracy does not support the registration process as it is unknown (unavailable) at the beginning of the registration process.

The goal of this paper is to investigate more deeply the registration accuracy of RS images both at the local and global scales so as to highlight the advantages that can be drawn from using this additional information in the registration process. For this purpose, we associate the potential registration accuracy with each correspondence between pairs of image fragments where this accuracy is intrinsically linked with the properties of registered image fragments. Potential registration accuracy is then utilized in the registration process as additional *a priori* information, and it also supports the outlier/inlier detection stage. Finally, at the output, we quantify registration accuracy at the global scale with no use of ground truth or manual control points possibly available.

The possibility to move forward in the direction of quantifying registration accuracy comes essentially from the CRLB $\mathrm{CRLB}_{\mathrm{fBm}}$ on image registration errors first introduced in [20] for pure translation model and further developed in our paper [26] for the rotation-scaling-translation (RST) model. Here, index fBm stands for fractal Brownian motion model used to derive $\mathrm{CRLB}_{\mathrm{fBm}}$ bound. Recently, the $\mathrm{CRLB}_{\mathrm{fBm}}$ bound has been extended to the case of optical-to-radar image registration by accounting for the spatial correlation properties of speckle noise and its strong signal dependence [27]. An interesting advantage of the $\mathrm{CRLB}_{\mathrm{fBm}}$ bound over state-of-the-art alternatives is that it can be applied to evaluate the registration accuracy of correspondences found by the area-based approach for all kinds of registration problems including monomodal, multiview, multitemporal, and multimodal cases. At the local scale, it takes into account the reference and template image fragment texture and noise properties such as texture amplitude and roughness, correlation between the reference and template images, noise spatial correlation, and signal dependence. This confers to the $\mathrm{CRLB}_{\mathrm{fBm}}$ bound the capability for accurately predicting the registration accuracy of known area-based registration methods including NCC, MI, and phase correlation [20], [26], [27].

The main contribution of this paper is a new area-based image registration scheme, called registration with accuracy estimation (RAE) that beneficially involves registration accuracy at all stages, including the preliminary search of putative correspondences (PC), followed by outlier detection and estimation of geometrical transformation parameters. At the local scale, we assign the registration accuracy estimated using the $\mathrm{CRLB}_{\mathrm{fBm}}$ bound to each PC between control fragments (CF) of registered images. At this stage, the main properties of texture and noise of a registered pair of image CFs are taken into account. Registration accuracy is directly exploited to range PCs in order of increasing contribution to registration. At the outlier detection stage, the leave-one-out cross-validation (LOOCV) approach is employed: the actual position of each PC is compared to its prediction based on other PCs. Detection is done by comparing the error of PC position prediction with a threshold. The threshold for each PC is calculated using PC registration accuracy previously derived. At the stage of estimating geometrical transform parameters, the covariance matrix of parameter estimates is obtained from the derived PC registration accuracy and outlier detection results. At the output, we provide the standard deviation (SD) of the registration error for each pixel of the reference image. This registration error is individual for each pair of the registered image fragments. It reflects more adequately the structure of the registered images that is a composition of areas suitable for registration in higher or less degree: urban, rural, forest, cloud cover, etc., and noise properties of both reference and template images.

For the sake of clarity, let us summarize the elements of our previous works that are inherited by the newly proposed RAE method.

1) The $\mathrm{CRLB}_{\mathrm{fBm}}$ lower bound on RST parameter estimation error derived in [20] and [26].
2) Estimator of the fBm-field parameters from the noisy fragments of the reference and template images aligned with the RST model [26].
3) Extension of the $\mathrm{CRLB}_{\mathrm{fBm}}$ to the case of more practically realistic noise model including signal dependence and nonnegligible spatial correlation [27]. This noise model is essential for multimodal registration cases, especially when radar and DEM images are involved.

The core of the proposed RAE method is a novel criterion for joint outlier detection and estimation of registration parameters. We extend the criterion proposed by Ma *et al.* in [23] to better take into account the estimated registration accuracy of each PC and allow a single template CF to have multiple PCs to the reference image fragments. The other two improvements allow us, on one hand, to get rid of additional constraint on geometrical transformation, namely, local linearity, and, on the other hand, to reformulate outlier detection from the point of view of the LOOCV approach. The new criterion that we propose assures that the found correspondences do not contain high-leverage points [28]. We will demonstrate that all of these improvements allow performing successful registration for general affine and second-order polynomial models when probability of finding the true correspondences is as low as 2%.

Asymptotically, the RAE method is characterized by linearithmic complexity with respect to image size (number of control fragments). However, in practice, it has high computational complexity as it relies on $\mathrm{CRLB}_{\mathrm{fBm}}$ bound. We reduce this complexity by an incremental scheme where each newly found correspondence between the reference and template image fragments shrinks the search zone in the geometrical transform parameter space. With this approach, we are able to reach acceptable registration time.

Let us underline that the input of the RAE method is a set of PCs found by any area- or feature-based approach provided that the registration accuracy of a method used can be precisely quantified, i.e., there is a way to predict registration error SD directly from image properties (SNR, noise signal dependence, noise spatial correlation, texture roughness, structural properties, etc.) without performing actually a registration with this method. In this paper, we follow the area-based approach using the NCC similarity measure [19], [29], [30] and intensity interpolation for reaching subpixel registration accuracy. The choice of the NCC method is justified by the fact that its performance is rather close to the $\mathrm{CRLB}_{\mathrm{fBm}}$ bound in those image areas where $\mathrm{CRLB}_{\mathrm{fBm}}$ is the most suitable, i.e., isotropic textures with normal increments [26]. In the experimental part of this paper, the capabilities of the proposed RAE method using NCC will be demonstrated by solving the most complex registration problems including radar-to-optical, optical-to-DEM, and DEM-to-radar real multimodal scenarios in a unified manner.

The first case, optical-to-radar registration, is a well-studied problem for which two dominant approaches exist at the moment: either methods utilizing MI similarity measure or feature-based methods [31]. Registration of high-resolution optical-to-radar images in urban areas using MI similarity measure with additional segmentation step was studied in [10]. The method can only deal with pure translation estimation. The reported registration RMSE is from 0.96 to 2.6 pixels for large control fragments of size 300 by 300 pixels. MI measure was used in [11] to find and localize correspondences between the optical and radar images at the local scale. The drawback of the method is that it finds a very limited number of correspondences.

Hui *et al.* explored the contour-based approach to optical-to-radar image registration in [13]. The method applies chain-based correlation of closed contours and salient features of open contours to estimate parameters of RST transform between the optical and radar images. The reported RMSE of control points is about $1.1, \ldots, 2.1$ pixels. However, the method can be applied only when initial registration error is small (less than about 5 pixels). A descriptor called shape-context is proposed for optical-to-radar image registration in [32]. It is based on distribution of edge features in log-polar space. For this method, the reported RMSE of the registered control points is about 1.8 pixels. The classical SIFT descriptor has been found not suitable when directly applied to RS images in general [33] and to optical and radar images in particular [11]. Improvements to the SIFT descriptor were proposed by Suri *et al.* in [34] and further developed by Bin *et al.* in [31]. A mixed approach utilizing the MI method at the coarse-registration stage and line features at the fine-registration stage [14] demonstrated a registration RMSE of 5 pixels. Overall, the reported RMSE of control points for optical-to-radar image registration varies from 1 to 5 pixels [4], [10], [14], [31], [32]. It is interesting to note that the lower boundary of this interval—1 pixel—was identified as a lower limit for optical-to-radar image registration accuracy (RMSE of correspondences) in [27]. In this paper, we will demonstrate that the RAE method is able to provide an RMSE of about $0.75, \ldots, 1.05$ pixels and subpixel registration accuracy at the global scale for images covering rather featureless areas without large-scale water objects where other methods (used in comparisons) either provide less accurate results or fail.

Optical-to-DEM registration is even more complex (we refer interested readers to the experimental section of [35] for discussion of such a registration scenario complexity). Successful optical-to-DEM image registration based on contour-based approach and nonuniform rational B-splines was reported in [4]. The method can deal with both affine and perspective geometrical models, and the achieved RMSE of control points is $2.3, \ldots, 2.8$ pixels. The drawback of this method is in the usage of manual segmentation stage and inability to provide control points in image areas without rivers or other water object boundaries. Murphy and Le Moigne in [35] utilized shearlet-based features to register multimodal RS data under affine transformation. This approach was reported to register optical to shaded DEM images with mean RMSE of about 3.5 pixels, but for the optical-to-DEM case, it was not successful. In turn, for the optical-to-DEM registration problem, our RAE method provides control points present at different land covers with an RMSE of about $0.62, \ldots, 0.75$ pixels. Similarly to the optical-to-radar registration case, subpixel registration accuracy at the global scale is achieved.

The last DEM-to-radar registration scenario is of even higher level of complexity, and to the best of our knowledge, no successful registration cases were reported in the literature. Our RAE method can handle this registration case with the RMSE of found control fragments of about $0.67, \ldots, 0.85$ pixels.

III. Image Registration Method Utilizing Potential Registration Accuracy at Local and Global Scales

This section formally introduces the proposed RAE registration method and discusses its main features. We start by recalling the constraints that can be considered for modeling a geometrical transformation at both local and global scales for RS applications. Then, we define and state the problem of interest. We describe the search for putative correspondences using NCC similarity measure, assignment of potential registration accuracy to each found PC, outlier detection, and estimation of registration parameters at the global scale. Then, registration accuracy at the global scale is derived. Finally, computational complexity of the RAE method is analyzed in detail.

A. Constraints on Geometrical Transformation Model Parameters for RS Applications

Geometrical errors encountered in RS applications have their own specificities. Due to remote sensors' linearity (initially assured in their design), affine transformation hypothesis can be accepted as a first-order approximation of geometrical transformation between two RS images: the errors nonlinearity caused by sensor deviations from ideal linear array camera model, Earth curvature, etc., are limited as compared to the main linear part of the sensor behavior. A preponderant error source is the translation error due to errors in satellite positioning and orientation. Scale and orientation errors are significantly smaller and can be locally treated as spatially varying translation errors (drift errors). Nonlinear distortions can be most of the time neglected at the local level [15]. As a result, at the local scale, the transformation between the reference and template images reduces itself to the rotation-isomorphic scaling-translation (RST) model with smaller estimation errors w.r.t. the rotation and scaling and outweighing translation estimation errors [33]. Such assumptions can be considered typical for the RS image registration field [31], [36].

Initial geopositioning of RS images is generally based on a rigorous sensor orbital model. The registration error provided with this method (no ground control points available) is called direct geopositioning error, and it defines the initial search zone w.r.t. the translation components. The value of the direct geopositioning error reduces gradually with advances in technology. However, this process is accompanied by improvement of sensor spatial resolution [5]. For example, for the OLI sensor of the Landsat 8 satellite, direct geopositioning error (CE90 or 2 sigma) is 65 m. Expressed in pixels (15-m spatial resolution for PAN band), this gives the error of 6.5 pixels (3 sigma). Direct geopositioning error is 56.56 for QuickBird/QuickBird-2, 16.50 for IKONOS, 15 for RADARSAT-2, and 3.81 pixels for RapidEye satellites (all values mentioned here were calculated using sensor specifications). Overall, direct geopositioning error expressed in pixels is more or less a stable value with a range of variation from 3 to about 60 pixels (3 sigma; from a few pixels to a few tens of pixels according to [35]).

In this paper, we assume that images to register are orthorectified (in the experimental part, we use preliminary orthorectified images or perform relief correction using DEMs for study areas). Under these conditions, we assume a polynomial model as a geometrical transformation model at the global scale. The proposed RAE method is applicable for arbitrary model order. In the experimental part of this paper, we provide results for first- and second-order models. A more accurate geometrical modeling is, however, possible using the rational polynomial coefficient (RPC) model. We prefer here to consider a simpler polynomial model to better illustrate the proposed method advantages, leaving the RPC model for future study.

At the local scale, the RST model is chosen as justified previously. The initial values of the model parameters are estimated from the metadata provided with the reference and template images (box corner coordinates). The search zone w.r.t. the translation parameters is set based on the direct geopositioning error of the reference and template image platforms (the procedure is described in the next section). Designing the RAE method, we pursue the goal to assure robust registration for quite large initial translation errors up to ± 100 pixels in order to cover direct geopositioning errors commonly met in practice.

B. Definitions, Constraints, and Problem Statement

The two images to register are the reference image (RI) with m_{RI} rows and n_{RI} columns and the template image (TI) with m_{TI} rows and n_{TI} columns. The coordinates of a pixel in the ith row and jth column of the RI/TI images will be referred to as $\mathbf{y} = (i_{\mathrm{RI}}, j_{\mathrm{RI}})^T$ and $\mathbf{x} = (i_{\mathrm{TI}}, j_{\mathrm{TI}})^T$, respectively.

The reference and template images are initially coarsely registered based on longitude and latitude of all four corners embedded in the imagery file metadata. The initial registration is thus described by the affine transformation

$$\mathbf{y}_{\mathrm{Init}} = \mathbf{A}_{\mathrm{Init}}\mathbf{x} + \mathbf{d}_{\mathrm{Init}}. \tag{1}$$

An affine transformation matrix $\mathbf{A}$ can be approximated by RST transformation matrix $\mathbf{A}_{\mathrm{RST}} = \Delta r\mathbf{R}$, where Δr is a scaling factor and $\mathbf{R} = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}$ is a rotation matrix through an angle α. The optimal values of Δr and $\mathbf{R}$ are found according to the method described in [23] (see Section C "Rigid Feature Matching" and the references therein): $\Delta r = \sqrt{|\mathbf{A}|}$ and $\mathbf{R} = \mathbf{U}\mathbf{V}^T$, where $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ is the SVD decomposition of $\mathbf{A}$ and $\mathbf{A}$ does not include reflections. The initial value for the rotation angle α_{Init} and the scaling factor Δr_{Init} between RI and TI images are estimated by applying this decomposition to $\mathbf{A}_{\mathrm{Init}}$.

