

# Adaptation and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images

Guillaume Brigot, Elise Colin-Koeniguer, Aurélien Plyer, and Fabrice Janez

**Abstract**—The coregistration of heterogeneous geospatial images is useful in various remote sensing applications. Since the number of available data increases and the resolution improves, it is interesting to have an approach as automated, fast, robust, and accurate as possible. In this paper, we present a solution based on optical-flow computation. This algorithm called GeFolki allows the registration of images in a nonparametric and dense way. GeFolki is based on a local method of optical flow derived from the Lucas-Kanade algorithm, with a multiscale implementation, and a specific filtering including rank filtering, rolling guidance filtering and local contrast inversion. The efficiency of our coregistration chain is shown on radar, LIDAR, and optical images on Remningstorp forest in Sweden. An analysis of the relevant parameters is investigated for several scenarios. Finally, we demonstrate the accuracy of our coregistration by proposing specific metrics for LIDAR/radar coregistration, and optics/radar coregistration.

**Index Terms**—Computer vision, forestry, geophysical image processing, geographic information systems, image fusion, image registration.

## I. INTRODUCTION

Coregistration of heterogeneous images is useful in various remote sensing image fusion applications, since one expects a gain from the synergy of sensors. Applications where data fusion is relevant are numerous, whether for land classification [1], for agriculture [2], or forestry applications. In the case of the forest, many works illustrate the benefit of the combination of LIDAR and radar images [3], [4], LIDAR and high-resolution optical images [5], or radar and optical images [6], [7], for the characterization of plant species or biomass assessment.

In all cases where the expected product is a map, geometrically aligning two or more images in order to combine pixels corresponding to the same objects is a crucial step of the fusion.

Most methods of remote sensing image coregistration are based either on geocoding, or on nonrigid image registration methods that use only the images as input.

In the case of geocoding, the accuracy of coregistration will be highly dependent on the availability and precision of both a digital terrain model and the orbit parameters [8].

On the other hand, nonrigid image registration without geocoding is widely investigated in various fields beyond the

scope of remote sensing, for example in computer vision and medical imaging [9]. In computer vision, video image coregistration has to meet the constraints of robustness and speed of execution, but often focuses on images taken from the same sensor with little delay in time. In medical imaging [10] or remote sensing, difficulty generally lies in the different nature of the images to compare. Moreover, the context of remote sensing is also changing today with larger quantities of time series data, and some time-sensitive applications require fast processing. This is the case for example for near real time change detection for rapid post disaster assessment, wildlife tracking, and surveillance across broad areas such as battlefields or border regions [11].

Most nonrigid registration methods are *parametric methods*, meaning that an assumption is made about a parametrized model that constrains the form of the expected deformations between processed images. Then a similarity function is optimized to find an approximation of a real underlying deformation [12]. Among them, feature-based approaches establish a correspondence between a number of especially distinct points in images [13]. Selection of these points can use SIFT [14], [15] or SURF methods [16]. Other methods handle more complex features, such as segments, or use the shape descriptors [17], [18].

The choice of an image similarity measure is a key point. One of the most widespread used for the registration of multimodal images is mutual information [19]. Already used in remote sensing image coregistration, the main drawback of this measure is that it is quite time consuming. Instead of employing a similarity metric, [20] proposes to exploit a low rank constraint to jointly register multiple hyper-spectral images. Although such a model appears to be stable with respect to occlusions and imaging artifacts, it is not directly applicable in our case. In practice such a model requires a stack of multiple images and is further quite slow.

Another family of nonrigid coregistration methods is *non-parametric*. Among them, *dense* methods compute a displacement for every pixel in the image. They are particularly interesting in the case of very local deformation due for example to terrain elevation that has a lot of influence on high-resolution images. Most of dense and nonparametric methods belong to *optical flow* estimation. Optical flow is the pattern of apparent motion of objects in a visual scene caused by the relative motion between the sensor and the scene. Optical flow methods have been developed in a context where the constraints of speed and robustness to environmental effects have led to intensive efforts

Manuscript received October 06, 2015; revised March 04, 2016, April 12, 2016, May 20, 2016, and May 26, 2016; accepted May 31, 2016. Date of publication July 14, 2016; date of current version August 12, 2016.

The authors are with the DTIM, ONERA, Palaiseau 92320, France (e-mail: guillaume.brigot@onera.fr; elise.koeniguer@onera.fr; aurelien.plyer@onera.fr; fabrice.janez@onera.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/JSTARS.2016.2578362

in producing algorithms that combine robustness, precision, and high computing speed.

In this paper, we want to show how one optical flow method can undertake such a task of coregistration of heterogeneous images.

Optical flow has already been considered by the remote sensing community. Vogel *et al.* [21] proposes a dense method to reconstruct the annual motion of glacier surfaces from orthophotos. In contrast, in this study we focus on the more general case of heterogeneous data, acquired by different sensor modalities, like the coregistration of LIDAR to synthetic aperture radar (SAR) or optical images. In comparison to other existing methods developed for heterogeneous images often based on the use of mutual information (e.g., [10]), our algorithm prevails clearly when the execution speed matters. Further, we demonstrate that our algorithm can be applied successfully to our needs, while providing added value compared to a method based on geocoding. We have already shown in [22] that an optical flow algorithm can successfully coregister SAR images in various conditions, including the demanding interferometric ones. In this paper, we present a new version of this algorithm, called **GeFolki**, adapted to the coregistration of heterogeneous data, such as SAR-LIDAR images, or SAR-optical images. The acronym **GeFolki** stands for Geoscience Extended Flow Optical Lucas–Kanade Iterative.

A flow algorithm computes the displacement between two images. Most classical methods use partial derivatives of image intensity with respect to the spatial and temporal coordinates. They are based on two key points:

- 1) First, an observation model where intensity remains the same from one image to another. This hypothesis is called the *brightness constancy constraint*.
- 2) Then, a linearization step in order to be able to solve the equation. Generally, it enables to write the variation of the associated intensity as a local gradient.

In the case of heterogeneous images, brightness constancy is not respected because images have very different characteristics. It does not even make sense to compare SAR and photometric data with regard to intensity. Hence, the adaptation of such an optical flow algorithm to heterogeneous images is a challenging task. The solution proposed in this paper is to apply additional functions on the images to make them as similar as possible.

Moreover, the multiscale strategy is used to estimate large displacements.

Our study is therefore based on the adaptation of a generic optical flow algorithm that includes this multiscale iterative strategy to compute the existing displacement between two remote sensing images of different types. The resulting new algorithm is called **GeFolki**. In Section II, we present the previously mentioned **GeFolki** algorithm, detailing the different steps. Different results are presented in Section III. Section IV assesses the underlying robustness of the result and deals with the parameter setting. In Section V, we present a specific method to evaluate the performance of the results on the coregistration of images over forest lands, before concluding.

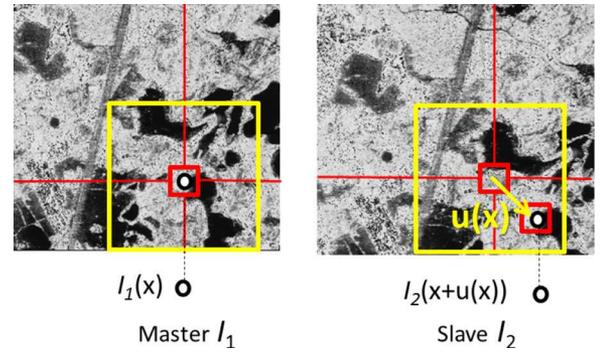


Fig. 1. Definition of variables used in our optical-flow algorithm: To register the slave image  $I_2$  to the master image  $I_1$ , we model the displacements between both images  $\mathbf{u}(\mathbf{x})$  for each pixel  $\mathbf{x}$  in the master image. Corresponding pixels  $I_1(\mathbf{x})$  and  $I_2(\mathbf{x} + \mathbf{u}(\mathbf{x}))$  are supposed to be close with respect to a similarity measure.

## II. DESCRIPTION OF GeFOLKI

### A. *GeFolki* Position Among Optical Flow Algorithms

Let us consider the registration of two images  $I_1$  and  $I_2$ , defined on a 2-D support  $S \in \mathbb{R}^2$ . The dense optical flow in computer vision is the pattern of apparent motion of objects. It corresponds to the displacement to find between both images. It is defined by  $\mathbf{u} : \mathbf{x} \rightarrow \mathbf{u}(\mathbf{x}) \in \mathbb{R}^2$ . Those variables introduced are illustrated in Fig. 1.

Most optical flow algorithms have a common starting point: it is to assume that pixel intensities remain constant from one frame to the next, what can be written as

$$I_1(\mathbf{x}) = I_2(\mathbf{x} + \mathbf{u}(\mathbf{x})). \quad (1)$$

This choice of *observation model* is called the *brightness constancy*. To compute the flow, the equality above needs usually a linearizing step. In most widely used techniques for optical flow computation, a change in intensity is written as

$$I_2(\mathbf{x} + \mathbf{u}(\mathbf{x})) = I_2(\mathbf{x}) + \nabla I_2 \cdot \mathbf{u}. \quad (2)$$

Equation (2) is often called the *linearized optical flow constraint*. This constraint is based on the assumption that the displacements are small.

In the case of remote sensing images, almost all assumptions are violated: there is no temporal proximity between images, they are acquired under uncontrolled acquisition conditions and even by different sensors. The classical algorithms are thus not suited to model or compensate these effects. In this study, we will show that with properly chosen modifications, the classical formulation can be adapted to these more general conditions.

As the optical flow system of (1) and (2) is underdetermined, optical-flow methods introduce in general additional conditions on the spatial distribution of the flow. Most prominently is to assume spatial smoothness in the flow field. This spatial additional condition can either be expressed globally or locally and relative methods can be classified respectively as local and global methods. Founding works are the global regularized framework of Horn and Schunck [23], and the local window-based approach of Lucas–Kanade [24].

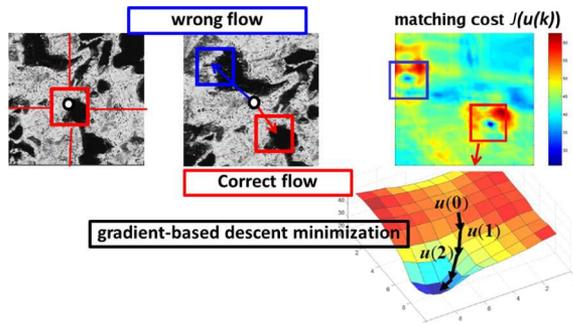


Fig. 2. General minimization scheme by an iterative method. On the left: master image with a local window under consideration. In the middle: Two candidate regions in the slave image. The correct match is highlighted in red, another window with high similarity in blue. On the right: The matching cost is in general a nonconvex function. Below: When started within sufficient proximity to the true minimum, gradient descent-based optimization can recover the correct solution.

Local optical flow methods are often considered as more time and memory efficient, while global techniques are based on a richer prior formulation on the estimated motion flow [25]. Therefore, global methods can produce piecewise constant flow fields, as for example in the case of TV regularization [26]. However they have a higher computational cost due to global formulation of the problem. More recent approaches combine both concepts.

However, in this paper, our aim is not to produce the most efficient gradient-based optical flow algorithm, but to show that it is possible to adapt one of them for the coregistration of heterogeneous remote sensing images. Our choice was to adapt eFolki, a fast and robust local method, developed in the perspective of the tradeoff between quality of the estimated flow and computational cost [27]. eFolki is derived from the Lucas–Kanade gradient-based method. Since eFolki has a remarkably simple and parallel structure, it is ideally suited to massively parallel computing architectures. Thus, the computation of the flow with our MATLAB implementation of **GeFolki** can be obtained in a few seconds for a  $5000 \times 5000$  pixel image on a Intel(R) Core(TM) i7-2620M CPU @ 2.80 GHz. A GPU implementation could provide this registration at the scale of hundreds of milliseconds.

Classical Lucas–Kanade algorithms assume that the flow is essentially constant in the local neighborhood of the pixel under consideration, and solves the basic optical flow equations for all the pixels in that neighborhood, by the least square criterion. Minimization is done by an iterative Gauss–Newton strategy based on first-order Taylor expansion of the intensity of image around a previous guess of the optical flow  $\mathbf{u}_k$ , as illustrated in Fig. 2.

Modern Lucas–Kanade algorithms are not only iterative, but also multiresolution. A pyramid of images is used to compute optical flow at varying scales following a coarse-to-fine strategy. This strategy enables to estimate large displacements despite the local constraint.

Our fast and robust algorithm can also be categorized as being of Lucas–Kanade type. Thus, it belongs to local or window-based approaches.

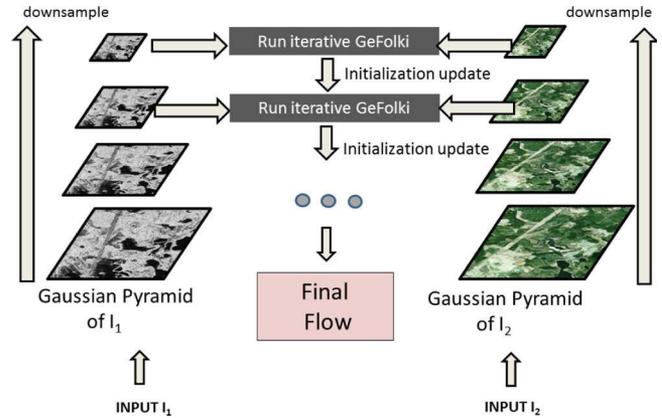


Fig. 3. To enable the recovery of large displacements, a multiscale strategy is employed: After the generation of image pyramids for master and slave image, the optical flow is computed from coarse to fine, by minimizing the cost function (3).

## B. General Scheme

**GeFolki** takes two input images resampled at the same rate, and outputs the two components of the flow at every pixel position. The aim of the algorithm is to minimize a given criterion  $J(\mathbf{u}; \mathbf{x})$  [28]

$$J(\mathbf{u}; \mathbf{x}) = \sum_{\mathbf{x}' \in S} \omega(\mathbf{x}' - \mathbf{x}) (f_1(I_1(\mathbf{x}')) - f_2(I_2(\mathbf{x}' + \mathbf{u}(\mathbf{x}))))^2 \quad (3)$$

where  $\omega$  defines a local window of radius  $r$  and size  $(2r + 1) \times (2r + 1)$ , thus  $\omega(\mathbf{x}) = 1$  iff  $|\mathbf{x}|_\infty \leq r$  and 0 else;  $f_1$  is a function applied to the master image  $I_1$ , and  $f_2$  is a function applied to the slave image  $I_2$ . These functions  $f_1$  and  $f_2$  are designed to project the images in a space where  $f_1(I_1)$  and  $f_2(I_2)$  are similar enough to validate the brightness constancy model. They significantly improve the robustness, but are also designed to retain the efficiency of our method.

Multiscale algorithms can deal with displacements larger than a few pixels because they initialize the flow by estimates from coarser image scales, where displacements are small enough to be estimated by local optimization. Unfortunately, too much down-sampling can remove information that may be vital for establishing the correct matches. Therefore, to converge to a correct solution, the initialization is important. We propose to start with a couple of images first computed in a common resolution grid. We compute the projective transformation based on the selection of four tie-points. The outcome from this first course registration becomes our input images for registration.

The first step is to create a pyramidal structure of down-sampled images, in order to be able to find large displacements. The number of pyramid levels  $L$  affects the amplitude of the desired movements. Then, we have to solve the minimization of the criterion  $J(u; x)$  iteratively in each scale, as represented in Fig. 3. The solutions are incrementally refined over the scales.

Then, for each pyramid level, from the coarsest resolution to the finer:

- 1) First, functions  $f_1$  and  $f_2$  are applied to the intensity images  $I_1$  and  $I_2$ . These functions are detailed below.

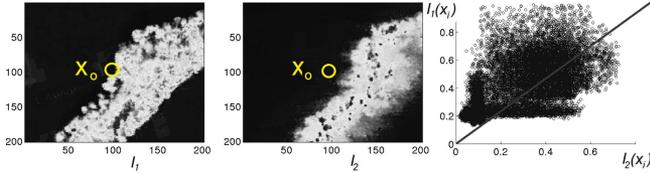


Fig. 4. Example of the contrast inversion decision. Left: Optical master image  $I_1$  for a local window. Middle: Radar slave image  $I_2$  for the same window. Right: Scatter Plot of  $I_1(x_i)$ ,  $I_2(x_i)$  for the example. Our criterion defined in (5) and (6) leaves the value of the central pixel  $x_0$  of the window unchanged.

- 2) Then, the algorithm minimizes the criterion expressed in (3) by first-order Taylor expansion of the residuals at the current flow estimate. At the first pyramid level, this initial current flow value is 0, and at the other levels, it derives from the expansion of a previous low resolution estimate.

The quadratic term of (3) is a composition of several functions with nontrivial derivatives, for example the rank function that is not continuous. Recall, that the Gauss–Newton iterations require a linearization around the current solution. Here, we cannot compute these derivatives explicitly. Instead we approximate the derivatives numerically, with regard to the current discretization of the image function. In other words, we directly consider  $f_1(I_1)$  and  $f_2(I_2)$  as functions defined on the image grid, after applying the similarity correcting transformations  $f_1$  and  $f_2$ , and operate with linear finite elements defined on the image grid. We did not experience any problems with this simplification.

In **GeFolki**, the function  $f_1$  and  $f_2$  are designed to ensure the robustness of the algorithm, as well as the execution speed. The success of the criterion lies in the presence of macroscopic components similar in both images. It is imperative to adjust the parameters in such a way that these structures are not affected by noise or by smaller details that differ in texture. Thus, in **GeFolki**,  $f_1$  and  $f_2$  are defined as follows:

- 1) For coregistration of SAR images, function  $f_1 = f_2 = R$  is a rank filter, as proposed in [22]. This function will be explained below. Note that in this case, we are brought back to the initial eFolki algorithm, evaluated in [22].
- 2) For coregistration of heterogeneous images,  $f_1$  and  $f_2$  will include, in addition to the rank filter:
  - a) A filtering step based on the rolling guidance filter, in order to make image textures more similar.
  - b) A function of local contrast inversion, for all cases where the contrast exhibited between the different sensors is not the same. It is very often the case for coregistration between optical and radar images, and more generally for images acquired in different parts of the electromagnetic spectrum, for which the scattering levels change, as shown in Fig. 5.

Therefore, in this case  $f_1$  and  $f_2$  are not strictly the same, and can be written as

$$\begin{aligned} f_1 &= R \circ g \\ f_2 &= C \circ R \circ g \end{aligned}$$

where  $g$  is the *rolling guidance filter*,  $R$  is the *rank-filter*, and  $C$  is the *local contrast inversion decision function*.

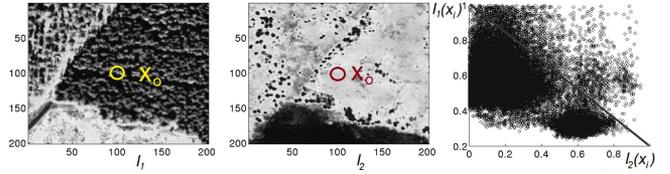


Fig. 5. Example of the contrast inversion decision. Left: Optical master image  $I_1$  for a local window. Middle: Radar slave image  $I_2$  for the same window. Right: Scatter Plot of  $I_1(x_i)$ ,  $I_2(x_i)$  for the example. Here, our criterion defined in (5) and (6) inverts the value of the central pixel  $x_0$  of the investigated window.

We choose for all of them fast execution functions to maintain the speed. They are described in the following sections.

### C. Rolling Guidance Filter $g$

An issue concerns the difference in texture and resolution in the original images. When the resolution of an image is too high compared to the other one, then the textures are generally different, and this makes the coregistration more difficult to do. Moreover, a detail present in one image, and not visible in the other one, can perturb the result of flow calculation, as well as the presence of speckle in SAR images. For example in the case of registration of SAR-LIDAR images, the main difficulty lies in the differences existing about texture and resolution effects. These differences make the flow calculation setting particularly challenging.

For these reasons, in order to maximize the efficiency of the algorithm, we propose to apply a prefiltering step that can make the intensity images more comparable. Among them, we have successfully tested the rolling guidance filter [29] that removes small-scale structures but maintains the edges in the image. To that end, the image is first blurred with a Gaussian filter. This step essentially removes small structures and artifacts. Starting from this blurred version, repeated bilateral filtering with the original image itself, then recovers the edge information.

In our experience, rolling guidance filter enables to filter images with the complete control of detail smoothing under a scale measure, which is especially interesting to use in **GeFolki** since it is also a multiscale strategy. Finally, the rolling guidance filter implementation achieves real-time performance and, then, can be integrated to our algorithm without degrading its computing performance. The complete description of the algorithm is given by [29], and the source code is provided by the authors.

### D. Rank Filter $R$

This rank function  $R(I)$  applied to the image  $I$ , is expressed as

$$R(I)(\mathbf{x}) = \#\{\mathbf{x}': \mathbf{x}' \in S_R(\mathbf{x}) \text{ with } |I(\mathbf{x})| > |I(\mathbf{x}')|\} \quad (4)$$

where  $S_R(\mathbf{x})$  is a neighborhood of the pixel  $\mathbf{x}$ .

We use  $\#$  to denote the number of elements of the set. This rank function consists in replacing the intensity value  $I(\mathbf{x})$  by the number of pixels in the neighborhood of  $\mathbf{x}$  with an intensity lower than  $I(\mathbf{x})$ . Thus, the rank transform depends on the relative order of pixel values within a local neighborhood, which is invariant under various transformations of the data, namely

those which preserve this order. The effect of the rank transform is a nonlinear filter that highly compresses the signal dynamics. By this compression effect on the signals gradient, the rank filter enhances the robustness of the motion estimation. This has been already studied in [30], and used in [31]. Other functions such as Census transform could also be used, but empirically deliver worse results. Moreover, rank filter is better suited to a massively parallel implementation.

### E. Local Contrast Inversion Decision Function $C$

In order to be compliant with the brightness constancy model, we need to address the possible differences in terms of contrasts. From one image to another, these contrasts can be locally reversed. This is particularly the case between a radar image and an optical image.

In order to decide whether the intensity value of the pixel under consideration  $\mathbf{x}_0$  should be inverted or not, we proceed as follows. We consider  $\mathcal{W}$  a  $(2M + 1) \times (2M + 1)$  sliding window centered around  $\mathbf{x}_0$ , defined by  $\mathcal{W}(\mathbf{x}_0) = \{\mathbf{x} : |\mathbf{x} - \mathbf{x}_0|_\infty \leq M\}$ .

We determine whether the set of points  $\{(I_1(\mathbf{x}), I_2(\mathbf{x} + \mathbf{u}(\mathbf{x}))) : \mathbf{x} \in \mathcal{W}(\mathbf{x}_0)\}$  is closer to the line defined by the equation  $y = x$  or the line defined by the equation  $y = 1 - x$ . To this aim, we have to compute the following criteria:

$$C_1(\mathbf{x}_0) = \sum_{\mathbf{x}_i \in \mathcal{W}(\mathbf{x}_0)} |I_1(\mathbf{x}_i) - I_2(\mathbf{x}_i + \mathbf{u}(\mathbf{x}_i))| \quad (5)$$

$$C_2(\mathbf{x}_0) = \sum_{\mathbf{x}_i \in \mathcal{W}(\mathbf{x}_0)} |1 - I_1(\mathbf{x}_i) - I_2(\mathbf{x}_i + \mathbf{u}(\mathbf{x}_i))|. \quad (6)$$

If  $C_1(\mathbf{x}_0) < C_2(\mathbf{x}_0)$ , we do not have to invert the intensity value of the second image, whereas if  $C_2(\mathbf{x}_0) < C_1(\mathbf{x}_0)$  we replace  $I_2(\mathbf{x}_0 + \mathbf{u}(\mathbf{x}_0))$  by  $1 - I_2(\mathbf{x}_0 + \mathbf{u}(\mathbf{x}_0))$  in image  $I_2$ .