The initial registration error with respect to both spatial coordinates is bounded previously by $d_{\max 0}$: $\sqrt{(i_{\mathrm{RI.Init}} - i_{\mathrm{RI0}})^2 + (j_{\mathrm{RI.Init}} - j_{\mathrm{RI0}})^2} \le d_{\max 0}$, where $\mathbf{y}_0 = (i_{\mathrm{RI0}}, j_{\mathrm{RI0}})$ are the coordinates of the true correspondence and $\mathbf{y}_{\mathrm{Init}} = (i_{\mathrm{RI.Init}}, j_{\mathrm{RI.Init}})$ is given by (1). The value of $d_{\max 0}$ can be set based on direct geopositioning error SDs $\sigma_{\mathrm{g.RI}}$ and $\sigma_{\mathrm{g.TI}}$ of the reference and template images, respectively: $d_{\max 0} = 3\sqrt{\sigma^2_{\mathrm{g.RI}} + \sigma^2_{\mathrm{g.TI}}\Delta r^2_{\mathrm{Init}}}$, where $d_{\max 0}$ and $\sigma_{\mathrm{g.RI}}$ are given in RI pixels and $\sigma_{\mathrm{g.TI}}$ is given in TI pixels. The values of $\sigma_{\mathrm{g.RI}}$ and $\sigma_{\mathrm{g.TI}}$ can be found in specifications of sensors which the RI and TI images were acquired with.

As it has been introduced previously, we consider the geometrical transformation model at the global scale in the polynomial form of degree n

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{c}_1, \mathbf{c}_2) \tag{2}$$

where $\mathbf{g}(\mathbf{x}, \mathbf{c}_1, \mathbf{c}_2) = (g(\mathbf{x}, \mathbf{c}_1), g(\mathbf{x}, \mathbf{c}_2))$, $g(\mathbf{x}, \mathbf{c}) = g(i_{\mathrm{TI}}, j_{\mathrm{TI}}, \mathbf{c}) = \mathbf{c} \cdot (1, i_{\mathrm{TI}}, j_{\mathrm{TI}}, \ldots, i^{k_1}_{\mathrm{TI}} j^{k_2}_{\mathrm{TI}}, \ldots, i^n_{\mathrm{TI}}, j^n_{\mathrm{TI}})$, $0 \le k_1 + k_2 \le n$, $\mathbf{c}$ is $n_c = (n+2)(n+1)/2$ column vector of coefficients, and $\mathbf{c}_1$ and $\mathbf{c}_2$ define a transformation with respect to horizontal and vertical directions. Initially, $\mathbf{c}_1 = (\mathbf{d}_{\mathrm{Init}}(1), \mathbf{A}_{\mathrm{Init}}(1,1), \mathbf{A}_{\mathrm{Init}}(1,2))$ and $\mathbf{c}_2 = (\mathbf{d}_{\mathrm{Init}}(2), \mathbf{A}_{\mathrm{Init}}(2,1), \mathbf{A}_{\mathrm{Init}}(2,2))$ for $n = 1$; $\mathbf{c}_1 = (\mathbf{d}_{\mathrm{Init}}(1), \mathbf{A}_{\mathrm{Init}}(1,1), \mathbf{A}_{\mathrm{Init}}(1,2), 0, \ldots, 0)$ and $\mathbf{c}_2 = (\mathbf{d}_{\mathrm{Init}}(2), \mathbf{A}_{\mathrm{Init}}(2,1), \mathbf{A}_{\mathrm{Init}}(2,2), 0, \ldots, 0)$ for $n > 1$.

For both registered images, relief influence is taken into account by introducing systematic correction factors $\Delta i(i, j, H(i,j))$ and $\Delta j(i, j, H(i,j))$, compensating the observed shift at point (i, j) due to relief with height $H(i, j)$ (obtained from DEM for the study area).

The $\mathrm{CRLB}_{\mathrm{fBm}}$ bound is currently restricted to the RST model; therefore, we should constraint transformation (2) to be well approximated by the RST model at the local scale (about ± 10 pixels for the proposed method). This constraint is natural for RS sensors as it has been discussed in Section III-A. In order to impose it mathematically, model (2) is first approximated by an affine transformation $\mathbf{y} = \mathbf{A}_{\mathrm{linear}}\mathbf{x} + \mathbf{d}_{\mathrm{linear}}$ in the neighborhood of a point $\mathbf{x}_0 = (i_{\mathrm{TI0}}, j_{\mathrm{TI0}})$ using first-order terms of its Taylor series expansion. We checked that, for all test cases considered in the experimental part of this paper, nonlinear

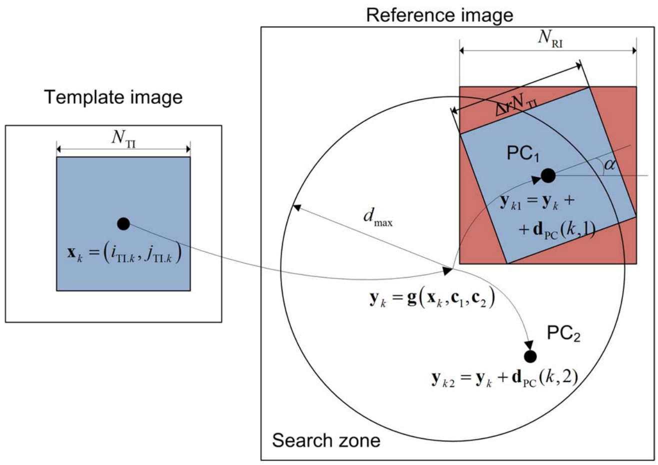


Fig. 1. Illustration of search of putative correspondences between the template and reference images.

effects at the distance of 10 pixels are of the magnitude order of 10^{-2} pixel, which can be reasonably neglected.

Then, the affine transformation matrix $\mathbf{A}_{\text{linear}}$ is further approximated by the RST transformation matrix $\mathbf{A}_{\text{RST}}$. In the neighborhood of $\mathbf{x}_0$, the uncompensated difference between RI and TI fragments at the RI coordinate system can be measured through matrix difference $\mathbf{A}_{\text{linear}} - \mathbf{A}_{\text{RST}}$. The maximum uncompensated error at the one pixel distance, defined as d_1, is obtained as the square root of the maximum eigenvalue of matrix $(\mathbf{A}_{\text{linear}} - \mathbf{A}_{\text{RST}})(\mathbf{A}_{\text{linear}} - \mathbf{A}_{\text{RST}})^T$. For a RI fragment of ± 10 pixel size, this error increases to $d_{10} = 10 d_1$. For image pairs considered in this paper, we found that d_{10} varies from 0.03 to 0.3 pixels. This error can be neglected as it is tolerated by area-based methods, such as the NCC method in our case [37], [38].

With these definitions, the registration problem is formulated as an estimation problem of parameter vectors $\mathbf{c}_1$ and $\mathbf{c}_2$.

C. Search for Putative Correspondences

Let us tile the template image by nonoverlapping CFs of size $N_{\text{TI}} \times N_{\text{TI}}$ pixels and define the center coordinates of these CFs by $\mathbf{x}_k$, $k = 1, \ldots, N_{\text{CF}}$, where N_{CF} is the total number of CFs.

Using the current estimate of transformation parameter vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ (see Fig. 1), the position of the center of the transformed template CF is predicted as $\mathbf{g}(\mathbf{x}_k, \mathbf{c}_1, \mathbf{c}_2)$. The true position could be anywhere inside the search zone—circle with radius $d_{\max}(k)$—centered at $\mathbf{g}(\mathbf{x}_k, \mathbf{c}_1, \mathbf{c}_2)$. Initially, the value of the radius is $d_{\max}(k) = d_{\max 0}$ for all k. Recall, that $\mathbf{c}_1$ and $\mathbf{c}_2$ are initialized based on the longitude and latitude of the four corners of the reference and template images as described in Section III-B.

We propose to perform the search of PCs within the search zone w.r.t. only the translation vector components in both directions, keeping the rotation angle and scaling factor values fixed at their initial values α_{Init} and Δr_{Init}. To justify this, let us evaluate the error d_{10} if the $\mathbf{A}_{\text{RST}}$ matrix is calculated based on the coarse-registration matrix $\mathbf{A}_{\text{Init}}$: $\mathbf{A}_{\text{RST.Init}} = \Delta r_{\text{Init}} \begin{pmatrix} \cos(\alpha_{\text{Init}}) & \sin(\alpha_{\text{Init}}) \\ -\sin(\alpha_{\text{Init}}) & \cos(\alpha_{\text{Init}}) \end{pmatrix}$. In this case, d_{10} varies from 0.05 to 0.7 pixels. This distortion is symmetrical w.r.t. the RI fragment center, and it will not cause bias of estimation of PC position. Its main effect will be a slightly reduced correlation between RI and TI fragments. At this stage of our research, we have neglected this error. Note that an iterative refinement of PC positions using more and more accurate RST model parameters obtained in the registration process can help in compensating this error.

We stress that the search w.r.t. translations does not mean that the RAE method cannot be applied to images with a higher degree of nonlinearity. In this case, additional search of PCs w.r.t. the rotation angle and scaling factor is to be considered at the expense of higher computational complexity of the PC search procedure. This does not also mean that we deal with global translation estimation because estimated translations are different for each CF and are estimated independently.

Accordingly, each template CF is projected into the reference image using initial rotation angle α_{Init} and the scaling factor Δr_{Init} and interpolated to the reference image grid so as to form a CF of size N_{RI} (pixels having no match at the template CF are filled with zero values and are no more used in subsequent processing). Similarity between a template CF and a reference CF is measured by NCC in this paper. Within the search zone, NCC may exhibit a multiextremal behavior, with one extremum relating to the true correspondence (if such a correspondence exists for a particular CF) and a number of local extrema relating to false correspondences. The pairs of reference and template CFs linked to each NCC local extremum will be later referred to as putative correspondences (PCs).

Putative correspondence search is implemented as a two-step procedure: detection of PCs and refinement of their position with subpixel accuracy. First, the NCC k_{RT} between the transformed template CF and all reference CFs in the search zone is calculated on the grid with the lag equal to 0.5 pixel in both directions. This particular value for the lag was found experimentally as a good compromise between computational complexity (that increases with lag decrease) and probability of finding local NCC maxima (that decreases with lag increase). Then, all local NCC extrema with $|k_{\text{RT}}| > 0.25$ are found. For a kth template CF, a pth PC is described by the pair of coordinates $\mathbf{x}_k$ and $\mathbf{y}_{kp}$ and the NCC value $k_{\text{RT}.kp}$. Second, the positions of found PCs are refined with subpixel accuracy. This is done using template image intensity interpolation, allowing calculation of NCC for arbitrary RST parameters [39]. The NCC value is then maximized w.r.t. $\mathbf{y}_{kp}$.

D. Proposed Method General Structure

Modern sensors tend to acquire images of a very large size. Hence, the number of control fragments and putative correspondences can be huge. Therefore, we propose an iterative registration scheme called RAE (Fig. 2) with successive refinement of estimates of geometrical transformation parameters and removal of false PCs.

At the initialization stage, the list of PCs is populated using NCC values. All found PCs are sorted according to the $|k_{\text{RT}}|$ value. If they lie within the current search zone, they are considered as active. Then, each iteration proceeds as follows: for the first unprocessed active PC, potential registration accuracy

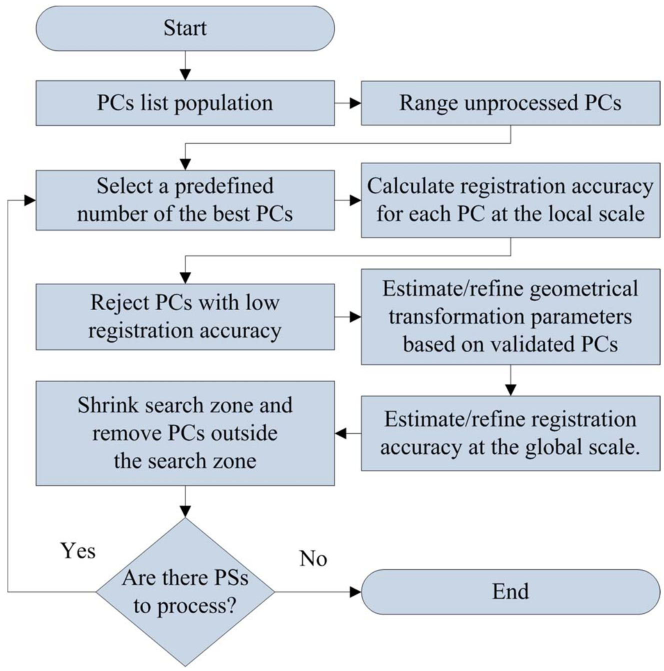


Fig. 2. Flowchart of the proposed RAE registration method.

between the reference and template CFs (in the form of CRLB on translation estimation error) is estimated. If the registration accuracy is high enough (translation estimation error is low enough), this PC is validated for being involved in the geometrical transformation parameter estimation; otherwise, it is discarded. When a predefined number of validated PCs are found, the estimates of geometrical transform parameter vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ are refined along with their estimation accuracy. Refined geometrical transform allows reducing the search zone, and it shrinks iteratively the list of active PCs. Thus, the computational complexity of each next iteration significantly reduces. Iterations continue until no active PC is left.

The core feature of the proposed RAE registration algorithm is its ability to quantify the registration accuracy. This is done in two ways: first, at local scale, by systematically estimating registration accuracy for each PC and, second, at global scale, by quantifying estimation accuracy of the transformation parameters $\mathbf{c}_1$ and $\mathbf{c}_2$. Let us next address these stages more in detail.

E. Registration Accuracy at the Local Scale

At the local scale, for a given PC, we distinguish two factors influencing the registration error. The first one is straightforwardly related to the inherent structure of the registered reference and template fragments and establishes the lowest possible or potential registration accuracy. This factor is described by the potential translation estimation error SD $\sigma_{\mathrm{PC.LB}}$. The value of $\sigma_{\mathrm{PC.LB}}$ is different for each PC and does not depend on the used registration algorithm. The second factor is the statistical efficiency [40] of the used registration method denoted as e_{est}. Recall that $0 \leq e_{\mathrm{est}} \leq 1$. When $e_{\mathrm{est}} = 1$, the estimator is called efficient, and it provides the highest possible registration accuracy equal to $\sigma_{\mathrm{PC.LB}}$. In practice, the registration accuracy is lower in some degree due to the finite sample size or implementation peculiarities intended for speeding-up. Naturally, higher values of e_{est} are preferable. Although the value e_{est} may depend on the PC (the registration method could be more effective for one texture and less effective for another one), such subtle effects are outside the scope of this paper. Following the previous study of the NCC estimator behavior in [20], [26], and [27], the value of e_{est} is set, in all experiments, equal to 0.1.