One example for each case is given in Figs. 4 and 5 with  $M = 100$ . The first case concerns images of a transition between a vegetated area and a water one. In both radar and optical images, the contrasts are the same, and the intensity point cloud is closer to the line  $y = x$  than the line  $y = 1 - x$ . As a consequence, no inversion will be applied to the central pixel. In the second case, we see a transition between a forest area and a bare soil area. In the optical image, forest is darker than the soil, whereas forest is brighter than the soil in the radar image. Thus, the corresponding point cloud is closer to the line  $y = 1 - x$ . The value of the central pixel  $\mathbf{x}_0$  of the slave image will be replaced by  $1 - I_2(\mathbf{x}_0 + \mathbf{u}(\mathbf{x}_0))$ .

One situation where this strategy does not work could be when one of the images contains an element, and the other image does not contain it. The strategy of parameter settings for the window size will be discussed in Section IV-A.

For the whole images, these criteria can be expressed as

$$C_1(\mathbf{x}) = (|I_1 - I_2| * \mathbf{W})(\mathbf{x}) \quad (7)$$

$$C_2(\mathbf{x}) = (|1 - I_1 - I_2| * \mathbf{W})(\mathbf{x}) \quad (8)$$

where  $*$  denotes the convolution between images and  $\mathbf{W}$  is a  $(2M + 1) \times (2M + 1)$  all-ones matrix. The advantage of these

criteria is that they can be rapidly computed by using integral images.

Then, the transformed slave image  $C(I_2)$  is computed from

$$C(I_2) = (C_2 < C_1) \bullet I_2 + (C_1 > C_2) \bullet (1 - I_2) \quad (9)$$

where  $(C_2 < C_1)$  and  $(C_1 > C_2)$  are simple logical arrays corresponding to the respective conditions expressed for each pixel, and  $\bullet$  is the Hadamard product.

## III. RESULTS

### A. Description of the Dataset

Now that we have presented the algorithm, we apply it to a concrete case of interest: the heterogeneous image registration on the site of Remningstorp, a Swedish forest. First, this site was chosen because the forest is a particularly interesting case for fusion and many open dataset exist for the scientific community. Second, it represents an interesting case for coregistration because forest images contain few specific features such as segments and isolated points. Finally, the images are airborne high-resolution images, where deformations are particularly interesting to estimate with a dense method. The dataset contains a radar image, an airborne laser scanning image, as well as an optical image.

The SAR acquisition is a P-band full-polarimetric image taken by the airborne Onera system SETHI [32] on 23th September 2010 during the BioSAR 2010 campaign. It corresponds to the first track of the BioSAR delivery. This acquisition has been chosen to maximize the common footprint with the LIDAR data provided by SLU (Swedish University of Agricultural Sciences). The quality of localization has been estimated on a trihedral corner with an accuracy of about 1 m. Accuracy of trajectory is 20 cm. Range and azimuth resolutions are 80 cm. Local incidence angle varies from  $25^\circ$  to  $60^\circ$ , and the image is  $12\,250 \times 7000$  pixels.

The LIDAR acquisition was carried out by the Helicopter-mounted TopEye Mk III system on 29th August 2010. All LIDAR data have been processed by SLU, i.e., going from the collected LIDAR raw data of point clouds to various products delivered as several raster files. Here, we use a raster image with a  $0.5\text{ m} \times 0.5\text{ m}$  resolution, which contains the difference of height value of the first LIDAR return and height value of the *digital elevation model* (DEM) at the LIDAR return horizontal position. This image is called *Digital Surface Model* (DSM) image and is georeferenced in WGS84.

For the georeferencing of the SAR data, another DEM was provided by FOI, the Swedish Defence Research Agency. This DEM covers the entire radar images. At the date of the BIOSAR 2010 campaign, this DEM was a commercial product with a single user license and not included in the deliverable distributed by ESA for public access. However, since July 2015 it is provided without license with simplified terms of use by the National Land Survey as open data. It contains ground height values at the intersections of a 50-m grid. This product is referred as GSD50+ and is described in [33]. Standard error in height is announced to be 2.5 m. To test the influence of this DEM on the coregistration, in particular on georeferencing, we have also

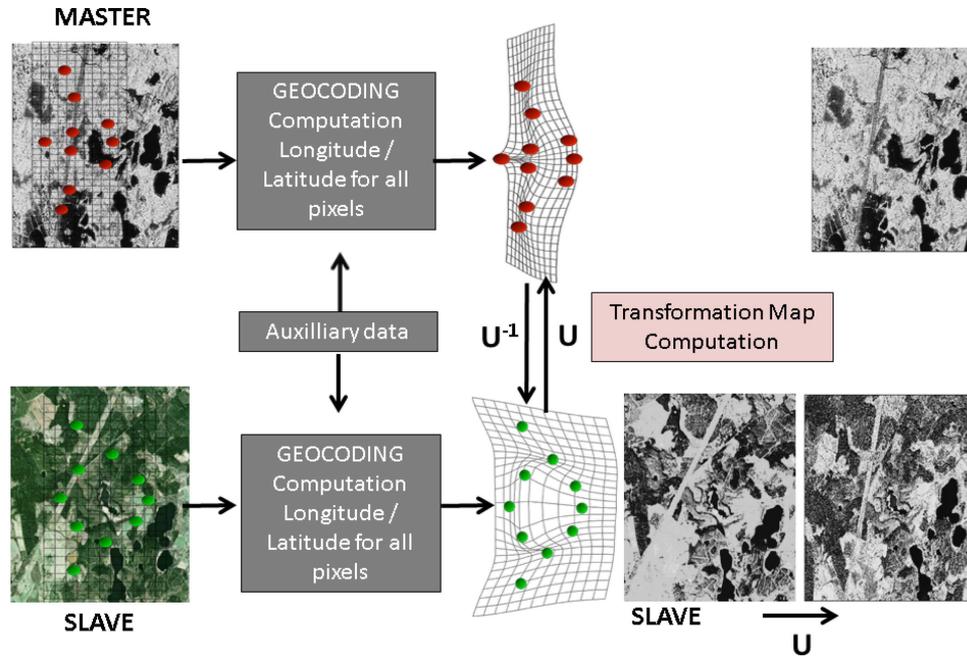


Fig. 6. First coregistration scheme by geocoding from a geometric point of view: the latitude and longitudes are computed for each pixel of the images according to topographic and track information. Then, the flow is computed between these both coordinates, and the slave image is resampled accordingly.

used the product delivered by SRTM90 data before 2015, which gives ground height values on a 90-m grid.

### B. Tested Coregistration Scenarios

For the coregistration of two images, we have to choose one as master and the other one as slave. When an SAR image is included in the dataset, we choose this one as the master image for the whole coregistration products for at least two reasons:

- 1) Projecting a radar image on the ground introduces artifacts, because of the special SAR geometric configuration [34]. These artifacts are further amplified for high relief terrains, and also when the incidence angle variation is great, such as encountered for our airborne acquisitions.
- 2) The choice of a master SAR image is in favor of advanced processing methods such as interferometry that will not be affected by a resampling of the radar image.

In order to enhance the usefulness of **GeFolki** for the coregistration, we have tested several scenarios as follows:

- 1) The SETHI SLC data have been delivered by ONERA with an auxiliary transformation grid file to be able to link precisely one pixel on the image to one geographic location. This grid file allows to link a pixel position with a height to a precise latitude and longitude. It can also be used to associate an image position and a height to a geographic position. Therefore, we have applied the process to link each pixel of our SAR image to a geographic location, by using different available DEM. This process is described in [35]. Then, this geocoding information is used in combination with the LIDAR one to find the differences of positions of image pixels in map coordinates. Thereafter, it is possible to resample one image onto the other one. Results obtained by this process will be called

**coregistered products by geocoding.** The corresponding synoptic is illustrated in Fig. 6. This method does not utilize the flow algorithm. It has been implemented in order to compare the results with those obtained using **GeFolki**. It also allows us to judge the influence of the quality of the DEM on the accuracy of registration. Indeed, we have applied it using the two types of DEM available on the whole image: SRTM90 and GSD50+ provided by the National Land Survey.

- 2) In parallel, we register images without using any georeferencing data. We begin with images after the initialization procedure explained in Section II-B, as represented in Fig. 7. Then, we apply **GeFolki** to compute the flow, and the slave image is resampled accordingly. We call this approach **GeFolki after coregistration by manual initialization.** This approach has been chosen because it enables us to attempt a coregistration even if no auxiliary data is available.

As a baseline, we use a method whose similarity metric is based on mutual information [36]. Because the baseline algorithm is more time consuming, we could not apply it at full resolution ( $12\,250 \times 7\,000$  pixel). For that reason, we have restricted the test on a part of the image only, and thereby, only to qualitative evaluation. The initial algorithm introduced by Thirion [37], called demon registration, is a popular algorithm for nonrigid image registration because of its linear computational complexity and ease of implementation. It is a nonparametric registration that approximately solves the registration problem by successively estimating force vectors that drive the deformation toward alignment, and smoothing the force vectors by Gaussian convolution. In this version, the method has been adapted to heterogeneous images. This algorithm is provided for

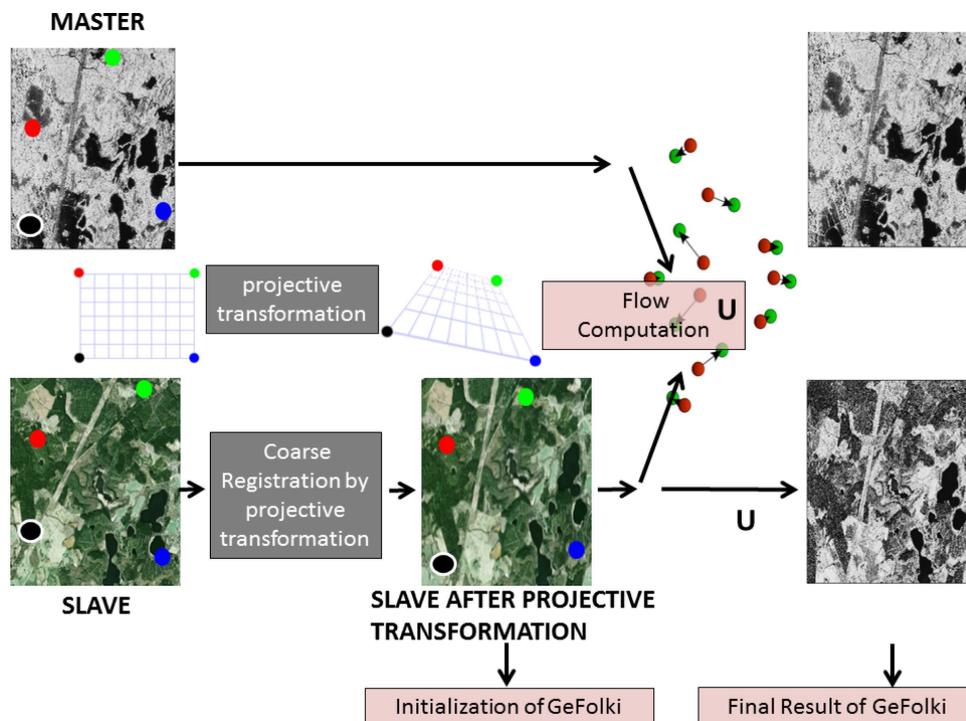


Fig. 7. Coregistration scheme by optical flow calculation: First a coarse coregistration is performed by warping the slave image on the master image through a projective transformation based on the manual selection of four tie-points. This first step constitutes the initialization of **GeFolki**. Then, **GeFolki** computes the flow for this couple of images, and the final slave image is resampled accordingly.

free by the authors in MATLAB language. Therefore, it is close in spirit to **GeFolki**, in the sense that it handles the conventional formulations of optical flow problem, adapts the similarity criterion to heterogeneous images, and is able to compute the flow for each pixel in the images.

The conclusions drawn from our various tests are

- 1) Geocoding methods with our dataset are insufficient to ensure a pixel precision. The impact of DEM accuracy on the result is clearly demonstrated in next section.
- 2) The improvement brought by **GeFolki** is significant. It can be particularly useful for all applications where the highest accuracy is required, for instance for the counting of trees.
- 3) **GeFolki** was successful even with its coarse initialization through a projective transformation based on a few tie-points.
- 4) With the same coarse initialization, the baseline algorithm based on mutual information was not successful. Therefore, we have managed to use it in combination with our best geocoding result as initialization. The results appear to be still of inferior quality and needs a 20 times more computation time.

In next sections, we detail these qualitative results, first, of the LIDAR DSM image at the best resolution (0.5 m) on the SAR image, and then of an optical aerial photography on the SAR image.

### C. 0.5 m × 0.5 m Elevation Map With the SAR Image

We consider here coregistration results of a 1000 × 1000 pixel part of the whole image SAR image, with the LIDAR

0.5 m × 0.5 m elevation map. This part contains forest stands, isolated trees, electrical cables, and pylons. Each result is presented as a colored composition in which the radar image is in the green channel and the coregistered LIDAR image in the magenta channel. When a pixel is white or grey, we have similar intensities in both images. If not, a pixel will tend to take the color of the channel with the highest intensity value. With this representation, it is then possible to appreciate the quality of a registration, in particular when we focus on isolated scatterers.

The two first results presented in Fig. 8 are coregistered products by geocoding. For the first one, we have used the SRTM90 DEM, and for the second one the GSD50+ DEM which is more accurate. For both cases, we can notice that a residual shift remains, notably between forest edges and small structures such as pylons. These results also demonstrate that the precision obtained by geocoding is better with the most accurate DEM of the Land National Survey, than with the SRTM product. Thus, the DEM choice is crucial on the final accuracy.

Then, Fig. 9 presents the results obtained by the baseline algorithm based on mutual information and by **GeFolki**. The baseline algorithm has been initialized by the previous result of best geocoding, because it was diverging for the coarse initialization based on the projective transformation.

For these two new results, the images show less saturated colors, which means that the coregistration results are better than using the geocoding method. However, some colored pixels remains, mainly because of the lateral viewing geometry of the radar, that implies that the return of the tree is often clearly double: the branches echoes appear in HV polarization at the first range cells, whereas the trunks double bounce echo appears

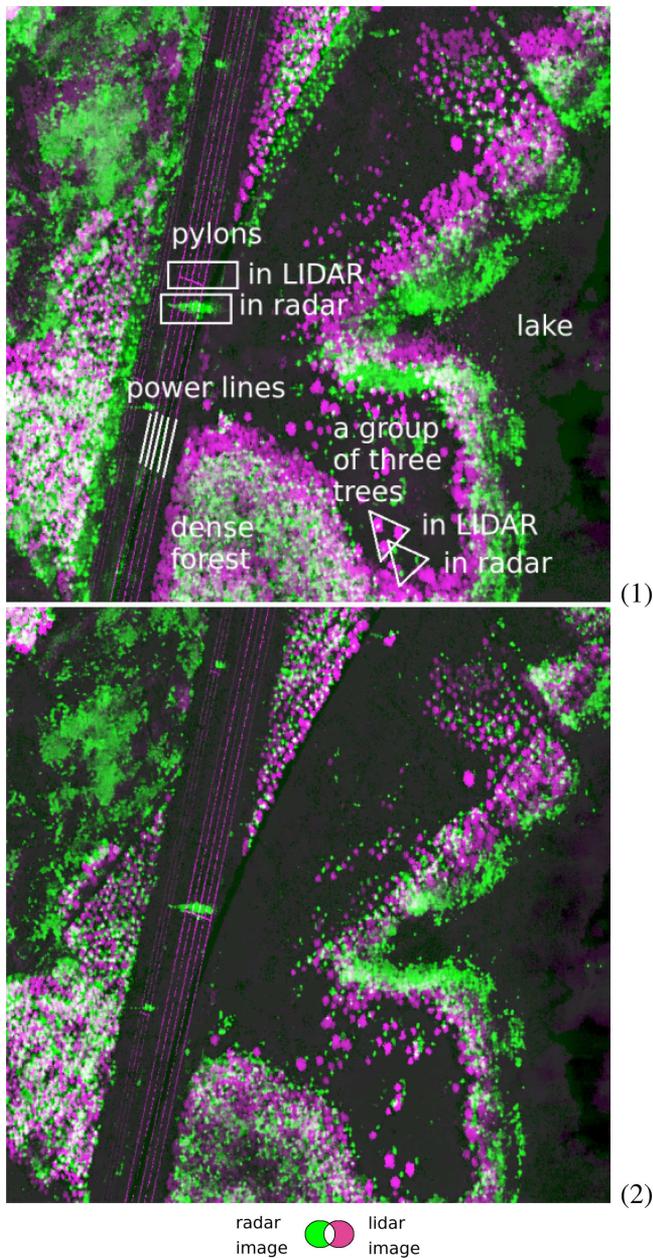


Fig. 8. Result of coregistration for DSM LIDAR data on a  $1000 \times 1000$  pixel SAR image sample over Remningstorp forest (Sweden) (1) after coregistration by geocoding using SRTM, (2) after coregistration by geocoding using GSD50 + DEM.

in HH-VV polarization at further range cells. Due to its vertical illumination, the LIDAR return of one tree is a single patch. These effects are illustrated on Fig. 10. Note that for this reason, we perform the computation of the flow by using the radar image calculated in HH-VV polarization, in order to make the double bounce echo of the ground matching to the LIDAR return echo.

Concerning the baseline algorithm based on mutual information, we notice that electrical cables are deformed. Two different image samples represented in Fig. 11 show that tree superposition is good for both **GeFolki** and baseline algorithm; however, cables are not straight any more in the case of the coregistration by baseline algorithm.

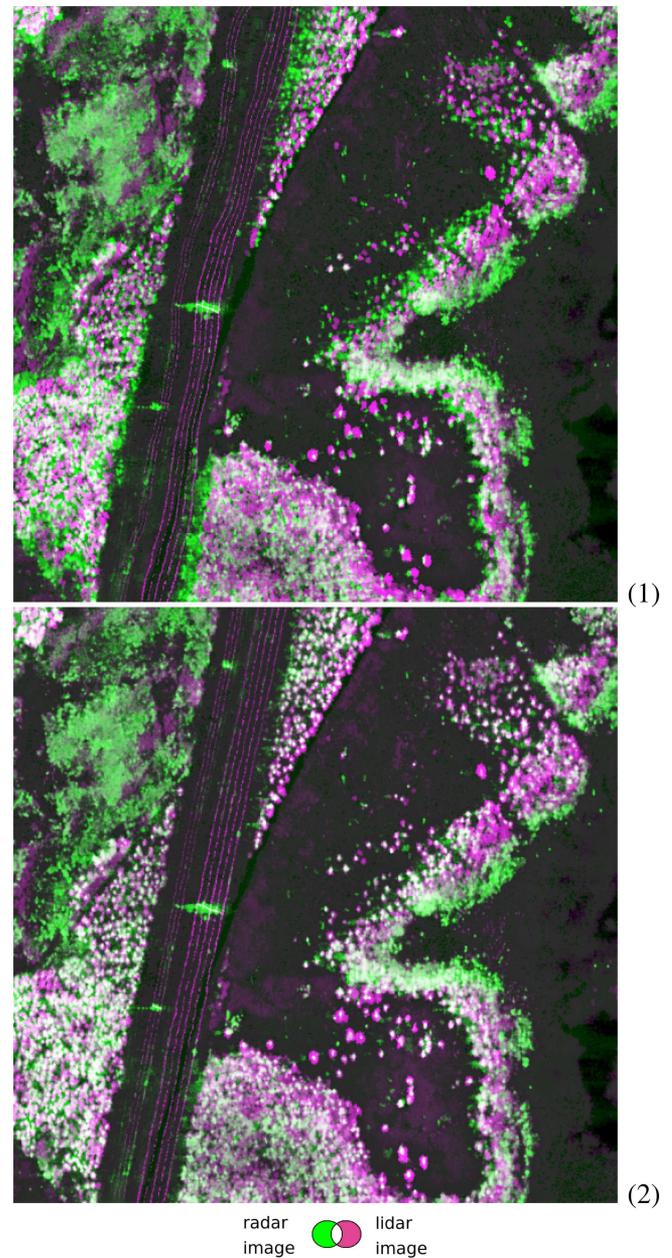


Fig. 9. Result of coregistration for DSM LIDAR data on a  $1000 \times 1000$  pixel SAR image over Remningstorp forest (Sweden) (1) after coregistration by an algorithm based on Mutual Information, (2) after coregistration by **GeFolki**.

The execution time of the baseline method in MATLAB is 248 s, while **GeFolki** needed only 7 s for the same images ( $1000 \times 1000$  pixels).

#### D. Optical and SAR Images

Our algorithm **GeFolki** is now applied for the registration of radar and optical images. The optical image used in this test is distributed by Lantmäteriet. Several areas of the scene present intensity contrasts that differ between the SAR image and the optical one, as we will show later in Fig. 16. Forests are dark patches in the optical one, and bright in the radar images. Water surfaces are dark in both cases. The intensity of the soil varies

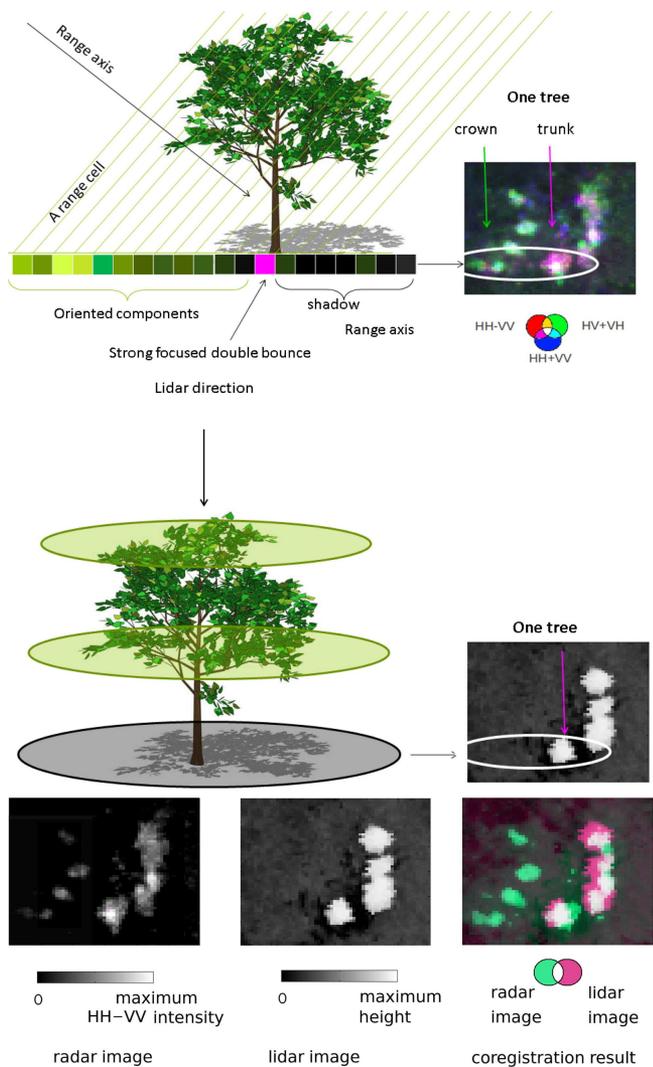


Fig. 10. Effects of geometry on the signature of a tree in the scene for polarimetric radar images and for LIDAR image. Example on a very small  $30 \times 40$  pixel extract of the different products containing four trees: polarimetric SAR image on the left, DSM LIDAR image on the middle, and superposition after our best coregistered product on the right.

greatly in both images but not necessarily in the same way for the radar and for the optical sensors. In this case, success of **GeFolki** lies mainly in the inclusion of the local contrast inversion.