Taking into account these two factors, for each PC, the registration accuracy is calculated as

$$\sigma_{\mathrm{PC}} = \frac{\sigma_{\mathrm{PC.LB}}}{\sqrt{e_{\mathrm{est}}}}. \tag{3}$$

To estimate $\sigma_{\mathrm{PC.LB}}$, we rely on our previous works [20], [26], [27], where the CRLB on the RST parameter estimation error was obtained and studied. For each PC, calculation of $\sigma_{\mathrm{PC.LB}}$ involves the following stages. At the initialization stage, noise parameters are defined separately for the reference and template fragments. We use the complex model applicable to the RS images of different types: spatially correlated normally distributed noise with signal-dependent variance [27]. Noise model parameters are estimated in advance according to a procedure specified in Section IV. At the first stage, PC texture parameter vector $\boldsymbol{\theta}_{\mathrm{PC.texture}} = (\sigma_{x.\mathrm{RI}}, \sigma_{x.\mathrm{TI}}, k_{\mathrm{RT}}, H)$ is estimated. Here, $\sigma_{x.\mathrm{RI}}, \sigma_{x.\mathrm{TI}}$ define the SD of the texture increments on unit distance for the reference and template CFs, and Hurst exponent $H \in (0, 1)$ characterizes texture roughness (a value less than 0.5 corresponds to rough and greater than 0.5—to smooth textures). The estimation of the value of k_{RT} is improved at this stage owing to the more relevant noise model used. At the second stage, the estimated texture parameter vector is appended by translation estimates $\mathbf{d}_{\mathrm{PC}} = \mathbf{y}_{kp} - \mathbf{g}(\mathbf{x}_k, \mathbf{c}_1, \mathbf{c}_2)$ to form 6 by 1 vector $\boldsymbol{\theta}_{\mathrm{PC}} = (\hat{\sigma}_{x.\mathrm{RI}}, \hat{\sigma}_{x.\mathrm{TI}}, \hat{k}_{\mathrm{RT}}, \hat{H}, \mathbf{d}_{\mathrm{PC}}(1), \mathbf{d}_{\mathrm{PC}}(2))$. The rotation angle and scaling factor values are the same as those for the PC search stage, and they are not included in $\boldsymbol{\theta}_{\mathrm{PC}}$ as being considered fixed. The corresponding Fisher information matrix (FIM) $\mathbf{I}_{\boldsymbol{\theta}_{\mathrm{PC}}}$ on the $\boldsymbol{\theta}_{\mathrm{PC}}$ vector is then evaluated and inverted to obtain CRLB matrix $\mathbf{C}_{\boldsymbol{\theta}_{\mathrm{PC}}} = \mathbf{I}^{-1}_{\boldsymbol{\theta}_{\mathrm{PC}}}$. The submatrix $\mathbf{C}_{\boldsymbol{\theta}_{\mathrm{PC}}}(5:6, 5:6)$ characterizes the potential translation estimation accuracy or PC registration accuracy. As it was shown in [20], the estimation errors of the two translation components might be correlated and have different SDs. However, these effects are not essential, and $\mathbf{C}_{\boldsymbol{\theta}_{\mathrm{PC}}}(5:6, 5:6)$ can be well approximated as a scalar multiple of an identity matrix. Therefore, $\sigma_{\mathrm{PC.LB}}$ is finally obtained as $\sigma_{\mathrm{PC.LB}} = \sqrt{(\mathbf{C}_{\boldsymbol{\theta}_{\mathrm{PC}}}(5,5) + \mathbf{C}_{\boldsymbol{\theta}_{\mathrm{PC}}}(6,6))/2}$.

For each kpth PC, the estimated value $\sigma_{\mathrm{PC.LB}.kp}$ is compared to the predefined threshold $\sigma_{\mathrm{PC.LB.\max}}$. If $\sigma_{\mathrm{PC.LB}.kp} < \sigma_{\mathrm{PC.LB.\max}}$, this PC is considered suitable for registration with respective registration accuracy $\sigma_{\mathrm{PC}.kp} = \sigma_{\mathrm{PC.LB}.kp}/\sqrt{e_{\mathrm{est}}}$. Such a PC is called CRLB-validated PC and denoted further as vPC. The number of vPCs for a kth CF is denoted as $n_{\mathrm{vPC}}(k)$; their total number is $n_{\mathrm{vPC}} = \sum_{k=1}^{N_{\mathrm{CF}}} n_{\mathrm{vPC}}(k)$. In the experimental part of this paper, we set $\sigma_{\mathrm{PC.LB.\max}} = 0.35$ pixel for all test cases. This choice is based on the analysis performed in [27], where it was shown that registration errors using NCC similarity measure were linearly related to $\sigma_{\mathrm{PC.LB}}$ up to values $\sigma_{\mathrm{PC.LB}} < 0.4$. For higher values of $\sigma_{\mathrm{PC.LB}}$ (related either to higher noise level or lower correlation between reference and template images), NCC registration becomes unreliable.

F. Geometrical Transformation Parameter Estimation. Registration Accuracy at Global Scale

To detect true correspondences and to estimate more reliably the transformation parameters, we follow the likelihood approach developed in [22]–[24]. Let us first describe the proposed solution and subsequently discuss its distinctive features and advantages.

At the global scale, we process only CFs with at least one CRLB-validated PC (CF with $n_{\mathrm{vPC}}(k) > 0$). Their number is N_{vCF}. For notation simplicity, we use the same indexes k and p for CRLB-validated CFs and PCs.

For each kth CF, we define a binary-valued vector $\mathbf{z}_k$ with $n_{\mathrm{vPC}}(k)$ elements. Unity in pth position of the vector $\mathbf{z}_k$ indicates that the pth PC is an inlier. In this case, the probability density function (pdf) of observing $\mathbf{y}_{\mathrm{kp}}$ value is $N(\mathbf{y}_{\mathrm{kp}} - \mathbf{g}(\mathbf{x}_k, \mathbf{c}_1, \mathbf{c}_2), \mathbf{\Sigma}_{\mathrm{kp}})$, where $\mathbf{\Sigma}_{\mathrm{kp}} = \begin{pmatrix} \sigma^2_{\mathrm{PC}.kp} & 0 \\ 0 & \sigma^2_{\mathrm{PC}.kp} \end{pmatrix}$ is the covariation matrix of registration error. Zeros in a pth position of the vector $\mathbf{z}_k$ indicate that the pth PC is an outlier. In this case, $\mathbf{y}_{\mathrm{kp}}$ is distributed uniformly within the respective search zone with pdf $1/S_{\mathrm{SearchZone}}(k)$, where $S_{\mathrm{SearchZone}} = \pi d^2_{\max}(k)$ is the area of the search zone. The probability of a pth PC to be an inlier is defined as $P(z_k(p) = 1) = P^{\mathrm{in}}_{\mathrm{kp}}$. Correspondingly, $P(z_k(p) = 0) = 1 - P^{\mathrm{in}}_{\mathrm{kp}}$. Since only one among PCs related to the same CF can be an inlier, only one element of $\mathbf{z}_k$ or none can be unity.

Events when the first, second,$\ldots, k$th$, \ldots,$ or $n_{\mathrm{vPC}}(k)$ PC is a true correspondence are disjoint and equiprobable with *a priori* probability $P^{\mathrm{in}}_{\mathrm{PC}}$. The value of $P^{\mathrm{in}}_{\mathrm{PC}}$ is derived using *a priori* probability $P^{\mathrm{in}}_{\mathrm{CF}}$ to find a true correspondence for a given CF as $P^{\mathrm{in}}_{\mathrm{PC}}(k) = 1 - (1 - P^{\mathrm{in}}_{\mathrm{CF}})^{1/n_{\mathrm{vPC}}(k)}$. Note that $P^{\mathrm{in}}_{\mathrm{CF}}$ is considered as a fixed value for all CFs, while $P^{\mathrm{in}}_{\mathrm{PC}}$ varies from CF to CF depending on $n_{\mathrm{vPC}}(k)$. The pdf of PCs related to a single CF conditional on $\boldsymbol{\theta} = (P^{\mathrm{in}}_{\mathrm{CF}}, \mathbf{c}_1, \mathbf{c}_2)$ is

$$\begin{aligned} f&\left(\mathbf{y}_{k1}, \mathbf{y}_{k2}, \ldots, \mathbf{y}_{kn_{\mathrm{vPC}}(k)}/\mathbf{x}_k, \boldsymbol{\theta}\right) \\ &= \sum_{\mathbf{z}} f\left(\mathbf{y}_{k1}, \mathbf{y}_{k2}, \ldots, \mathbf{y}_{kn_{\mathrm{vPC}}(k)}, \mathbf{z}/\mathbf{x}_k, \boldsymbol{\theta}\right) \\ &= P^{\mathrm{in}}_{\mathrm{PC}}(k) \sum_{p=1}^{n_{\mathrm{vPC}}(k)} N\left(\mathbf{y}_k - \mathbf{g}(\mathbf{x}_k, \mathbf{c}_1, \mathbf{c}_2), \mathbf{\Sigma}_{\mathrm{kp}}\right) \\ &\quad + \left(1 - P^{\mathrm{in}}_{\mathrm{CF}}\right)/S_{\mathrm{SearchZone}}(k). \end{aligned} \tag{4}$$

Using the independence between CFs, the pdf for all CFs is obtained as

$$\begin{aligned} f&\left(\mathbf{y}_{11}, \mathbf{y}_{12}, \ldots, \mathbf{y}_{N_{\mathrm{vCF}} n_{\mathrm{vPC}}(N_{\mathrm{vCF}})}/\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_{\mathrm{vCF}}}, \boldsymbol{\theta}\right) \\ &= \prod_{k=1}^{N_{\mathrm{vCF}}} f\left(\mathbf{y}_{k1}, \mathbf{y}_{k2}, \ldots, \mathbf{y}_{kn_{\mathrm{vPC}}(k)}/\mathbf{x}_k, \boldsymbol{\theta}\right). \end{aligned} \tag{5}$$

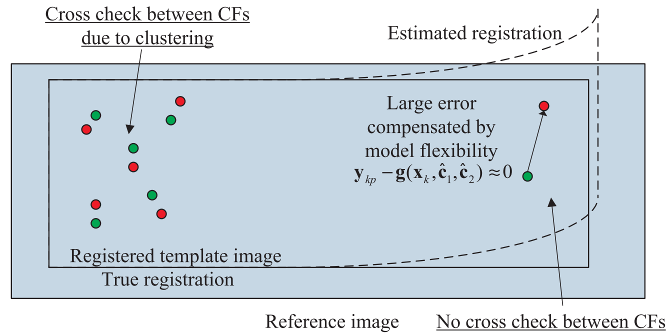


Fig. 3. Illustration of CF clustering effect.

The pdf (5) has one drawback that we will illustrate with an example. Fig. 3 shows a specific composition of CFs where all but one are in one cluster while one CF stands aside. For the clustered CFs, the estimates $\hat{\mathbf{c}}_1$ and $\hat{\mathbf{c}}_2$ will average the influence of all CFs in the cluster: no CF has a preponderant role. For the isolated CF, flexibility of the transformation $\mathbf{g}(\mathbf{x}, \mathbf{c}_1, \mathbf{c}_2)$ allows approximating a large difference between $\mathbf{y}_{\mathrm{kp}}$ and $\mathbf{g}(\mathbf{x}_k, \mathbf{c}_1, \mathbf{c}_2)$, making $\mathbf{y}_{\mathrm{kp}} - \mathbf{g}(\mathbf{x}_k, \hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2) \approx 0$ at the same time not altering the cluster registration. Therefore, the probability of the isolated PC to be an inlier defined by pdf $N(\mathbf{y}_{\mathrm{kp}} - \mathbf{g}(\mathbf{x}_k, \mathbf{c}_1, \mathbf{c}_2), \mathbf{\Sigma}_{\mathrm{kp}})$ will be high irrespective of whether this point is an inlier or an outlier. This effect is called leverage in statistics (particularly in regression analysis) [28], where isolated points lacking neighboring observations are called high-leverage points. Notice that researchers in the image registration area are also well aware of this problem. The benefit that can be drawn from control fragment clustering has been systematically outlined and used for improving the registration accuracy [41], [42].

We propose to solve this problem directly by introducing a modification to pdf (4) according to the LOOCV method [12]. The idea is to define the pdf of observing the inlying PC $\mathbf{y}_{\mathrm{kp}}$ as $N(\mathbf{y}_{\mathrm{kp}} - \mathbf{g}(\mathbf{x}_k, \mathbf{c}_{1k}, \mathbf{c}_{2k}), \mathbf{\Sigma}_{\mathrm{kp}})$, where $\mathbf{c}_{1k}$ and $\mathbf{c}_{2k}$ are the transformation parameters defined by all PCs excluding the ones related to the kth CF. In this manner, $\mathbf{g}(\mathbf{x}_k, \hat{\mathbf{c}}_{1k}, \hat{\mathbf{c}}_{2k})$ is a prediction of $\mathbf{y}_{\mathrm{kp}}$ based on PCs related to the neighboring CFs. Such a prediction has a low probability to be accurate for outliers and excludes convergence of $\mathbf{y}_{\mathrm{kp}} - \mathbf{g}(\mathbf{x}_k, \hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ to small values for isolated CFs. In statistical terms, our score function assures absence of high-leverage points (or points with high Cook's distance [28] that is also based on the LOOCV method) among inliers.

The corrected pdf of PCs related to a single CF conditional on $\boldsymbol{\theta}_k = (P^{\mathrm{in}}_{\mathrm{CF}}, \mathbf{c}_{1k}, \mathbf{c}_{2k})$ becomes

$$\begin{aligned} f&\left(\mathbf{y}_{k1}, \mathbf{y}_{k2}, \ldots, \mathbf{y}_{kn_{\mathrm{vPC}}(k)}/\mathbf{x}_k, \boldsymbol{\theta}_k\right) \\ &= P^{\mathrm{in}}_{\mathrm{PC}}(k) \sum_{p=1}^{n_{\mathrm{vPC}}(k)} N\left(\mathbf{y}_k - \mathbf{g}(\mathbf{x}_k, \mathbf{c}_{1k}, \mathbf{c}_{2k}), \mathbf{\Sigma}_{\mathrm{kp}}\right) \\ &\quad + \frac{\left(1 - P^{\mathrm{in}}_{\mathrm{CF}}\right)}{S_{\mathrm{SearchZone}}(k)} \end{aligned} \tag{6}$$

and the pdf for all CFs is now given by

$$f\left(\mathbf{y}_{11},\mathbf{y}_{12},\ldots,\mathbf{y}_{N_{\mathrm{vCF}}n_{\mathrm{vPC}}(N_{\mathrm{vCF}})}/\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{N_{\mathrm{vCF}}},\boldsymbol{\theta}\right) = \prod_{k=1}^{N_{\mathrm{vCF}}} f\left(\mathbf{y}_{k1},\mathbf{y}_{k2},\ldots,\mathbf{y}_{kn_{\mathrm{vPC}}(k)}/\mathbf{x}_k,\boldsymbol{\theta}_k\right) \quad (7)$$

where $\boldsymbol{\theta} = (P_{\mathrm{CF}}^{\mathrm{in}}, \mathbf{c}_{11}, \mathbf{c}_{21}, \ldots, \mathbf{c}_{1N_{\mathrm{vCF}}}, \mathbf{c}_{2N_{\mathrm{vCF}}})$. Note that the parameter vector $\boldsymbol{\theta}$ now includes the pairs $\mathbf{c}_{1k}$ and $\mathbf{c}_{2k}$ for all N_{vCF} control fragments.

The problem (7) is solved via the EM algorithm. The complete-data log-likelihood takes the following form:

$$\begin{aligned} Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}) = \sum_{k=1}^{N_{\mathrm{vCF}}} \sum_{p=1}^{n_{\mathrm{vPC}}(k)} & \times\Bigg[-\frac{P_{\mathrm{kp}}^{\mathrm{in}}(\hat{\boldsymbol{\theta}})}{2\sigma_{\mathrm{PC}.kp}^2}\left\|\mathbf{y}_{\mathrm{kp}} - \mathbf{g}(\mathbf{x}_k,\mathbf{c}_{1k},\mathbf{c}_{2k})\right\|^2 - \\ & - \ln\left(\sigma_{\mathrm{PC}.kp}^2\right) P_{\mathrm{kp}}^{\mathrm{in}}(\hat{\boldsymbol{\theta}}) + \ln\left(P_{\mathrm{PC}}^{\mathrm{in}}(k)\right) P_{\mathrm{kp}}^{\mathrm{in}}(\hat{\boldsymbol{\theta}}) \\ & + \ln\left(1 - P_{\mathrm{CF}}^{\mathrm{in}}\right)\left(1 - P_{\mathrm{kp}}^{\mathrm{in}}(\hat{\boldsymbol{\theta}})\right) \\ & + \ln\left(\frac{1}{S_{\mathrm{SearchZone}}(k)}\right)\Bigg]. \end{aligned} \quad (8)$$

At the E-step of the EM method, the probabilities $P_{\mathrm{kp}}^{\mathrm{in}}(\hat{\boldsymbol{\theta}})$, i.e., *a posteriori* probability for each correspondence to be an inlier, are estimated. Using Bayes' rule, we get

$$\hat{P}_{\mathrm{kp}}^{\mathrm{in}} = \frac{P_{\mathrm{PC}}^{\mathrm{in}}(k) N\left(\mathbf{y}_{\mathrm{kp}} - \mathbf{g}(\mathbf{x}_k,\hat{\mathbf{c}}_{1k},\hat{\mathbf{c}}_{2k}),\boldsymbol{\Sigma}_{\mathrm{kp}}\right)}{P_{\mathrm{PC}}^{\mathrm{in}}(k) N\left(\mathbf{y}_{\mathrm{kp}} - \mathbf{g}(\mathbf{x}_k,\hat{\mathbf{c}}_{1k},\hat{\mathbf{c}}_{2k}),\boldsymbol{\Sigma}_{\mathrm{kp}}\right) + \frac{\left(1-\hat{P}_{\mathrm{CF}}^{\mathrm{in}}\right)}{S_{\mathrm{SearchZone}}(k)}}. \quad (9)$$

The M-step of the algorithm is responsible for maximization of $Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$. The average probability to find a true correspondence for a single CF is found as

$$\hat{P}_{\mathrm{CF}}^{\mathrm{in}} = \frac{1}{N_{\mathrm{vCF}}} \sum_{k=1}^{N_{\mathrm{vCF}}} \hat{P}_{\mathrm{CF}}^{\mathrm{in}}(k)$$

where $\hat{P}_{\mathrm{CF}}^{\mathrm{in}}(k) = [1 - \prod_{p=1}^{n_{\mathrm{vPC}}(k)}(1 - \hat{P}_{\mathrm{kp}}^{\mathrm{in}})]$ is the *a posteriori* probability that at least one of the PCs of the kth CF is an inlier. The inlying PCs are found at this stage according to the rule $\hat{P}_{\mathrm{kp}}^{\mathrm{in}} > P_{\mathrm{th}}^{\mathrm{in}}$, where $P_{\mathrm{th}}^{\mathrm{in}}$ is a threshold. If more than one PC belonging to the same CF is recognized as an inlier, the one with maximal $\hat{P}_{\mathrm{kp}}^{\mathrm{in}}$ is selected. CFs with inlying PC are called inlying CFs. In the experiments, we set $P_{\mathrm{th}}^{\mathrm{in}} = 0.9$ for all test cases.

Maximization of (8) with respect to $\mathbf{c}_{1k}$ and $\mathbf{c}_{2k}$ is a weighted polynomial regression problem with weights of each PC inversely proportional to $w_{\mathrm{kp}} = \sigma_{\mathrm{PC.kp}}^2/\hat{P}_{\mathrm{kp}}^{\mathrm{in}}$. The solution of this problem is found by equating the first derivatives of (8) to zero with respect to the elements of $\mathbf{c}_{1k}$ and $\mathbf{c}_{2k}$

$$\hat{\mathbf{c}}_{1k} = \mathbf{I}_k^{-1} \cdot \mathbf{b}_{1k}\ \hat{\mathbf{c}}_{2k} = \mathbf{I}_k^{-1} \cdot \mathbf{b}_{2k} \quad (10)$$

where $\mathbf{I}_k = \sum_{t\neq k}\sum_{p=1}^{n_{\mathrm{vPC}}(t)} w_{\mathrm{tp}}^{-1}\mathbf{e}_{\mathrm{tp}}^T\mathbf{e}_{\mathrm{tp}}$, $\mathbf{b}_{1k} = \sum_{t\neq k} i_{\mathrm{RI.t}} \sum_{p=1}^{n_{\mathrm{vPC}}(t)} w_{\mathrm{tp}}^{-1}\mathbf{e}_{\mathrm{tp}}^T$, $\mathbf{b}_{2k} = \sum_{t\neq k} j_{\mathrm{RI.t}} \sum_{p=1}^{n_{\mathrm{vPC}}(t)} w_{\mathrm{tp}}^{-1}\mathbf{e}_{\mathrm{tp}}^T$, and $\mathbf{e}_{\mathrm{tp}} = (1,\ldots,i_{\mathrm{TI.tp}}^{k_1} j_{\mathrm{TI.tp}}^{k_2},\ldots,i_{\mathrm{TI.tp}}^{n}, j_{\mathrm{TI.tp}}^{n})$. The first sum is over all inlying CFs except the kth CF. The maximization step fails if the number of inlying CFs is not large enough to solve linear equations in (10), i.e., less than $n_c + 1$.

The complexity of (10) for one CF is $O(n_{\mathrm{vPC}}) = O(N_{\mathrm{vCF}}) = O(N_{\mathrm{CF}})$; for all CFs—$O(N_{\mathrm{CF}}^2)$. Quadratic complexity is undesirable. Therefore, we propose to estimate $\mathbf{c}_{1k}$ and $\mathbf{c}_{2k}$ using a fixed number of inlying CFs instead of all CFs. We set this number to the minimum sufficient value $n_c + 1$. Using k-d trees [43], the complexity of search for a finite number of closest neighbors is $O(N_{\mathrm{CF}}\ln(N_{\mathrm{CF}}))$. As a result, the complexity of the global transformation parameter estimation procedure is linearithmic.

The final geometrical transformation parameters $\hat{\mathbf{c}}_1$ and $\hat{\mathbf{c}}_2$ are estimated after the EM algorithm convergence according to the same formula (10) but using all validated CFs. In this case, the matrix $\mathbf{I}_k$ is replaced by $\mathbf{I}$ in (10). The estimation errors of $\hat{\mathbf{c}}_1$ and $\hat{\mathbf{c}}_2$ are independent of each other and characterized by the same covariation matrix $\mathbf{R_c}$ equal to $\mathbf{I}^{-1}$ (as the $\mathbf{I}$ matrix is, in essence, a FIM).

The matrix $\mathbf{R_c}$ characterizes the registration accuracy of the proposed method at the global scale. Note that it is a function of registration accuracy at the local scale $\sigma_{\mathrm{PC}.kp}^2$. Given covariation matrix $\mathbf{R_c}$, the registration accuracy at point $(i_{\mathrm{TI}}, j_{\mathrm{TI}})$ at TI image is obtained as

$$\sigma_{\mathrm{reg}}(i_{\mathrm{TI}}, j_{\mathrm{TI}}) = \sqrt{\mathbf{e}(i_{\mathrm{TI}}, j_{\mathrm{TI}})\mathbf{R}_c\mathbf{e}(i_{\mathrm{TI}}, j_{\mathrm{TI}})^T},\ 0 \le k_1 + k_2 \le n$$

where $\mathbf{e}(i_{\mathrm{TI}}, j_{\mathrm{TI}}) = [1, i_{\mathrm{TI}}, j_{\mathrm{TI}}, \ldots, i_{\mathrm{TI}}^{k_1} j_{\mathrm{TI}}^{k_2}, \ldots, i_{\mathrm{TI}}^{n} j_{\mathrm{TI}}^{n}]$. Note that $\sigma_{\mathrm{reg}}(i_{\mathrm{TI}}, j_{\mathrm{TI}})$ is in RI pixels. The respective point at RI image is defined by estimated parameter vectors $\hat{\mathbf{c}}_1$ and $\hat{\mathbf{c}}_2$ according to model (2). Later, we use the notion $\sigma_{\mathrm{reg}}(k) = \sigma_{\mathrm{reg}}(i_{\mathrm{TI}.k}, j_{\mathrm{TI}.k})$ to describe the registration accuracy for the kth CF.

One drawback of the EM algorithm is that it does not assure finding the global maximum of a likelihood function [44]. It might converge to a local maximum depending on an initial guess for global transformation parameters. To increase the probability of finding the global maximum, we propose to use a multistart approach [45]. As discussed previously, RS sensors are highly linear (neglecting relief influence), and an affine transformation as an initial guess is a reasonable choice.

Multistart optimization runs as follows. Random triples of vPCs belonging to different CFs are selected without repetition. For each triple, affine transformation parameters are estimated. The found transformation is considered as valid if the maximum difference between initial and newly found transformations over the reference image area does not exceed $d_{\max 0}$. Each valid transformation is used as the initial guess for the EM algorithm. The algorithm stops when a fixed number of validated initial guesses are found. The solution that maximizes the complete-data log-likelihood is then chosen. We have found experimentally that ten starts are sufficient.

Let us summarize in the following the key differences between the proposed registration method and the method [23] published recently.

1) The multiple putative correspondences for each CF are taken into account.
2) The LOOCV approach is utilized, and the registration method is formulated purely as a likelihood maximization without introducing any additional empirical regularization terms like the local linear transformation term in [23]. Owing to this, the RAE method is less restrictive with respect to the geometrical transformation properties.
3) The registration accuracy for each putative correspondence (in the form of CRLB) is introduced into the likelihood function as additional *a priori* information. By doing this, we take into account the structural differences between PCs affecting registration accuracy.

The most important consequence of these modifications is that it becomes possible to obtain registration accuracy (variance) at the global scale. In the experimental part of this paper, we will demonstrate that the registration accuracy estimated in this manner is very accurate even when only a few correspondences are available.

G. Reduction of Computational Complexity

Processing of a single PC according to the proposed method is a computationally intensive task (although all processing stages except for the EM optimization stage are of linear complexity w.r.t the number of CFs). First of all, CRLB calculation involves operations with joint correlation matrix of reference and template image fragments. Second, the proposed method emphasizes registration accuracy; therefore, time-consuming area-based methods with subpixel registration accuracy are needed for PC search and registration (like NCC). To cope with the arising complexity problem, in the following, we propose an implementation of the RAE method allowing us to drastically reduce the number of processed PCs, making the registration complexity acceptable.

The basic idea under speeding-up the proposed method is that each newly found CRLB-verified PC can, in principle, contribute to reduce the uncertainty about global transformation parameters. The refined estimation of global transformation parameters, in turn, allows reducing search zone and removing PCs outside the shrunk search zone. Such an alternative scheme requires precise prediction of registration accuracy at each stage—this is the main feature of the proposed method.

The RAE method is detailed in Algorithm 1. It includes the population stage of PC list (2 and 3), PC processing stage (5 and 6), global transformation parameter refinement stage (8), and PC list truncation stage (9). Finding all PCs in advance may be unreasonable as some of them will be rejected later at the global transformation refinement stage. Therefore, we balance time between PC list population and CRLB calculation (stage 4). In this manner, the PC population list becomes asynchronous. PCs are processed starting from the one with the highest $|k_{\mathrm{RT.kp}}|$ value. At each tth iteration, the PC processing stage runs until a predefined number of CRLB-validated PC $n_{\mathrm{PCnew}}(t)$ are found. After this, the estimate of geometrical transformation parameters is refined, and the registration SD $\sigma_{\mathrm{reg}}(k)$ for each CF is calculated and used to shrink the search zone according to the expression $d_{\max}(k) = 6 \cdot \sigma_{\mathrm{reg}}(k) + 2$. Here, $d_{\max}(k)$ is the 6-sigma zone enlarged by 2 pixels to account for the width of NCC lobes (even for perfectly registered images, PCs identified by the NCC measure deviate more or less from the true correspondence). We used 6-sigma interval instead of 3-sigma interval to account for the experimental results indicating that $\sigma_{\mathrm{reg}}(k)$ underestimates real RAE method performance. Finally, the PC list is truncated by rejecting all PCs outsize the newly defined search zone.