Optical and SAR images now imply even more different geometries. The optical scene is viewed from a given camera position where the main source of light is the sun. In the case of our images, the camera is almost at the top of the tree, the sun is low in the sky and consequently the shadows are relatively large. The optical signature of a tree will consist of the crown, which partially hides the foot of the trunk, and the shadow. The crown has little contrast with the ground; it is especially the shadow that is visible in the case of isolated trees, as seen as in Fig. 12 on the top.

Also, in order to create a visualization of the differences between both radar and optical images, we first reverse the contrast of the optical image before assigning the magenta color channel in the output image. The green color is assigned to the

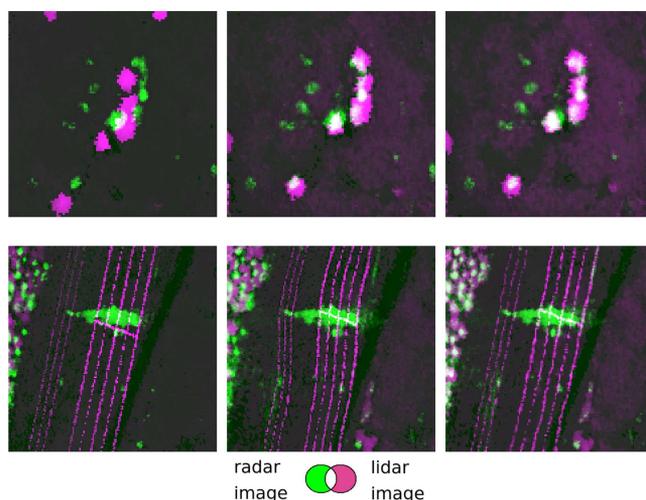


Fig. 11. Result of coregistration for LIDAR data and SAR image over Remningstorp forest (Sweden) for two  $100 \times 100$  pixel extracts 1) after coregistration by geocoding, 2) after coregistration by the baseline algorithm based on mutual information, and 3) after coregistration by **GeFolki**.

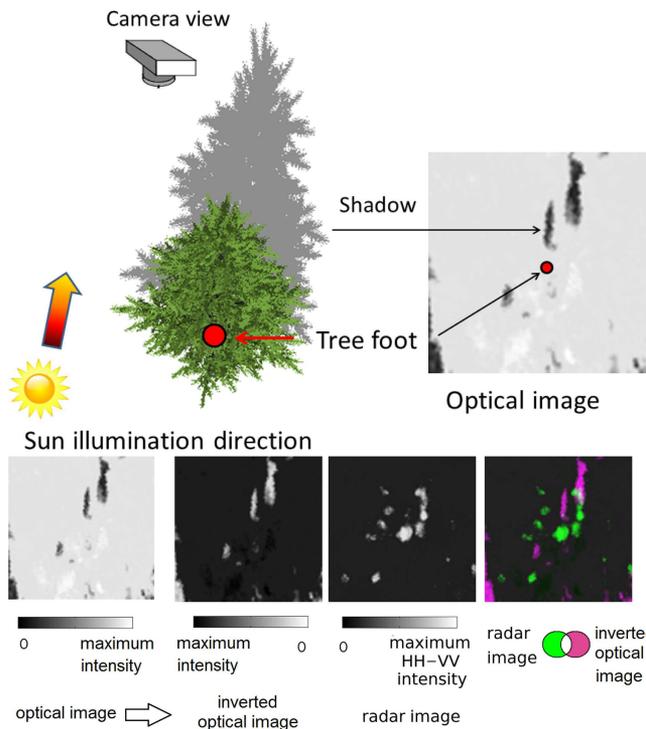


Fig. 12. Effects of geometry on the signature of a tree in the scene for the optical image. Example on a very small  $50 \times 50$  pixel extract of the different products containing some trees: initial optical image on the left, then inverted optical image, SAR image, and superposition after our best coregistered product on the right.

radar image in the final composite visualization. In the case of a perfect coregistration, the colored overlay makes us see the optical shadow in magenta, whose direction crosses the radar double bounce echo, colored green. The choice of the colors is explained in Fig. 12.

Using the same procedure as in the case of LIDAR to coregister the optical image with the SAR image, we obtain the different

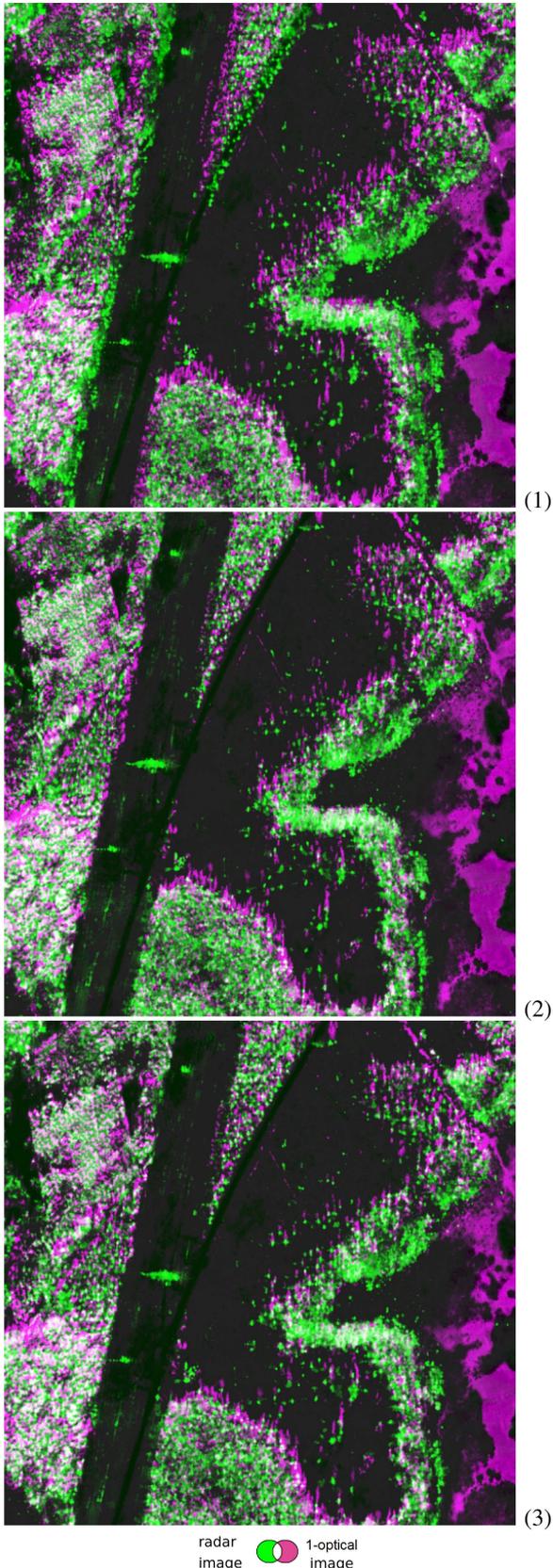


Fig. 13. Result of coregistration of an optical image on a  $1000 \times 1000$  pixel SAR image over Remningstorp forest (Sweden) 1) after coregistration by geocoding, 2) after baseline coregistration based on mutual information, and 3) after coregistration by **GeFolki**.

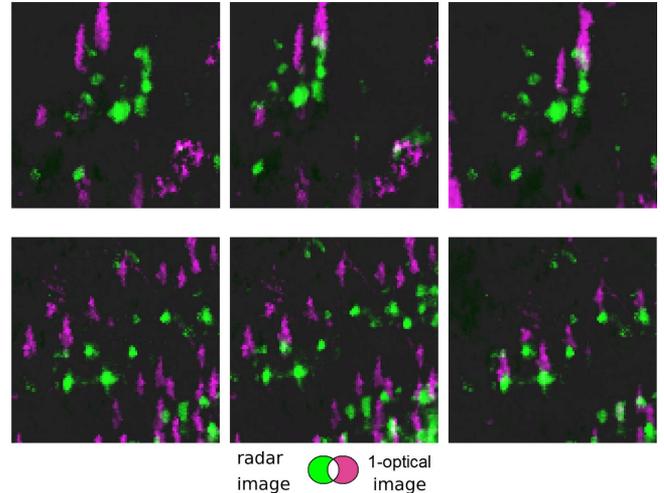


Fig. 14. Result of coregistration for optical data and SAR image over Remningstorp forest (Sweden) for two  $100 \times 100$  pixel samples 1) after coregistration by geocoding, 2) after coregistration by an algorithm based on mutual information, and 3) after coregistration by **GeFolki**.

results presented in Fig. 13 for geocoding, baseline registration, and **GeFolki**.

Once again, the coregistration algorithm applied to remote sensing heterogeneous images results in an interesting product in terms of spatial accuracy. The algorithm based on mutual information also improves the coregistration results in most areas, albeit unevenly, as shown in Fig. 14. On the first extract, both baseline registration and **GeFolki** improve the matching of trees, whereas in the second case, only **GeFolki** is successful. On the  $1000 \times 1000$  image, coregistration by **GeFolki** requires 22 s, whereas baseline coregistration by mutual information is done in 263 s.

Let us have a look at another example of the devil in the details. A mosaic image of a result of coregistration obtained by geocoding is given on the top of Fig. 15. Then, we can see on the same figure below how the flow computation through **GeFolki** and a rough manual initialization, improves the geocoding coregistration on some details. Examples of areas where the coregistration is better done by **GeFolki** are highlighted by yellow ellipses.

#### IV. ROBUSTNESS AND PARAMETER SETTING ANALYSIS

Several parameters have to be set for the whole coregistration process: the window size parameter for the local inversion contrast, and the parameters used for the calculation of the flow. We study these two kinds of parameters independently, to evaluate their impact on the quality of results.

##### A. Discussion About the Local Contrast Inversion

The first point is that for the coregistration of optical and radar images, the algorithm fails to estimate the flow without the local contrast inversion. Indeed, estimation of the flow without the contrast inversion function diverges, even if a good initialization is done. Therefore, this function is crucial in the algorithm.