Algorithm 1 RAE registration

Input: Reference and template images, product corner coordinates, direct geopositioning errors;
Output: Coefficients of polynomial geometrical transform model, covariation matrix of the coefficients estimation error;

1. Set $t = 1$, set initial value n_{PCnew} (1);
2. Randomly select one template CF among unprocessed ones;
3. Find PC for the selected CF and populate the PCs list;
4. Repeat 2, and 3 until the processing time exceeds time for calculating $\mathrm{CRLB}_{\mathrm{fBm}}$ for one PC. Skip the stages 2–3 if all CFs are processed (the list of PCs has been populated);
5. Select PC with the maximum value of $|k_{\mathrm{RT.kp}}|$;
6. Calculate CRLB $\sigma_{\mathrm{PC.kp}}$ on registration error;
7. Repeat the steps from 2 to 6 until new $n_{\mathrm{PCnew}}(t)$ CRLB-validated PCs are found;
8. Estimate the global transform parameters. Evaluate the registration accuracy $\sigma_{\mathrm{reg}}(k)$ of each CF;
9. Determine the new search zone for each CF as $d_{\max}(k) = \min(6 \cdot \sigma_{\mathrm{reg}}(k) + 2, d_{\max.0})$, remove PCs outside the newly calculated search zone;
10. Set $t = t + 1$. Calculate $n_{\mathrm{PCnew}}(t)$;
11. Repeat the steps from 2 to 10 until all PCs are processed.

As it was shown, the global transform parameter estimation procedure is of $O(\ln(N_{\mathrm{CF}})N_{\mathrm{CF}})$ complexity. The iterative nature of the proposed registration method suggests recalculation of the global transform parameters after each $n_{\mathrm{PCnew}}(t)$ processed CF and gradual reduction of the search zone with respect to spatial coordinates. Let us first consider a fixed step size $n_{\mathrm{PCnew}}(t) = n_{\mathrm{PCnew}}$ and define the expected number of iterations as $r_{\mathrm{const}} = n_{\mathrm{vPC}}/n_{\mathrm{PCnew}}$, where n_{vPC} is the total number of CRLB-validated PC. In this case, the complexity for determining the expected number of iterations $O(r_{\mathrm{const}}) = O(n_{\mathrm{vPC}}) = O(N_{\mathrm{CF}})$ and the total complexity of the global transform parameter estimation procedure exceed quadratic

$$
\begin{aligned}
&O\left[n_{\mathrm{PCnew}} \ln(n_{\mathrm{PCnew}}) + 2n_{\mathrm{PCnew}} \ln(2n_{\mathrm{PCnew}}) + \cdots\right.\\
&\quad \left.+ r_{\mathrm{const}} n_{\mathrm{PCnew}} \ln(r_{\mathrm{const}} n_{\mathrm{PCnew}})\right]\\
&\quad = O\left[\ln(n_{\mathrm{PCnew}}) n_{\mathrm{PCnew}}(1 + 2 + \cdots + r_{\mathrm{const}}) + n_{\mathrm{PCnew}}\right.\\
&\qquad\quad \left.\cdot\ (1 + 2\ln(2) + \cdots + r_{\mathrm{const}} \ln(r_{\mathrm{const}}))\right]\\
&\quad = O\left(r_{\mathrm{const}}^2 \ln(r_{\mathrm{const}})\right) = O\left(N_{\mathrm{SW}}^2 \log(N_{\mathrm{SW}})\right).
\end{aligned}
$$

TABLE I
CHARACTERISTICS OF THE TEST DATA SETS

Case	Image modality reference / template	Sensor/dataset	Acquisition date	Site Latitude/ Longitude, degrees	Spatial resolution, m	Scale	Initial registration error $d_{\max 0}$, pixels
1	Optical	Hyperion (band #25) EO1H1800252002116110KZ	26.04.2002	49.4339/32.0678	30.38	1	125
	Optical	Landsat8, OLI (band #1) LC81770252014065LGN00	06.03.2014	48.8497/31.6597	30	1	
2	Optical	The same Hyperion band as in case 1				1/2	130 (65 at scale 0.5)
	DEM	ASTER GDEM-2* ASTGTM2_N48-49E031-032	2009**	49/32	1 arc-second (≈30m at the equator)	1/2, 1/3 (vertical, (horizontal)	
3	Optical	Landsat8, OLI (band #8) LC81990262014363LGN00	29.12.2014	48.8666/2.3488	15	1/2	110 (55 at scale 0.5)
	Radar	SIR-C (HH polarization) pr41419_ldr_ceos	05.10.1994	48.9584/2.8732	12.5	1/2	
4	DEM	ASTER GDEM-2 ASTGTM2_N48-49E002-003	2009	49/3	1 arc-second	1/2, 1/3	60 (30 at scale 0.5)
	Radar	The same as in case 3				1/4	
5	Optical	Hyperion (band #155) EO1H2010262006218110PZ	06.08.2006	48.3892/-1.1613	30.38	1	45
	Optical	Landsat8, OLI (band #5) LC82010262013342LGN00	08.12.2013	48.8662/-0.7878	30	1	
6	Optical	The same Landsat8 band as in case 5				1/2	60 (30 at scale 0.5)
	DEM	ASTER GDEM-2 ASTGTM2_N47-48W001-002	2009	48/-1	1 arc-second	1/2, 1/3	
7	Radar	SIR-C (HH polarization) pr43020_ldr_ceos	05.10.1994	49.8250/36.7498	12.5	1/2	120 (60 at scale 0.5)
	Optical	Landsat8, OLI (band #8) LC82010262013342LGN00	16.02.2016	50.2810/36.9146	15	1/2	
8	Radar	The same as in case 7				1/5	60 (30 at scale 1/2)
	DEM	ASTER GDEM-2 ASTGTM2_N49-50E036-037	2009	50/ 37	1 arc-second	1/2, 1/3	

*ASTER GDEM is a product of METI and NASA ; hyphen in the tiles name indicates range of longitudes/latitudes
**ASTER GDEM-2 release date

Frequent recalculation of the global transform parameters is needed only at the beginning of the registration process when each newly found correspondence can significantly improve the overall registration accuracy. Therefore, as a second and more effective strategy, we propose linear increase of $n_{\mathrm{PCnew}}(t)$ step size: $n_{\mathrm{PCnew}}(t+1) = qn_{\mathrm{PCnew}}(t)$, $n_{\mathrm{PCnew}}(1) = 1$.

The last iteration index is $r_{\mathrm{linear}} = \ln(n_{\mathrm{vPC}})/\ln(q)$. Note that $O(r_{\mathrm{linear}}) = O(\ln(n_{\mathrm{vPC}})) = O(\ln(N_{\mathrm{CF}}))$ and $O(q^{r_{\mathrm{linear}}}) = O(n_{\mathrm{vPC}}) = O(N_{\mathrm{CF}})$. In this case, complexity of the global transform parameter estimation remains linearithmic

$$
\begin{aligned}
&O\left(q\ln(q) + q^2\ln(q^2) + \cdots + q^{r_{\mathrm{linear}}}\ln(q^{r_{\mathrm{linear}}})\right)\\
&= O\left(\ln(q)(q + 2q^2 + \cdots + r_{\mathrm{linear}}q^{r_{\mathrm{linear}}})\right)\\
&= O(r_{\mathrm{linear}}q^{r_{\mathrm{linear}}}) = O\left(N_{\mathrm{SW}}\ln(N_{\mathrm{SW}})\right).
\end{aligned}
$$

Thus, the overall linearithmic complexity of the RAE method w.r.t. the number of CFs is assured.

IV. EXPERIMENTAL PART

A. Test Data Description

Eight registration cases are considered to analyze the capabilities of the proposed RAE method: (1, 5) optical-to-optical, (2, 6) optical-to-DEM, (3, 7) optical-to-radar, and (4, 8) DEM-to-radar image pairs. A detailed description of each case is given in Table I. The first and fifth cases relate to monomodal registration, and others relate to complex multimodal registration. All cases correspond to multitemporal framework with differences in acquisition time from 7 to 22 years.

Test cases 1 and 2 share the same band of the reference Hyperion image, and test cases 3 and 4 exploit the same SIR-C template image (HH polarization channel). These images after registration are shown in Figs. 4 and 5. These images and data in Table I reveal the complexity for each test case. The test images for test cases $5, \ldots, 8$ are not shown to avoid a too lengthy paper. The used test images exhibit a wide variety of land covers: urban, rural, forest, agricultural, rivers, and snow cover. They have very scarce water coverage; no complex water–land boundaries are present, except for test case 6. It is well known that the presence of such boundaries facilitates registration and *vice versa*. Therefore, we prefer to deal with complicated practical situations.

Coarse registration of the test pairs of images was performed based on metadata information provided with each image: longitudes and latitudes of image corners. For each test case, reference affine transformation between RI and TI was obtained using about 25 manually found control points. Initial registration errors (see Table I) were measured by comparing coarse and manual registration. For test cases 1 and 2, the initial error in one direction (along-track of the reference Hyperion image) turned out to exceed significantly the assumed interval ± 100 pixels and reached values up to 300 pixels. We attribute this, at least, partly to the different reference geodetic data of Hyperion, Landsat 8, and ASTER GDEM data. To relax this

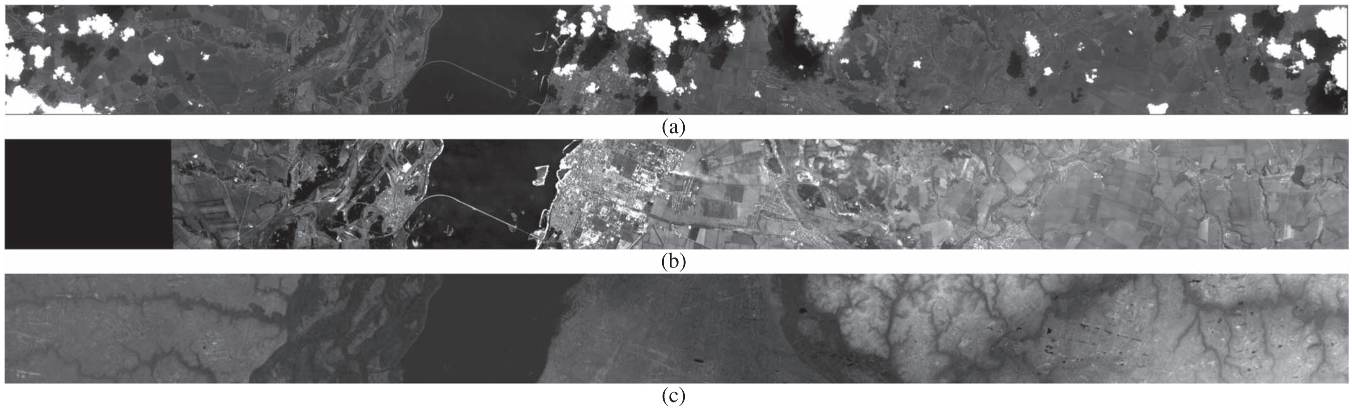


Fig. 4. Registered (a) Hyperion band #25, (b) Landsat 8 band B1, and (c) ASTER GDEM. Gray levels ranging from black to white cover the intensity ranges $1100, \ldots, 3800$ for Hyperion, $8500, \ldots, 9600$ for Landsat 8, and $50, \ldots, 250$ m for ASTER GDEM. Image size is 256 by 3129 pixels.

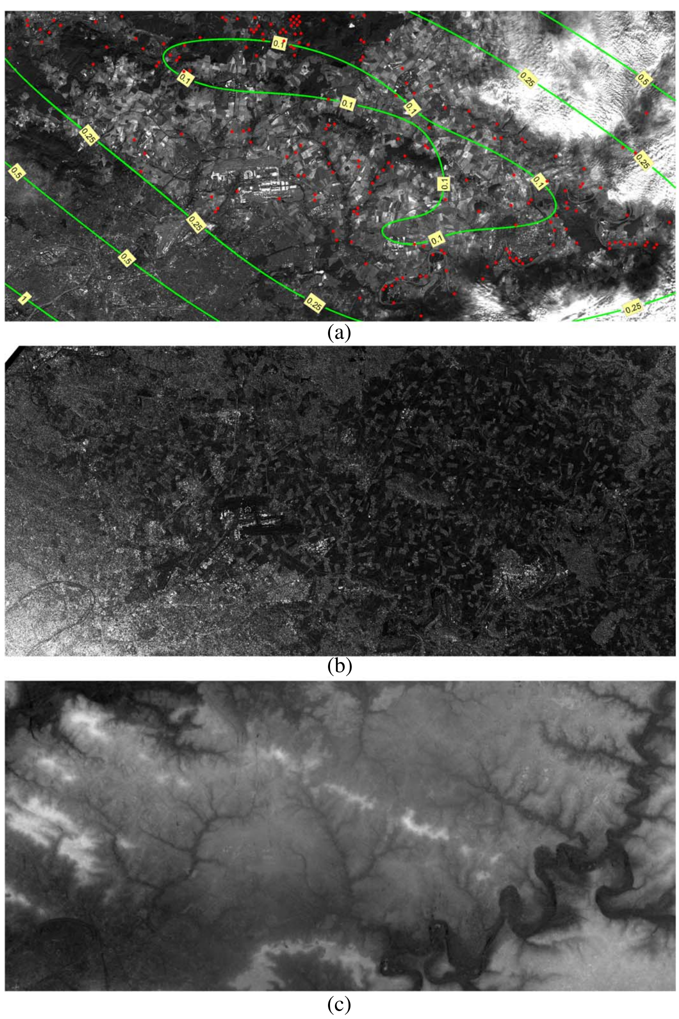


Fig. 5. Registered reference and template images for test cases 3 and 4: (a) Landsat 8, (b) SIR-C, and (c) ASTER GDEM images. Gray levels ranging from black to white cover the intensity ranges $5800, \ldots, 8000$ for Landsat 8, $0, \ldots, 255$ for SIR-C, and $0, \ldots, 255$ m for ASTER GDEM. Image size is 2151 by 951 pixels. The curves of constant registration accuracy (pixels) and registered CPs found by the RAE method for test case 3 and the second-order polynomial model are shown in (a) as green curves and red dots.

exceedingly high initial error, the corresponding correction shift was introduced: 180 pixels for test case 1 and 110 pixels for test case 2. This correction affected only the term $\mathbf{d}_{\text{Initial}}$ in (1), leaving the matrix $\mathbf{A}_{\text{Initial}}$ unchanged, e.g., initial rotation angle and scale factor. Even after this preliminary correction, the initial registration error magnitude exceeded the values expected for RS imagery (± 125 pixels for test case 1, ± 130 pixels for test case 2, and ± 110 pixels for test case 3). Nevertheless, we kept such an initial error to better illustrate the strength and benefits of the proposed approach.

Image scaling was applied in test cases $2, \ldots, 4$ and $6, \ldots, 8$. For cases 2, 4, 6, and 8, the aim was to correct the difference in spatial resolution of DEM images in latitudinal and longitudinal directions. Indeed, ASTER GDEM has the same 1 arc-second spatial resolution in both latitudinal and longitudinal directions, but these resolutions are different when they are expressed in meters. At the considered latitudes, the resolution in latitudinal direction is about 1.5 times higher than in longitudinal direction. Such a difference violates isotropic scaling hypothesis assumed to derive CRLB_{fBm} bound. That is why an anisotropic scaling with the factors 1/2 in longitudinal direction and 1/3 in latitudinal direction was applied to the ASTER GDEM data to correct its initial spatial resolution and to make it almost the same in both directions (60 m). The optical and radar images in these test cases were rescaled in order to match the DEM spatial resolution. Scaling with the factor 1/2 in test cases 3 and 7 was applied to take into account the fact that CRLB_{fBm} bound becomes less adequate at the main scale for the SIR-C radar image, as it was shown in [27].

For each image, noise is characterized by spatial correlation function width σ_c pixels, and signal-dependent noise variance in the form $\sigma_n^2 = \sigma_a^2 + \sigma_P^2 I + \sigma_\mu^2 I^2$, where σ_a^2 is the additive noise variance, σ_P^2 is a coefficient defining the Poisson component (applicable for optical data), and σ_μ^2 is a coefficient responsible for multiplicative noise component (applicable for radar data).

Modifications of CRLB_{fBm} bound allowing it to deal with such a complex noise model were introduced in our recent paper [27]. Using the methods proposed in [46]–[48], the following estimates were obtained (at the main scale of each image): $\sigma_c = 0$ pixels, $\sigma_n^2 = 69.64 + 0.071 \cdot I$ and $\sigma_n^2 = 69.6363 + 0.0714 \cdot I$ for Hyperion bands #25 and #155, respectively; $\sigma_c = 0.57$ pixels, $\sigma_n^2 = 35.55 + 0.021 \cdot I$, $\sigma_n^2 = 81.71 + 0.013 \cdot I$, and $\sigma_n^2 = 68.80 + 0.033 \cdot I$ for Landsat 8 bands B1, B5, and B8, respectively; and $\sigma_c = 1.2$ pixels, $\sigma_n^2 = 37.23 + 0.316 \cdot I^2$

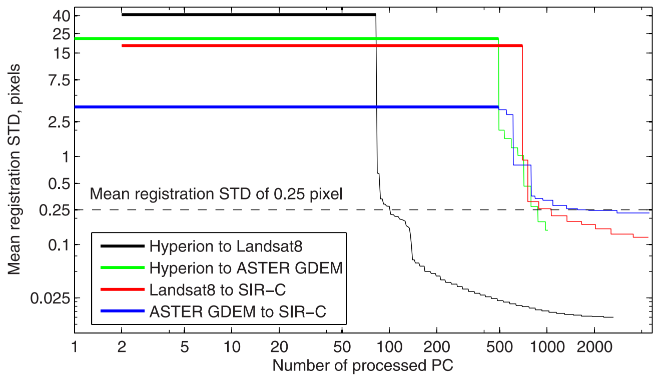


Fig. 6. RAE method registration accuracy improvement with iteration number for the four test cases.

and $\sigma_n^2 = 0.250 \cdot I^2$ for SIR-C radar images pr41419 and pr43020, respectively.

Noise in DEM images (data) is of different nature and mainly relates to elevation measurement error. Therefore, σ_n is measured in meters in this case. We have assumed a pure additive model for the DEM measurement error that follows Gaussian distribution and exhibits spatial correlation. The following estimates were obtained: $\sigma_c = 1.64$ pixels (at resolution of 30 m), $\sigma_n^2 = 9.45$ m^2. The obtained value $\sigma_n = 3.07$ m is consistent with mean SD of ASTER GDEM error of 3.52 m reported by Becek in [49].

Relief influence is a factor that might have an impact on the analysis of registration results. Landsat 8 images in the test cases are orthorectified and free from relief influence as well as ASTER GDEM images. Hyperion and SIR-C images were corrected for relief influence based on sensor parameters, viewing angles, and DEM information (using the same ASTER GDEM from the respective test cases).

B. *RAE Method Performance Analysis*

Let us first analyze the performance of the RAE method and illustrate its behavior and distinctive features for the eight test cases. It is important to mention that all results were obtained using the same RAE settings specified in the previous sections and $N_{\mathrm{TI}} = N_{\mathrm{RI}} = 17$ pixels.

The first feature—incremental registration accuracy improvement—is illustrated in Fig. 6. For test cases $1, \ldots, 4$ and affine registration model, the dependence of the mean registration error SD $\bar{\sigma}_{\mathrm{reg}}$ is shown versus the number of processed PCs, where $\bar{\sigma}^2_{\mathrm{reg}}$ is defined as the mean of $\sigma^2_{\mathrm{reg}}(i_{\mathrm{TI}}, j_{\mathrm{TI}})$ over the whole reference image area. At the initialization stage, $\sigma_{\mathrm{reg}}(i_{\mathrm{TI}}, j_{\mathrm{TI}})$ is set to $d_{\max 0}/3$ for all CFs. During the initialization stage, $\bar{\sigma}_{\mathrm{reg}}$ does not evolve until the first valid geometrical transform model parameters are found.

The complexity of the initialization stage depends on the registration problem complexity, and it can be measured as the necessary number of PCs processed before completion. We can see that, for the simplest case of optical-to-optical registration, the initialization stage is very short; it takes only 75 PCs to complete (see Table II for the numerical analysis results). Complexity increases for the multimodal test cases $2, \ldots, 4$, which require from about 550 to 900 PCs to initialize geometrical transform parameters. Somewhat unexpectedly, the registrations of radar and optical images to DEM are simpler than optical-to-radar registration in terms of initialization stage complexity.

After the initial registration stage, $\bar{\sigma}_{\mathrm{reg}}$ decreases fast due to having the search zone reduced and the increased probability of finding right correspondences. When the registration error gets small enough (<1 pixel), the search zone becomes so narrow that the PC search procedure is no longer needed: majority of the new PCs processed by the RAE method are true correspondences, but their registration accuracy $\sigma_{\mathrm{PC.LB}}$ can vary. In this operation mode, the mean registration error SD decreases approximately as a reciprocal square root of the number of processed PCs. This stage of the RAE method can be viewed as a fine-registration stage. The same observations were made from test cases $5, \ldots, 8$ (see Table III for the quantitative data), except that optical-to-DEM registration test case 6 has now the longest initialization stage.

In practice, three distinctive stages of RAE operation can be outlined: the initialization stage where the true PC search procedure is the most time-consuming, the intermediate stage where the search zone is permanently decreased to reach the size of the NCC lobe, and the fine-registration stage where the search zone is virtually annihilated and the whole processing time is spent to increase the registration accuracy.

Qualitatively, the RAE method performance for the first- and second-order polynomial models is summarized in Table II for TC1, ..., TC4 and in Table III for TC5, ..., TC8. The registration is successful in all cases. The number of registered CFs varies from 50 to about 1600, being larger for optical-to-optical and optical-to-radar cases and smaller for optical- and radar-to-DEM cases. This number is enough to provide subpixel registration over the entire registered image area (as indicated by the "Min/Mean/Max registration SD" line in Table II) in all test cases with affine model. While this accuracy is normal for monomodal cases 1 and 5, the ability to perform subpixel registration for the rest of the multimodal cases outlines the strong capability of the RAE method to succeed in registering complex multimodal image pairs.

The most visible characteristic of registration complexity is the percentage of inlying CFs, previously denoted as $P^{\mathrm{in}}_{\mathrm{CF}}$. According to this criterion, test cases $1, \ldots, 8$ can be ordered in terms of increased complexity: optical-to-optical, optical-to-radar, DEM-to-radar, and the most complex case is optical-to-DEM registration. Recall that $P^{\mathrm{in}}_{\mathrm{CF}}$ defines the probability that there is, at least, one true correspondence for a CF that has, at least, one CRLB-validated PC. The value of $P^{\mathrm{in}}_{\mathrm{CF}}$ increases as more PCs are tested, and it takes the lowest value at the end of the initialization stage (these values for the first-/second-order models are 75.8/51.9% and 59.8/91.3% for optical-to-optical TCs 1 and 5, 5.8/1.9% and 0.5/2.4% for optical-to-DEM TCs 2 and 6, 63.4/85.2% and 47.9/31.8% for optical-to-radar TCs 3 and 7, and 21.6/63.1% and 46.2/4.5% for radar-to-DEM TCs 4 and 8; see data in Tables II and III). Among all test cases, the lowest value of $P^{\mathrm{in}}_{\mathrm{CF}}$ of about 2% (percentage of outliers about 98%) is obtained for the optical-to-DEM case. The RAE method succeeds even for this extremely low $P^{\mathrm{in}}_{\mathrm{CF}}$ value. To the best of our knowledge, there are no successful

TABLE II
RAE METHOD PERFORMANCE CHARACTERISTICS FOR TEST CASES 1, . . . , 4

Parameter		Test case 1	Test case 2	Test case 3	Test case 4
Registration problem type		optical-to-optical	optical-to-DEM	optical-to-radar	DEM-to-radar
Overall number of CF		2169	1224	9234	4321
Percentage of inlying CF, P_{CF}^{in}		95.1801	31.4109	64.7286	68.3713
Affine model					
Length of the initialization stage		84	495	702	496
Number of processed PC		2648	1019	4406	4473
Percentage of inlying CF, P_{CF}^{in} (init)		75.7917	5.8243	63.4428	21.5689
Number of registered CF		1288	87	210	52
Image area registered with SD less than 0.25 pixels, %		100	100	100	52.2183
Min/Mean/Max registration SD		0.008/0.015/0.025	0.053/0.108/0.200	0.056/0.115/0.237	0.111/0.256/0.572
RMSE (SD of absolute error), pixels (number of points)	$\sigma_{PC.LB} < 0.35$	0.5904 (1288)	0.8436 (87)	1.2529 (210)	0.82621 (52)
	$\sigma_{PC.LB} < 0.225$	0.5080 (1186)	0.7453 (64)	0.8879 (24)	0.46818 (6)
	$\sigma_{PC.LB} < 0.15$	0.3394 (956)	0.4869 (26)	---	---
SD of normalized error, σ_{norm}		3.9043	4.2437	4.6807	3.0896
Second order polynomial model					
Length of the initialization stage		157	2009	3140	1428
Number of processed PC		2788	11991	7118	6453
Percentage of inlying CF, P_{CF}^{in} (init)		51.8646	1.9125	85.2335	63.0566
Number of registered CF		1293	65	195	59
Image area registered with SD less than 0.25 pixels, %		100	50.56	64.0626	30.1379
Min/Mean/Max registration SD		0.013/0.024/0.062	0.086/0.298/0.998	0.092/0.267/1.395	0.158/0.491/2.364
RMSE (SD of absolute error), pixels, x/y (number of points)	$\sigma_{PC.LB} < 0.35$	0.5921 (1293)	0.62215 (65)	1.065 (195)	0.672 (59)
	$\sigma_{PC.LB} < 0.25$	0.5090 (1189)	0.52815 (47)	0.84431 (35)	0.38554 (6)
	$\sigma_{PC.LB} < 0.15$	0.3371 (960)	0.46619 (24)	---	---
SD of normalized error, σ_{norm}		3.8549	3.3291	4.0768	2.4606

TABLE III
RAE METHOD PERFORMANCE CHARACTERISTICS FOR TEST CASES 5, . . . , 8

Parameter		Test case 5	Test case 6	Test case 7	Test case 8
Registration problem type		optical-to-optical	optical-to-DEM	radar-to-optical	radar-to-DEM
Overall number of CF		3307	10709	3396	2507
Percentage of inlying CF, P_{CF}^{in}		97.7874	39.0515	98.6447	82.0548
Affine model					
Length of the initialization stage		22	5730	56	94
Number of processed PC		2550	11911	5247	2561
Percentage of inlying CF, P_{CF}^{in} (init)		59.8048	0.5	47.9450	46.2770
Number of registered CF		1337	434	1585	205
Image area registered with SD less than 0.25 pixels, %		100	100	100	100
Min/Mean/Max registration SD		0.010/0.019/0.034	0.028/0.064/0.130	0.013/0.024/0.045	0.046/0.097/0.194
RMSE (SD of absolute error), pixels (number of points)	$\sigma_{PC.LB} < 0.35$	0.63388 (1337)	0.85119 (434)	0.72821 (1585)	1.0139 (205)
	$\sigma_{PC.LB} < 0.225$	0.57903 (1256)	0.81275 (328)	0.5678 (1076)	0.73816 (84)
	$\sigma_{PC.LB} < 0.15$	0.42427 (888)	0.60457 (47)	0.47738 (399)	0.52898 (18)
SD of normalized error, σ_{norm}		4.1813	4.2451	3.6400	4.2591
Second order polynomial model					
Length of the initialization stage		60	6162	66	2286
Number of processed PC		2263	12840	5617	7157
Percentage of inlying CF, P_{CF}^{in} (init)		91.3188	2.4361	31.8714	4.5512
Number of registered CF		1155	783	1594	192
Image area registered with SD less than 0.25 pixels, %		100	100	100	94.4609
Min/Mean/Max registration SD		0.016/0.036/0.113	0.033/0.052/0.162	0.020/0.035/0.103	0.069/0.140/0.392
RMSE (SD of absolute error), pixels, x/y (number of points)	$\sigma_{PC.LB} < 0.35$	0.59135 (1155)	0.75155 (783)	0.74459 (1594)	0.85119 (192)
	$\sigma_{PC.LB} < 0.225$	0.53047 (1089)	0.72007 (651)	0.55863 (1079)	0.60198 (84)
	$\sigma_{PC.LB} < 0.15$	0.37541 (773)	0.52004 (108)	0.44259 (410)	0.44945 (18)
SD of normalized error, σ_{norm}		3.8297	3.8719	3.6166	3.5437

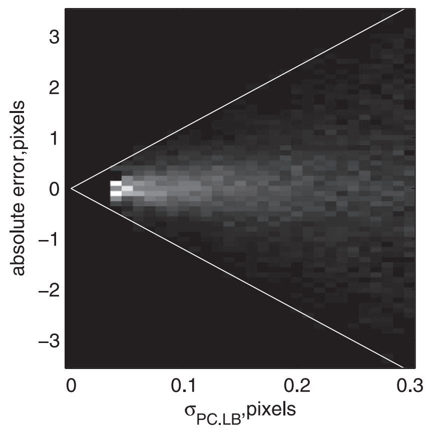


Fig. 7. 2-D histogram of absolute registration error (w.r.t. both directions) and $\sigma_{\rm PC.LB}$ for all correspondences found for test cases 1–8. White color corresponds to higher probability density.

registration examples yet with percentage of outliers exceeding 90% for methods available in the literature.