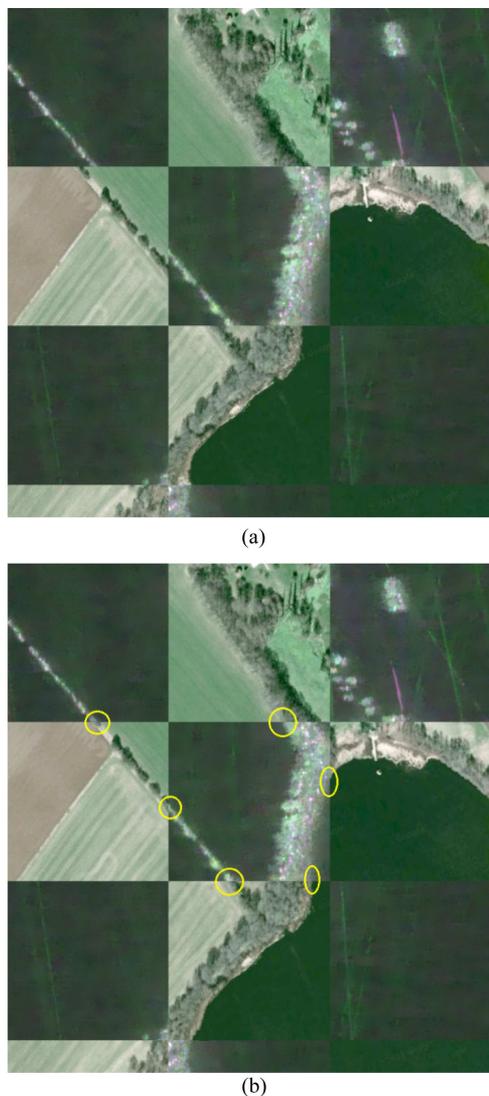
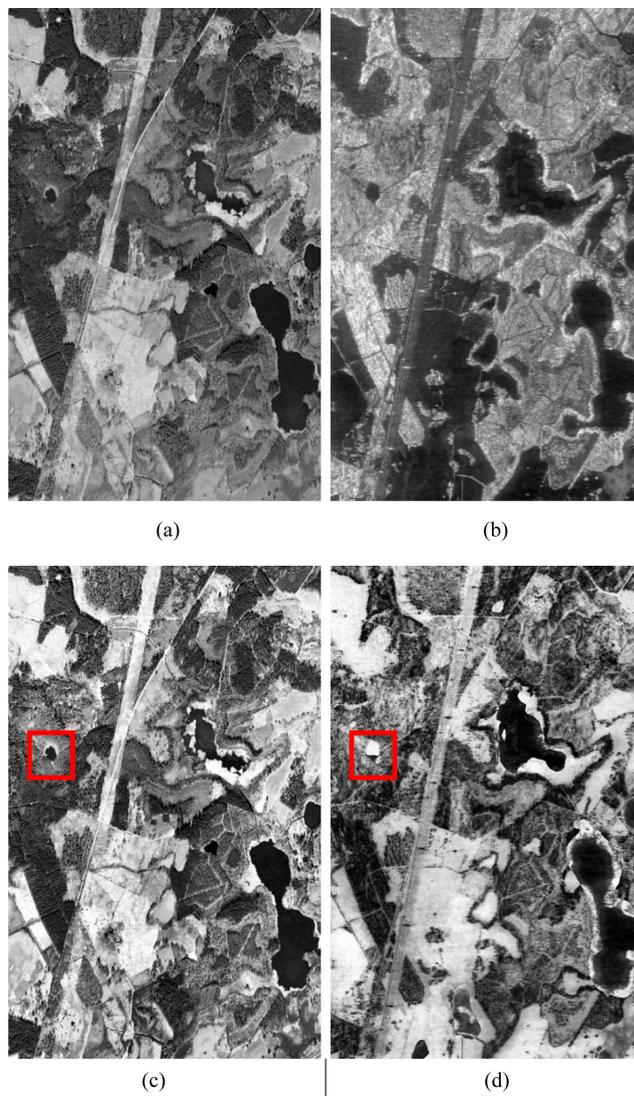


Fig. 15.  $1600 \times 1500$  pixel mosaic of the polarimetric SAR image and the RGB aerial photography. (A) by geocoding, (B) by **GeFolki**.

The criterion for inversion depends on the size of a local window. Therefore, we must keep in mind that this size is the outcome of a compromise between

- 1) The scale of structures offering contrasts that we want to correct: the size of the window has to be in the same order than the size of these structures
- 2) The flow to estimate: when a significant displacement between the two images exists, the inversion process must be performed at a larger scale than this displacement, otherwise the algorithm may fail on the transition areas.

Ultimately, through several tests on different images, we can confirm what we expect intuitively: similar to the flow calculation, the contrast inversion is very effective when conducted in a multiscale manner. At a given pyramid level, and for window size equal to that of the local window  $\omega$  in (3), the contrast inversion is conducted at a scale large enough compared to the flow and, therefore, is always carried out with an appropriate setting.



Scale 1:8, window 17x17

Scale 1:8, window 17x17

Fig. 16. Top left (A): the Red Component of the optical image - Top right (B): the SAR image in HH channel. Bottom left (C): the Red Component of the optical image after local inversion procedure - Bottom right (D): the SAR image in HH channel after contrast enhancement.

To illustrate this, we take a sample of our optical and radar images. The initial image pair is presented in Fig. 16, simply in native gray scale formats.

At the bottom, we show intermediate results of contrast inversion, obtained for images that have been scaled down by a scale factor 8 through a dyadic pyramid reduction of three levels ( $2^3 = 8$ ). The images have been successfully transformed by the local contrast inversion, for a radius equal to 8, that corresponds to a  $17 \times 17$  local window ( $2r + 1 \times 2r + 1$ ). Globally, the obtained images are very similar in terms of intensity levels. Some small areas are still different in contrast, such as the small lake highlighted by red rectangles. Nevertheless, these images depict the solution at an intermediate stage and when we apply the inversion algorithm to images at the initial resolution,

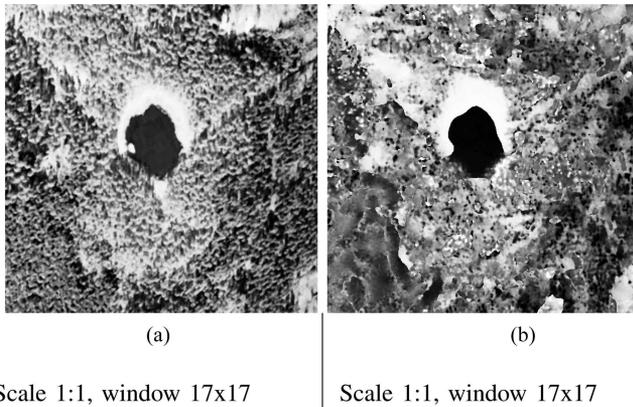


Fig. 17. Zoom of contrast inversion on images at initial resolution. (a) (C) optical Scale 1:1 window  $17 \times 17$ . (b) (D) Radar after contrast inversion. Scale 1:1 window  $17 \times 17$ .

we will see that the result is effective on these areas, as shown on Fig. 17.

In conclusion, a multiscale strategy is essential for the contrast inversion decision, because this one is robust only for local window large enough, and displacements small enough also, and the multiscale strategy is sufficient to satisfy these two requirements.

### B. *GeFolki* Parameter Settings

Now that the sensibility of the contrast inversion step has been analyzed, we want to discuss the influence of the parameters of **GeFolki** in the case of remote sensing heterogeneous image coregistration. The main features are

- 1)  $r$ , called the radius that describes the size of the local window.
- 2)  $L$ , the number of pyramid levels.
- 3)  $K$ , the number of iterations.

Empirically, the found solutions were not very sensitive to  $K$ . A prior sensitivity analysis was conducted in the case of SAR images in [22]. It was shown in particular that it was preferable to use large radii in order to minimize the impact of the speckle, this latter being a level of texture that provides no useful information for the correlation between images.

However, the SAR images were acquired in very similar conditions. The flow does not vary much locally. In the case where the images are taken with different geometries, the terrain projection effects can induce more local deformation between the two images, and in this case, it could be necessary to use a smaller radius to be more precise.

Still in the case of SAR images in close acquisition configurations, the desired displacements are small. In this case, it is sufficient to reduce the number of pyramid levels to only two. In the other cases investigated in this paper such as heterogeneous images, we encountered situations where even the coarse coregistration in first stage leads to remaining errors of several tens of pixels to be compensated after. In such cases, more pyramid levels are needed to estimate these major shifts.

However, these two parameters, radius and number of levels, are intrinsically linked. If the overall displacement to find is

large, a high number of pyramid levels is required. Moreover, a larger window leads to improved matching robustness but also constrains the motion which can be estimated, what would not be appropriate especially at a coarse resolution. This confirms that the quality of initialization plays a significant role in the successful outcome of **GeFolki**.

Overall, we therefore recommend the following procedure in order to optimize the effectiveness of our flow algorithm

- 1) The initial pyramid level is chosen directly based on the expected residual displacement after initialization. For example, if the flow to estimate remains greater than several tens of pixels, we have to choose a rather high pyramid level number, typically equal to 6, equivalent to a 1:64 scale. On the contrary, if images are already very close after the initialization step, the pyramid level can be only 2.
- 2) To choose relatively a rather large matching window. An effective strategy is to explore several values to ensure robustness, for example a list of 6 radii ranking from 4 to 24 by step of 4.

## V. COREGISTRATION PERFORMANCE

To date, there is no universal method to quantify the accuracy of a registration algorithm for heterogeneous images. Also, we propose a methodology for quantitative assessment of performance. The key concept is to measure distances between paired features in the two registered images. These features will be different according to the image pairs under consideration.

### A. Metrics for LIDAR-SAR Coregistration

The main idea is to detect trees in both images, to match them, and to measure the distance in pixels for each defined pair. This pixel-wise distance can be converted into physical distance. Depending on the number of points considered, it is also possible to carry out a statistical analysis of the results. Given the nature of our images, we propose to calculate performance by considering the positioning on the tree bases on the ground.

In practice, we proceed in three steps, described in Fig. 18

- 1) First in the LIDAR image, we extract the individual positioning of each tree. This positioning of individual trees from the LIDAR DSM is performed using the package *rL-iDAR* [38]. The two processing steps are a mean smoothing in order to eliminate spurious local maxima caused by tree branches then nonmaximum suppression within 2 m distance. Both functions have been used with a fixed  $5 \times 5$  windows.
- 2) We extract in the polarimetric radar image the isolated double bounce echoes. This detection is carried out starting from intensity image polarization HH-VV. A first simple threshold detection is performed on this image. The binary mask thus created is dilated by using a 3 radius disk structuring element object. We then consider all connected components found after this dilatation. We keep among these components, only those whose number of pixels does not exceed 100. Otherwise, we suspect several trees to be aggregated. For each detection area, the

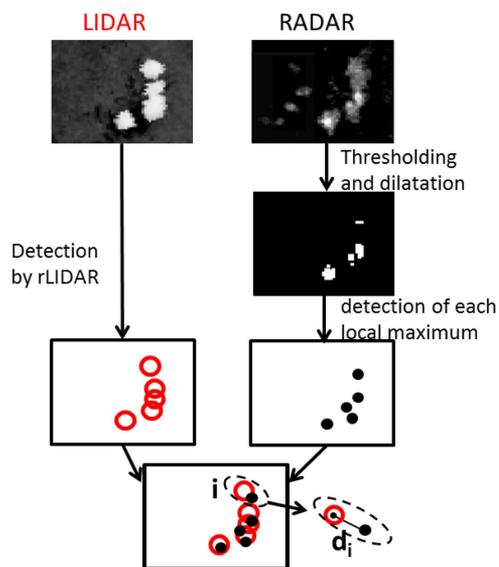


Fig. 18. Different steps to evaluate metrics on LIDAR-RADAR coregistration results.

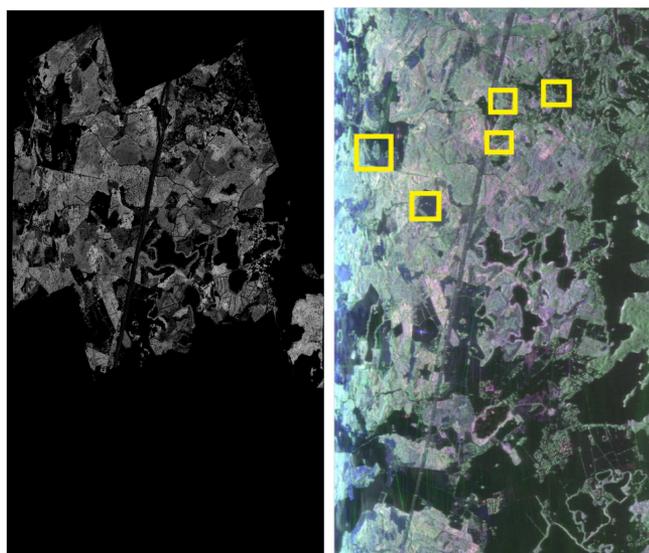


Fig. 19. Selection of areas for metrics in coregistration of LIDAR and SAR images- Left: the LIDAR image, Right: the 12 250 × 7000 pixel polarimetric image with areas selected for our metric highlighted by yellow rectangles.

coordinates are calculated at the maximum intensity pixel. If several adjacent pixels have the same peak intensity, we take the center of gravity of these pixels.