For the affine model, the RMSE of the found correspondences takes the smallest value of about 0.6 pixels for optical-to-optical registration cases 1 and 5, and it is from 0.72 to 1.25 pixels for test cases $2, \ldots, 4, 6, \ldots, 8$. This RMSE is calculated using all found correspondences. As it was discussed previously, each PC is characterized by its own registration accuracy, and the RAE method is able to both characterize it and take it into account in its operation principle to improve registration performance. Let us next show the validity of $\sigma_{\rm PC.LB}$ estimates for characterizing PC registration accuracy at the local scale.

Fig. 7 shows the 2-D histogram of absolute registration errors (w.r.t. both horizontal and vertical directions) as a function of $\sigma_{\rm PC.LB}$ for all correspondences found in the whole set of eight test cases. For a fixed $\sigma_{\rm PC.LB}$ value, each histogram row is normalized to represent the experimental pdf of the absolute registration error. The values of this error at the level $\pm 12\sigma_{\rm PC.LB}$ are shown as white lines. These lines depict approximately the decision threshold separating inliers from outliers in the RAE method (recall that separation is achieved by comparing $\hat{P}_{\rm kp}^{\rm in}$ with the selected threshold 0.9; in turn, $\hat{P}_{\rm kp}^{\rm in}$ defined by (9) is dependent on $\sigma_{\rm PC.LB}$). It is seen that the distribution of absolute errors concentrates more and more toward zero as $\sigma_{\rm PC.LB}$ decreases. Moreover, for all $\sigma_{\rm PC.LB}$ values, the registration error distribution remains nonuniform (close to normal) and decaying at the $\pm 12\sigma_{\rm PC.LB}$ level. This behavior confirms that the observed concentration of registration errors is not due to the inlier detection procedure but the structural difference of registered image textures reflected by $\sigma_{\rm PC.LB}$.

We can conclude that the $\sigma_{\rm PC.LB}$ value reflects well the registration accuracy of the found correspondences and can be used to select the most accurate of them. This is demonstrated in Tables II and III, where the registration results in terms of RMSE and upper threshold on $\sigma_{\rm PC.LB}$ are given on the same line (at the bottom of the tables for each model). In all cases, the RMSE value of the found correspondences can be significantly reduced at the expense of getting fewer correspondences.

The normalized error values are obtained by dividing the registration error of each found correspondence by the respective value of $\sigma_{\rm PC.LB}$. For an effective estimator, the normalized errors SD, denoted as $\sigma_{\rm norm}$, should be close to unity. For a real estimator, as the NCC considered in this study, the $\sigma_{\rm norm}$ value and the estimator efficiency are related to each other as $e = 1/\sigma_{\rm norm}^2$. For the NCC estimator, $\sigma_{\rm norm}$ varies from 2.5 to 4.25 for all test cases (the lower value is obtained for the more accurate second-order polynomial model as shown in Tables II and III), which corresponds to an efficiency value of about 10%. Therefore, we can justify again and validate the choice of the NCC estimator, selected in Section III just based on its efficiency observed in previous works [20], [26], [27]. We see in addition that the NCC estimator, although relatively simple, can be applied to a variety of complex multimodal registration problems with acceptable efficiency (such a possibility was mentioned by Mikolajczyk and Schmid in the discussion section in [30]).

The main differences between the registration results using the second-order polynomial model as compared to the first-order model are as follows: the RMSE of the registered PCs remains at the same level or slightly decreases due to a more complex geometrical model. The initialization stage length tends to increase significantly (up to 20 times). The reason for this is that a more complex model has lower predictive capability, and tighter PC clustering is needed to initialize the geometrical transform parameters with the RAE method. This also leads to reduced number of the registered PCs for test cases 2 and 3 as some of the PCs do not belong to tight clusters. However, if the second-order model is significantly more adequate as compared to the first-order model, the number of registered PCs can increase. This is the case for TC6. The overall registration process complexity increases as well due to the same reason: the search zone reduction is less effective for the second-order polynomial model with lower predictive capability. While the mean registration error SD remains almost the same, the maximal error increases significantly due to fast error divergence in regions not enclosed by the registered PCs.

Another feature of the RAE method that needs to be checked is its ability to predict the registration accuracy at the global scale, which is correctness of the estimates $\sigma_{\rm reg}(k)$. For all found correspondences, we have calculated the error between their position estimated by RAE and the one obtained using the reference transformation. We have normalized these errors by $\sigma_{\rm reg}(k)$. For TCs $1, \ldots, 8$, they lie approximately within ± 6-sigma interval with an SD value about 1.5. SD of normalized errors exceeding unity means a slight underestimation of the registration error, but overall, we can conclude that $\sigma_{\rm reg}(k)$ is a correct estimate of the registration accuracy provided by the RAE method.

For test cases $1, \ldots, 4$, five registered CFs corresponding to the lowest value of $\sigma_{\rm PC.LB}$ are shown in Fig. 8. We see that, for the optical-to-optical case, RI and TI CFs are almost identical. For the optical-to-DEM case, the intensities of RI and TI CFs are mostly inverted. The similarity between the pairs of CFs is lower than that for test case 1 but is still obviously visible. For the optical-to-radar case, a very high level of speckle noise affecting radar image CFs is observed. For most of the pairs, the intensity inversion between the registered optical and radar images takes place as well. The registered fragments from DEM

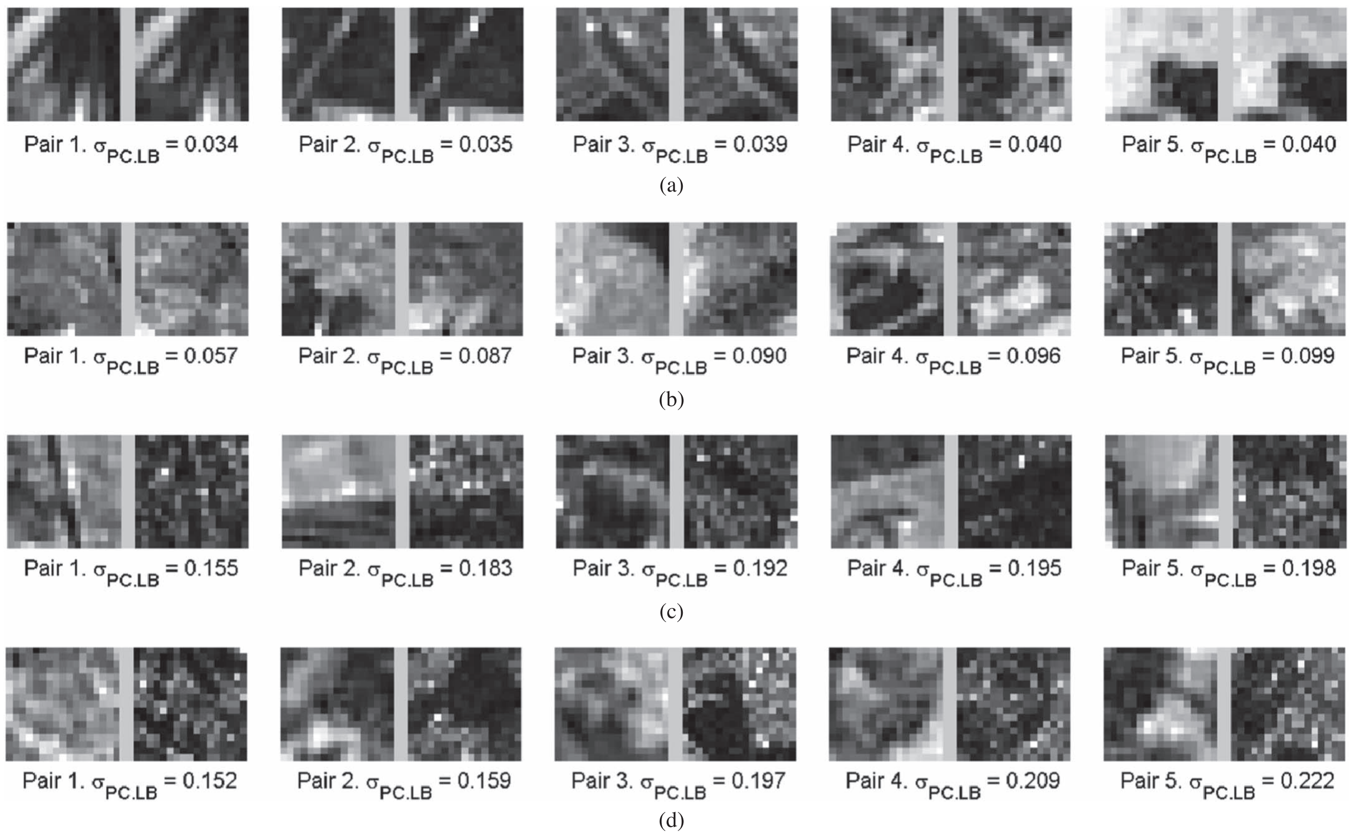


Fig. 8. Five control fragments with the lowest $\sigma_{\mathrm{PC.LB}}$ value (given in pixels) found by the RAE method for test cases 1–4: (a) optical-to-optical, (b) optical-to-DEM, (c) optical-to-radar, and (d) DEM-to-radar.

and radar images have very strong structural similarity. It is interesting to notice that many found correspondences do not belong to river basins but to forest edges. The effect is that both radar and DEM images are sensitive to forest edges but in a different manner. The radar images reflect the difference in surface roughness, the DEM images—in height.

The pairs of the registered correspondences demonstrate the complexity and variability of registration scenarios that can be successfully processed with the RAE method. An example of the registration result by the RAE method for test case 3 and second-order polynomial model is shown in Fig. 5(a) (with presented curves of constant registration accuracy, σ_{reg}, and positions of the registered CFs).

C. Comparative Analysis

The following registration methods were chosen for comparison: 1) the method for optical-to-radar image registration based on the improved SIFT descriptor (referred to as ImprovedSIFT) [31]; 2) the locally linear transforming image registration method (referred to as LLT) [23]; and 3) the registration based on MI similarity measure [50]. We have also tested an LLT method variant where the SIFT descriptor is replaced with a more robust ImprovedSIFT descriptor (referred to as LLT+ImprovedSIFT). The comparison was done w.r.t. the reference affine transformation. We stress that this reference transformation cannot be considered as very accurate since manual selection of control points for multimodal image pairs is subjective and inaccurate.

The code for the LLT method is available from the authors' webpage, and the ImprovedSIFT method was implemented using VLfeat library, strictly following the original paper with one difference. We have found that, among the three modifications of SIFT proposed in [31], namely, 1) skipping the first octave; 2) skipping the step of preponderant orientation assignment; and 3) using multiple support regions, the first one does not improve the registration quality for our test images. Therefore, we used only modifications 2 and 3 to obtain the results shown as follows.

Registration based on MI similarity measure was implemented using MATLAB's *imregister* function that includes MI calculated according to [50]. The *imregister* function allows finding local MI maxima using the gradient descent algorithm [51] in the neighborhood of user-supplied starting guess. We have used the reference affine transformation as such a starting guess but with additional translation in the interval $-50, \ldots, 50$ pixels w.r.t. both vertical and horizontal translations (with steps of 2.5 pixels) such that one starting guess exactly corresponds to the reference affine transformation. For each starting guess, the registration has been performed w.r.t. the affine transformation model, and the result with maximal MI value has been taken as the MI-based registration output. Such a multistart approach also checks whether MI measure has global maxima close to the true registration parameters.

The performance of the ImprovedSIFT and LLT methods is comparatively evaluated based on the following criteria: the number of found correspondences, the number of true correspondences (with the registration error not exceeding

TABLE IV
PERFORMANCE OF REGISTRATION METHODS IN COMPARISON ON TEST CASES $1, \ldots, 4$

	Method	Test case 1 optical to optical	Test case 2 optical to DEM	Test case 3 optical to radar	Test case 4 DEM to radar
RAE	Number of correspondences, all/true	**1288/1288**	**87/86**	**210/194**	**52/41**
	RMSE	**0.66**	**1.42**	**1.70**	**1.67**
	mean/max registration error	**0.64/1.41**	**1.557/2.913**	**1.9127/4.127**	**2.910/5.089**
Improved SIFT	Number of correspondences, all/true	**131/115**	7/0	10/0	15/0
	RMSE	**1.80**	12.15	283.34	593.98
	mean/max registration error	**0.73/1.74**	24.10/43.23	403.83/481.50	825.71/851.99
LLT	Number of correspondences, all/true	**60/53**	13/0	10/0	11/0
	RMSE	**1.99**	430.15	868.03	578.67
	mean/max registration error	**0.74/1.47**	480.72/1182.66	912.27/1821.18	751.23/1551.9
LLT+ Improved SIFT	Number of correspondences, all/true	**114/99**	6/0	10/0	40/0
	RMSE	**1.83**	303.63	515.95	658.95
	mean/max registration error	**0.82/1.91**	338.29/981.46	820.53/1887.27	712.43/1769.7
MI	mean/max registration error	**2.21/5.00**	27.88/61.67	**1.53/2.35**	51.826/76.34

TABLE V
PERFORMANCE OF REGISTRATION METHODS IN COMPARISON ON TEST CASES $5, \ldots, 8$

	Method	Test case 5 optical to optical	Test case 6 optical to DEM	Test case 7 optical to radar	Test case 8 DEM to radar
RAE	Number of correspondences, all/true	**1337/1337**	**407/344**	**1585/1575**	**205/197**
	RMSE	**0.76**	**2.13**	**1.26**	**1.71**
	mean/max registration error	**0.60/1.45**	**3.31/7.49**	**1.04/3.07**	**1.54/2.90**
Improved SIFT	Number of correspondences, all/true	**109/95**	51/2	8/0	**59/39**
	RMSE	**1.88**	8.24	117.28	**2.71**
	mean/max registration error	**1.21/2.79**	15.11/27.28	152.55/195.59	**3.62/7.60**
LLT	Number of correspondences, all/true	23/0	11/0	15/0	9/0
	RMSE	1302.07	553.11	624.34	408.51
	mean/max registration error	1383.08/2960.88	858.49/1954.50	894.22/1850.68	517.06/1229.8
LLT+ Improved SIFT	Number of correspondences, all/true	13/5	25/0	71/0	4/2
	RMSE	829.61	786.87	566.98	375.23
	mean/max registration error	1010.67/2517.15	880.48/2131.65	448.62/940.35	483.73/951.22
MI	mean/max registration error	**0.86/1.38**	63.01/152.42	**1.53/3.50**	**4.42/6.45**

4 pixels), the RMSE of the found correspondences, and the mean and maximal registration error. The two latter criteria were calculated w.r.t. the reference affine transformation. The MI-based registration accuracy is characterized only by the mean and maximal registration error. We did not use processing time as an additional criterion as our RAE method is one order of magnitude slower as compared to other analyzed methods.