- 3) The last step is to search for each trunk detected in the master image, position of its correspondent, supposed the nearest in the second image. Finally, we calculate the distance between the master pixel position and its corresponding position.

The LIDAR detection performance seems very good, while radar detection fails once the trunk density is high. For this reason, we have manually selected areas where detection results are good. It is also possible to automatically select the areas where the tree densities found in the LIDAR data are low.

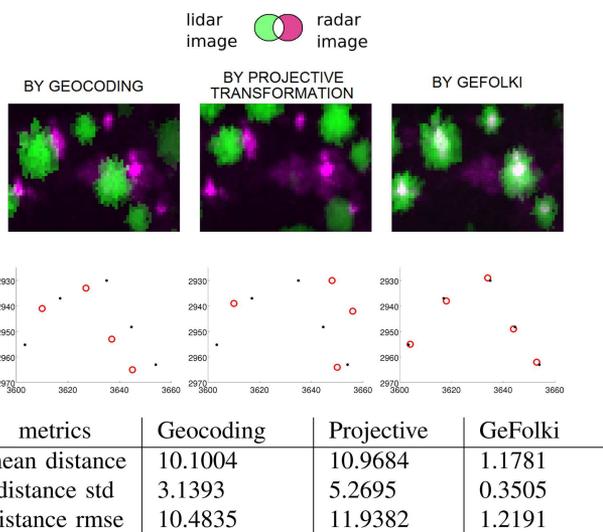


Fig. 20. Performance evaluation for area 1.

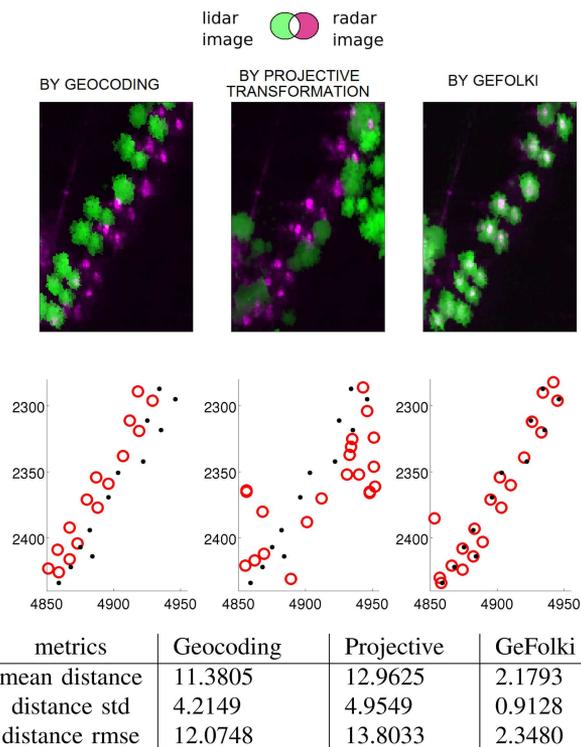


Fig. 21. Performance evaluation for area 2.

Five areas selected are indicated on Fig. 19 on the radar image on the right.

From Figs. 20 to 24, we present the results relative to these areas. For each of them, we compare the results in terms of distance, after a coregistration performed

- 1) By geocoding following the scheme described in Fig. 6.
- 2) After a simple projective transformation based on four Ground Control Points, as the first step of the scheme described in Fig. 7. This result constitutes the initialization of **GeFolki**.

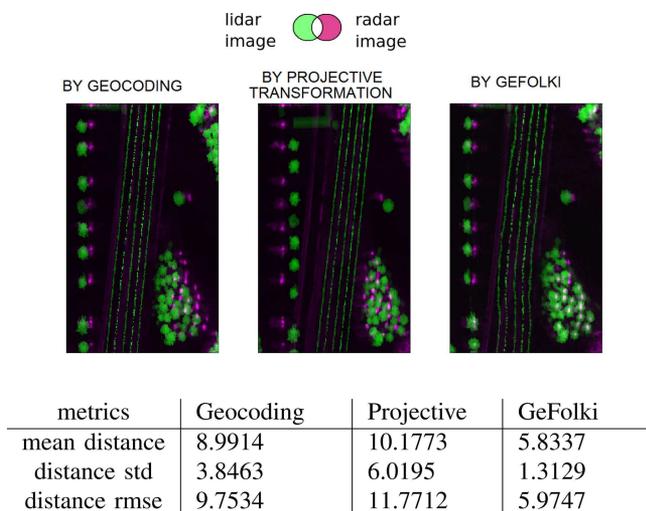


Fig. 22. Performance evaluation for area 3.

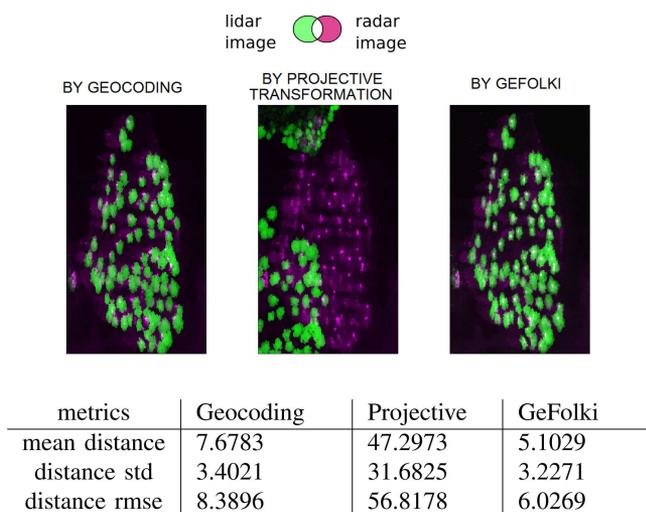


Fig. 23. Performance evaluation for area 4.

- 3) After coregistration of this latter product by **GeFolki**, as the final product described in Fig. 7.

Each image is as follows:

- 1) The first line shows the colored compositions of the LIDAR and radar images: one codes the magenta channel, and the other the green channel. This representation allows us to visualize the effect of any mismatch between images.
- 2) The second line for Figs. 20 and 21 indicates the trees location of the LIDAR image (red circles) and the radar image (black dots).
- 3) The last line is a table where we summarize the mean distance, the standard deviation, and the root mean square error obtained for all detected trunks.

Thus, Fig. 20 concerns a simple case of five well isolated trees whose LIDAR or radar signatures correspond to easily identifiable echoes. Clearly on the colored composition, we see that **GeFolki** allows to match the radar echoes with the crown that can be seen in the LIDAR image. Quantitatively,

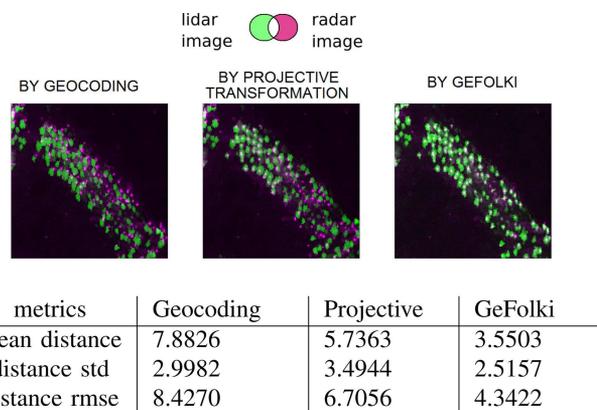


Fig. 24. Performance evaluation for area 5.

the geocoding method brings up a residual gap between the two images of about 10 pixels. On the image obtained by projective transformation, the calculated distances are of the same order of magnitude. However, in this case, it is very likely that the error is largely underestimated: the found matches are not correct because the gaps between trees are less than the error of existing offset in this area. Finally, **GeFolki** leads to a good match with a root mean square error around one pixel. The error is below the expected accuracy of our trunk detection chain.

On the second example represented on Fig. 21, we see a row of trees arranged in pairs. The order of magnitude of error is the same than previously.

However, we can see in this case that our evaluation can be more pessimistic for a precise algorithm than the real quality of coregistration: in fact, some trees are not detected in the radar image. For these missing trees, the distance measured on the LIDAR image will be done with the nearest tree, but not the true corresponding one. This will tend to increase the average distance calculated for **GeFolki**. On the contrary, this matching error has the opposite effect for the projective case: in this case, the nearest tree is not the real one and the metrics leads to underestimate the error, as in Fig. 21.

Fig. 22 is an example chosen along power lines. On the left there is one row of trees, and on the right side we see a group of trees. The evaluation leads to a slightly worse score for **GeFolki** than in the previous examples, but still better than the other two methods. Visually, it is mostly trees arranged on the left that makes them slightly less good statistics.

Another example of grove is given in Fig. 23. It is located in a part of the image very far from the previous one, at a distance of about 5 km. In this area, the projective transformation fails completely. Geocoding and **GeFolki** have equivalent statistics, slightly in favor of the latter, in terms of root mean square error about 8 pixels for Geocoding and 6 pixels for **GeFolki**. Note that in this example, the displacement obtained after the simple projective model is rather strong, and **GeFolki** can still recover from such an inaccurate initialization.

A final example with more trees is shown in Fig. 24, still in favor of the **GeFolki** method.



Fig. 25. Selection of areas for metrics in coregistration of optical and radar images- Left: the polarimetric image, Right: the optical image with seven pairs of pylons highlighted by yellow rectangles.



Fig. 26. Example of a common pylon selection: on the polarimetric SAR image on the left, on the optical image coregistered by **GeFolki** in the middle, and coregistered by geocoding on the right. Red circles indicate the position of the pylon bases in the SAR master image.

### B. Metrics for Optics-SAR Coregistration

For the optical images, the detection of trees is much more difficult to handle. For this reason, we first go through a manual selection of some reference points, at the foot of the various pylons detectable in the images. Most of them are found in pairs. In this case, the coordinates of each of them are then averaged. We then get seven coordinates, highlighted by yellow rectangles in Fig. 25, each corresponding to a pair of pylons. Note that in the case of radar, the echo at the foot of the tower is very strong, as can be seen in Fig. 26. Furthermore, the selection of the point is further facilitated by the polarimetric color signing the double bounce mechanism. Contrariwise, the position of the pylon base is more uncertain in the optics image, because these objects provide little contrast with the ground. Despite these difficulties, we are able to locate these targets which are well distributed throughout the image extent; furthermore echoes are sufficiently focused.

Then, we calculate for each pylon the difference between the coordinates of the radar and the optical image. This procedure is applied on the optical image coregistered by geocoding, and the one coregistered by **GeFolki**. The results are given in Fig. 27. We can note that **GeFolki** presents the best results.

Finally, we also looked at the errors of coregistration on other types of objects: the outline of the lakes and some rows of

Geocoding	GeFolki
Position difference in pixel	Position difference in pixel
17.61	4.95
19.42	6.96
21.21	5
12.04	7.16
18.90	1.80
15.53	0
12.74	3.53
Mean	Mean
16.78	4.20
STD	STD
3.47	2.63
RMSE	RMSE
17.08	4.86

Fig. 27. Examples of flow errors at reference points.

trees. We performed manual reading distances in the six areas circled in yellow on the mosaic of Fig. 15. Four distances are recorded at the transition on rows of trees, and two distances are measured on the edges of the lake. These distances lead to a root mean square error of 18.5 pixels in the case of data coregistered by geocoding, and 0.8 pixels in the case of data coregistered by **GeFolki**.