The obtained quantitative results are presented in Tables IV and V, where successful registration outcomes are marked in bold. The LLT methods with classical SIFT descriptor and LLT+ImprovedSIFT were able to register the simplest optical-to-optical TC1 and failed for TCs $2, \ldots, 8$. ImprovedSIFT shows better results, and it registered TC1 and multimodal radar-to-DEM TC8. However, its accuracy and number of found correspondences are lower as compared to the RAE method. MI demonstrates even better performance. For monomodal TC1, 5 and multimodal optical-to-radar TC3, 7, its performance is close to that of the RAE. For multimodal radar-to-DEM case TC 8, MI has global maxima close to the reference affine transformation but with higher mean/max errors as compared to RAE. For the most complex three TCs 2, 4, and 6, all involving DEM, the global MI maxima did not correspond to a solution close to the reference affine transformation, and MI registration was unsuccessful.

We conclude that the proposed RAE method is superior in terms of number of found correspondences and registration accuracy as compared to the set of state-of-the-art methods assessed here. It is able to cope with very complex registration scenarios where other methods in comparison fail to provide correct results.

V. CONCLUSION

In this paper, a new fully automatic area-based registration method has been proposed and proved suitable for a wide variety of RS applications including such complex multimodal scenarios as registration of optical image to DEM, optical-to-radar images, and DEM-to-radar image. The main features of the proposed method are its ability to quantify registration accuracy, deal with both linear and nonlinear geometrical models, reach linearithmic complexity with respect to image area (number of control fragments available), and attain compromise between processing time and area-based method accuracy.

The registration accuracy has been emphasized and used at both local and global scales for determining correspondences between registered images (control fragments) and for estimating geometrical transformation model parameters, respectively.

Unlike previous studies in the image registration field that mainly characterize *a posteriori* registration accuracy of correspondences at the output of the registration process, we have derived and next introduced the knowledge on local registration accuracy as an additional *a priori* information in the registration process. Such a possibility essentially comes from our previous efforts to quantify the potential registration accuracy of textural noisy images achievable by area-based registration methods: $\mathrm{CRLB}_{\mathrm{fBm}}$ bound. Having such additional information, we have been able to improve the efficiency of both the outlier detection stage and the geometrical transformation parameter estimation stage. The most important benefit of our approach is that the registration accuracy at the global scale can be evaluated as the covariance matrix of estimates of polynomial geometrical model coefficients. We would like to outline that this registration accuracy does not need any ground truth to be determined and characterizes individual pairs of registered images, taking into account their inherent structure. The validity of the registration accuracy estimates at both scales has been experimentally confirmed. To the best of our knowledge, such a result was not published previously in the literature.

Local parametric image texture model and complex noise model (spatially correlated signal-dependent model) exploited within the proposed RAE registration method make it flexible and applicable to a wide range of registration problems including well-studied optical-to-optical and optical-to-radar cases and scarcely studied DEM-to-optical and DEM-to-radar scenarios. Comparison with state-of-the-art methods has shown that the RAE method can handle the most complex registration cases where other methods fail to provide accurate and reliable results.

Future work is intended to reduce the high computational complexity of the RAE method. For this purpose, either the use of a multiscale approach or the search for simpler approximations of $\mathrm{CRLB}_{\mathrm{fBm}}$ bound can be considered. Another interesting direction could be in the use of more advanced similarity measures than NCC, e.g., MI measure or even feature-based descriptors.

REFERENCES

[1] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, pp. 977–1000, 2003.

[2] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, *Image Registration for Remote Sensing*. Cambridge, MA, USA: Cambridge Univ. Press, 2011.

[3] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[4] C. Pan, Z. Zhang, H. Yan, G. Wu, and S. Ma, "Multisource data registration based on NURBS description of contours," *Int. J. Remote Sens.*, vol. 29, no. 2, pp. 569–591, Jan. 2008.

[5] A. Goshtasby and J. Le Moign, *Image Registration: Principles, Tools and Methods*. London, U.K.: Springer-Verlag, 2012.

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

[7] D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1185–1199, Sep. 2004.

[8] I. S. Yetik and A. Nehorai, "Performance bounds on image registration," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1737–1749, May 2006.

[9] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2127–2137, Sep. 2005.

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

[11] M. A. Siddique, M. S. Sarfraz, D. Bornemann, and O. Hellwich, "Automatic registration of SAR and optical images based on mutual information assisted Monte Carlo," in *Proc. IEEE IGARSS*, 2012, pp. 1813–1816.

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

[13] L. Hui, B. S. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 320–334, Mar. 1995.

[14] B. Younggi, C. Jaewan, and H. Youkyung, "An area-based image fusion scheme for the integration of SAR and optical satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2212–2220, Oct. 2013.

[15] B. C. Robertson, "Rigorous geometric modeling and correction of QuickBird imagery," in *Proc. IGARSS*, 2003, vol. 2, pp. 797–802.

[16] S. Dawn, V. Saxena, and B. Sharma, "Remote sensing image registration techniques: A survey," in *Image and Signal Processing*. Berlin, Germany: Springer-Verlag, 2010, pp. 103–112.

[17] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surveys*, vol. 24, no. 4, pp. 326–376, Dec. 1992.

[18] R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," *Biophys. J.*, vol. 86, no. 2, pp. 1185–1200, Feb. 2004.

[19] J. Inglada, "Similarity measures for multisensor remote sensing images," in *Proc. IGARSS*, 2002, vol. 1, pp. 104–106.

[20] M. L. Uss, B. Vozel, V. A. Dushepa, V. A. Komjak, and K. Chehdi, "A precise lower bound on image subpixel registration accuracy," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3333–3345, Jun. 2014.

[21] E. Cohen, D. Kim, and R. Ober, "Cramer–Rao lower bound for point based image registration with heteroscedastic error model for application in single molecule microscopy," *IEEE Trans. Med. Imag.*, vol. 34, no. 12, pp. 2632–2644, Dec. 2015.

[22] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, Apr. 2000.

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

[24] J. Zhao, J. Ma, J. Tian, J. Ma, and D. Zhang, "A robust method for vector field learning with application to mismatch removing," in *Proc. IEEE Comput. Soc. Conf. CVPR*, 2011, pp. 2977–2984.

[25] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, Aug. 2010.

[26] M. L. Uss, B. Vozel, V. V. Lukin, and K. Chehdi, "Efficient rotation-scaling-translation parameter estimation based on the fractal image model," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 197–212, Jan. 2016.

[27] M. Uss, B. Vozel, V. Lukin, and K. Chehdi, "Potential accuracy of translation estimation between radar and optical images," in *Proc. SPIE*, 2015, pp. 1–12.

[28] G. A. Seber and A. J. Lee, *Linear Regression Analysis*, vol. 936. New York, NY, USA: Wiley, 2012.

[29] I. Karybali, E. Psarakis, K. Berberidis, and G. Evangelidis, "Efficient image registration with subpixel accuracy," in *Proc. EUSIPCO*, 2006, pp. 1–5.

[30] 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.

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

[32] H. Lei, L. Zhen, and Z. Rui, "SAR and optical images registration using shape context," in *Proc. IGARSS*, 2010, pp. 1007–1010.

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

[34] S. Suri, P. Schwind, J. Uhl, and P. Reinartz, "Modifications in the SIFT operator for effective SAR image matching," *Int. J. Image Data Fus.*, vol. 1, pp. 243–256, 2010.

[35] 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.

[36] H. Gonçalves, J. Gonçalves, L. Corte-Real, and A. Teodoro, "CHAIR: Automatic image registration based on correlation and Hough transform," *Int. J. Remote Sens.*, vol. 33, no. 24, pp. 7936–7968, 2012.

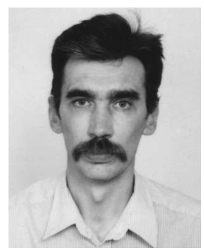
[37] H. Mostafavier and F. W. Smith, "Image correlation with geometric distortion part 1: Acquisition performance," *IEEE Trans. Aeros. Electron. Syst.*, vol. AES-14, no. 3, pp. 487–493, May 1978.
[38] H. Mostafavier and F. W. Smith, "Image correlation with geometric distortion part II: Effect on local accuracy," *IEEE Trans. Aeros. Electron. Syst.*, vol. AES-14, no. 3, pp. 494–500, May 1978.
[39] M. Debella-Gilo and A. Kääb, "Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation," *Remote Sens. Environ.*, vol. 115, no. 1, pp. 130–142, Jan. 2011.
[40] S. Wilks, *Mathematical Statistics*, vol. 1. New York, NY, USA: Wiley, 1962, p. 1971.
[41] L. Hairong and Y. Shuicheng, "Common visual pattern discovery via spatially coherent correspondences," in *Proc. CVPR*, 2010, pp. 1609–1616.
[42] B. Fan, F. Wu, and Z. Hu, "Towards reliable matching of images containing repetitive patterns," *Pattern Recognit. Lett.*, vol. 32, no. 14, pp. 1851–1859, Oct. 2011.
[43] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Trans. Math. Softw.*, vol. 3, pp. 209–226, 1977.
[44] R. Neal and G. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, vol. 89, M. Jordan, Ed. Dordrecht, The Netherlands: Springer-Verlag, 1998, pp. 355–368.
[45] L. Liberti and N. Maculan, *Global Optimization: From Theory to Implementation*, vol. 84. New York, NY, USA: Springer-Verlag, 2006.
[46] M. L. Uss, B. Vozel, V. V. Lukin, and K. Chehdi, "Local signal-dependent noise variance estimation from hyperspectral textural images," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 469–486, Jun. 2011.
[47] M. L. Uss, B. Vozel, V. V. Lukin, and K. Chehdi, "Maximum likelihood estimation of spatially correlated signal-dependent noise in hyperspectral images," *Opt. Eng.*, vol. 51, 2012, Art. no. 111712.
[48] M. Uss, B. Vozel, V. Lukin, S. Abramov, I. Baryshev, and K. Chehdi, "Image informative maps for estimating noise standard deviation and texture parameters," *EURASIP J. Adv. Signal Process.*, vol. 2011, 2011, Art. no. 806516.
[49] K. Becek, "Assessing global digital elevation models using the runway method: The advanced spaceborne thermal emission and reflection radiometer versus the Shuttle Radar Topography Mission case," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4823–4831, Aug. 2014.
[50] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Nonrigid multimodality image registration," in *Proc. SPIE 4322, Med. Imag., Image Process.*, 2001, pp. 1609–1620.
[51] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 3rd ed. New York, NY, USA: Springer-Verlag, 2008.

Mikhail L. Uss received the Diploma in radio engineering (with honors) and the Candidate of Technical Science degree in DSP for remote sensing from the National Aerospace University, Kharkov, Ukraine, in 2002 and 2006, respectively, and the Ph.D. degree in signal processing and telecommunications from the University of Rennes 1, Rennes, France, in 2011.

From 2005 to 2015, he was with the Department of Aircraft Radioelectronic Systems Design, National Aerospace University, as Docent and Department Head (2012–2015). He is currently with the Department of Transmitters, Receivers and Signal Processing, National Aerospace University. His research interests include statistical theory of radiotechnical systems, digital signal/image processing, image registration, blind estimation of noise characteristics, and theory of fractal sets with applications to image processing and remote sensing.

Benoit Vozel received the State Engineering degree, the M.Sc. degree in control and computer science, and the Ph.D. degree from Ecole Centrale de Nantes, Nantes, France, in 1991, 1991, and 1994, respectively.

As a Ph.D. student, he was with the Signal Processing Group, Institut de Recherche en Communication et Cybernétique de Nantes, where he worked on the detection of abrupt changes in signals. Since 1995, he has been with the Engineering School of Applied Sciences and Technology (ENSSAT), where he is currently with the Institute of Electronics and Telecommunications of Rennes (IETR). His research interests generally concern blind estimation of noise characteristics, image filtering and restoration, and adaptive image and remote sensing data processing.



Vladimir V. Lukin received the Diploma in radio engineering (with honors) from Kharkov Aviation Institute (now National Aerospace University), Kharkov, Ukraine, in 1983.

Since then, he has been with the Department of Transmitters, Receivers and Signal Processing, National Aerospace University. He defended the thesis of Candidate of Technical Science in 1988 and the thesis of Doctor of Technical Science in 2002 in DSP for remote sensing. Since 1995, he has been in cooperation with Tampere University of Technology, Tampere, Finland, where he is currently the Department Vice-Chairman and a Professor. His research interests include digital signal/image processing, remote sensing data processing, image filtering, and compression.

Kacem Chehdi received the Ph.D. and "Habilitation à diriger des recherches" degrees in signal processing and telecommunications from the University of Rennes 1, Rennes, France, in 1986 and 1992, respectively.

At the university, he was an Assistant Professor from 1986 to 1992, has been a Professor of signal and image processing with the Engineering School of Applied Sciences and Technology (ENSSAT) since 1993, was the head of the Laboratory of Systems Analysis for Information Processing (LASTI) from 1998 to 2003, and has been the Head of the Signal and Multicomponent/Multimodal Image Processing Research Team (TSI2M) within the Institute of Electronics and Telecommunications of Rennes (IETR) since 2004. His research interests include adaptive processing at every level in the pattern recognition chain by vision, blind restoration, and blind filtering (particularly the identification of the physical nature of image degradation and the development of adaptive algorithms) and segmentation and registration topics (particularly the development of unsupervised, cooperative, and adaptive systems). The main applications currently under investigation are multispectral and hyperspectral image processing and analysis.