In summary, the **GeFolki** method provides more accurate results than geocoding, with pixel-precision, compared to the ten pixels from our method of geocoding.

## VI. CONCLUSION

This paper introduces the use of an optical flow algorithm called **GeFolki** to register heterogeneous remote sensing images. This consists in a dense flow method derived from the Lucas–Kanade algorithm, with a multiscale strategy, applied to images transformed by a rank filtering, and, for heterogeneous images, also by a rolling guidance filter and a contrast inversion decision method.

The algorithm has been applied and qualitatively compared to coregistration made using only geocoding information, and also to an alternative method of baseline registration based on the use of mutual information. Data considered are airborne forest land images with resolutions down to the meter for radar, 50 cm for optical image, 10 m or 50 cm for LIDAR data. **GeFolki** has been shown to be

- 1) More accurate than geocoding on all our examples.
- 2) Free from some defects such as deformation induced by the baseline registration.
- 3) More robust than this same method with respect to the initialization.

Quantitative performances have been derived, both for a coregistration based on geocoding, and for **GeFolki**. Performances are based on detection of trees for the SAR/LIDAR coregistration, and on manual measurements on some pylons, tree rows and lake edges for the SAR/optical coregistration.

The registration precision of **GeFolki** is about a few pixels, (1 pixel = 80 cm), compared to ten of pixels by geocoding. This precision opens new opportunities for studies related to data fusion, for example, with applications involving the counting of trees. It appears that the fusion of different images may improve detection performance, but only if we can trust the accuracy of registration.

Finally, **GeFolki** has been proven to be faster than the baseline registration method, by a factor of 10 to 20.

Our next efforts will focus on

- 1) Taking into account the several bands of multivariate images, as opposed to a panchromatic image or single polarization radar image. Indeed, Hyperspectral or LIDAR data contains several bands having different contrasts, as well as the polarimetric radar images with the different polarimetric channels. To date, the choice of the band has not been reasoned. But a work could be undertaken on how to make the most of this data diversity.
- 2) The possible extension to more complex scenarios, for example with two images containing mountainous terrains or buildings, and acquired with different headings. This case is particularly difficult because these objects will be projected on the ground differently following the acquisition mode.

#### ACKNOWLEDGMENT

All data used in this paper come from various institutions:

- 1) LIDAR data have been provided by Johan Fransson from SLU, Swedish University of Agricultural Sciences, and Lars Ulander, from FOI, Swedish Defence Research Agency, for whom we are deeply grateful. All LIDAR images have been processed by SLU.
- 2) Polarimetric SAR dataset over Remningstorp is provided from the BIOSAR 2010 campaign funding by ESA, acquired by Onera.
- 3) The optical image over Sweden has been provided by Lantmäteriet.

One of the authors, G. Brigot, is a Ph.D. student jointly supervised by E. Koeniguer in Onera and M. Simard in JPL. His Ph.D. degree is supported jointly by TOTAL and Onera. The authors are thankful to M. Davidson and M. Fomelis from ESA-ESRIN for their very informative recommendations. We are also immensely grateful to G. Le Besnerais for his comments on an earlier version of the manuscript, although any errors are our own. Finally, we warmly thank C. Smetanka for her careful reading of the manuscript.

#### REFERENCES

- [1] J. D. Wegner, S. Nezam, S. Müller, and U. Sörgel, "Comparison of land cover classification using high-resolution terrasars-x and optical imagery," in *Proc. 8th Eur. Conf. Synthetic Aperture Radar*, 2010, pp. 1–4.
- [2] J. Qi, C. Wang, Y. Inoue, R. Zhang, and W. Gao, "Synergy of optical and radar remote sensing in agricultural applications," in *Proc. 48th Annu. Meeting Opt. Sci. Technol.*, 2004, pp. 153–158.
- [3] P. Hyde, R. Nelson, D. Kimes, and E. Levine, "Exploring LiDAR RaDAR synergy predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR and InSAR," *Remote Sensing Environ.*, vol. 106, no. 1, pp. 28–38, 2007.
- [4] S. Kaasalainen *et al.*, "Combining Lidar and synthetic aperture radar data to estimate forest biomass: Status and prospects," *Forests*, vol. 6, no. 1, pp. 252–270, 2015.
- [5] C. Paris and L. Bruzzone, "A three-dimensional model-based approach to the estimation of the tree top height by fusing low-density Lidar data and very high resolution optical images," *IEEE Trans. Geosci. Remote Sensing*, vol. 53, no. 1, pp. 467–480, Jan. 2015.
- [6] T. G. Yitayew, C. Brekke, and A. P. Doulgeris, "Multisensor data fusion and feature extraction for forestry applications," in *Proc. IEEE Int. Geosci. Remote Sensing Symp.*, 2012, pp. 4982–4985.
- [7] T. Hame *et al.*, "Improved mapping of tropical forests with optical and SAR imagery, part i: Forest cover and accuracy assessment using multi-resolution data," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sensing*, vol. 6, no. 1, pp. 74–91, Feb. 2013.
- [8] A. Schubert, D. Small, M. Jehle, and E. Meier, "COSMO-SkyMed, TerraSAR-X, and RADARSAT-2 geolocation accuracy after compensation for earth-system effects," in *Proc. IEEE Int. Geosci. Remote Sensing Symp.*, 2012, pp. 3301–3304.
- [9] B. Jian, B. C. Vemuri, and J. L. Marroquin, "Robust nonrigid multimodal image registration using local frequency maps," in *Proc. 19th Int. Conf. Inf. Process. Med. Imag.*, 2005, pp. 504–515.
- [10] T. Gaens, F. Maes, D. Vandermeulen, and P. Suetens, "Non-rigid multimodal image registration using mutual information," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 1998, pp. 1099–1106.
- [11] L. L. Coulter, D. A. Stow, C. D. Lippitt, and G. W. Fraley, "Repeat station imaging for rapid airborne change detection," in *Time-Sensitive Remote Sensing*, C. D. Lippitt, D. A. Stow, and L. L. Coulter, Eds. New York, NY, USA: Springer, 2015, pp. 29–43. [Online]. Available: [http://dx.doi.org/10.1007/978-1-4939-2602-2\\_3](http://dx.doi.org/10.1007/978-1-4939-2602-2_3)
- [12] A. Wong and D. Clausi, "ARRSI: Automatic registration of remote-sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 45, no. 5, pp. 1483–1493, May 2007.
- [13] Z. Li and J. Wang, "Least squares image matching: A comparison of the performance of robust estimators," *ISPRS Ann. Photogrammetry, Remote Sensing Spatial Inf. Sci.*, vol. 1, pp. 37–44, 2014.
- [14] H. Goncalves, L. Corte-Real, and J. Goncalves, "Automatic image registration through image segmentation and SIFT," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 7, pp. 2589–2600, Jul. 2011.
- [15] B. Fan, C. Huo, C. Pan, and Q. Kong, "Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT," *IEEE Geosci. Remote Sensing Lett.*, vol. 10, no. 4, pp. 657–661, Jul. 2013.
- [16] Z. L. Song and J. Zhang, "Remote sensing image registration based on retrofitted SURF algorithm and trajectories generated from lissajous figures," *IEEE Geosci. Remote Sensing Lett.*, vol. 7, no. 3, pp. 491–495, Jul. 2010.
- [17] Q. Shi *et al.*, "Robust image registration using structure features," *IEEE Geosci. Remote Sensing Lett.*, vol. 11, no. 12, pp. 2045–2049, Dec. 2014.
- [18] S. Cao, X. Zhu, Y. Pan, and Q. Yu, "A stable land cover patches method for automatic registration of multitemporal remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sensing*, vol. 7, no. 8, pp. 3502–3512, Aug. 2014.
- [19] H.-M. Chen, M. K. Arora, and P. K. Varshney, "Mutual information-based image registration for remote sensing data," *Int. J. Remote Sensing*, vol. 24, no. 18, pp. 3701–3706, 2003.
- [20] T. Hu, H. Zhang, H. Shen, and L. Zhang, "Robust registration by rank minimization for multiangle hyper/multispectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sensing*, vol. 7, no. 6, pp. 2443–2457, Jun. 2014.
- [21] C. Vogel, A. Bauder, and K. Schindler, "Optical flow for glacier motion estimation," *ISPRS Ann. Photogrammetry, Remote Sensing Spatial Inf. Sci.*, vol. 1, pp. 359–364, 2012.
- [22] A. Plyer, E. Colin-Koeniguer, and F. Weissgerber, "A new coregistration algorithm for recent applications on urban SAR images," *IEEE Geosci. Remote Sensing Lett.*, vol. 12, no. 11, pp. 2198–2202, Nov. 2015.
- [23] B. K. Horn and B. G. Schunck, "Determining optical flow," in *Proc. 1981 Tech. Symp. East*, 1981, pp. 319–331.
- [24] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, vol. 81, pp. 674–679.
- [25] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunck: Combining local and global optic flow methods," *Int. J. Comput. Vis.*, vol. 61, no. 3, pp. 211–231, 2005.

- [26] C. Zach, T. Pock, and H. Bischof, "A duality based approach for real-time tv-l1 optical flow," in *Proc. 29th DAGM Conf. Pattern Recog.*, 2007, pp. 214–223. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1771530.1771554>
- [27] A. Plyer, G. Le Besnerais, and F. Champagnat, "Massively parallel Lucas Kanade optical flow for real-time video processing applications," *J. Real-Time Image Process.*, vol. 11, pp. 713–718, 2014.
- [28] F. Champagnat *et al.*, "Fast and accurate PIV computation using highly parallel iterative correlation maximization," *Experiments Fluids*, vol. 50, no. 4, pp. 1169–1182, 2011.
- [29] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 815–830.
- [30] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. Eur. Conf. Comput. Vis.*, 1994, vol. 801, pp. 151–158.
- [31] G. Le Besnerais and F. Champagnat, "Dense optical flow by iterative local window registration," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2005, vol. 1, pp. 137–140.
- [32] G. Bonin and P. Dreuillet, "The airborne SAR-System: SETHI - Airborne microwave remote sensing imaging system," in *Proc. 7th Eur. Conf. Synthetic Aperture Radar*, Jun. 2008, pp. 1–4.
- [33] "GSD-Elevation data, grid 50+ hdb". (2015). [Online]. Available: [http://www.lantmateriet.se/globalassets/kartor-och-geografisk-information/hojddata/produktbeskrivningar/eng/e\\_grid50\\_plus\\_hdb.pdf](http://www.lantmateriet.se/globalassets/kartor-och-geografisk-information/hojddata/produktbeskrivningar/eng/e_grid50_plus_hdb.pdf)
- [34] I. Pétillet *et al.*, "Radar-coding and geocoding lookup tables for the fusion of GIS and SAR data in mountain areas," *IEEE Geosci. Remote Sensing Lett.*, vol. 7, no. 2, pp. 309–313, Apr. 2010.
- [35] L. Ulander *et al.*, "BIOSAR 2010 technical assistance for the development of airborne SAR and geophysical measurements during the BioSAR 2010 experiment; final report," Eur. Space Agency, Paris, France, Tech. Rep. Contract No. 4000102285/10/NL/JA/ef, 2011.
- [36] D. J. Kroon and C. H. Slump, "MRI modality transformation in demon registration," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano Macro*, Jun. 2009, pp. 963–966.
- [37] J.-P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.
- [38] C. A. Silva, N. L. Crookston, A. T. Hudak, L. A. Vierling, and M. C. A. Silva. (2015). Package rlidar. [Online]. Available: <https://cran.r-project.org/web/packages/rLiDAR/index.html>



**Guillaume Brigot** received the Graduate degree from the French Supélec High Engineer School, Paris, France, in 2014. He is currently working toward the Ph.D. degree on topic of the fusion of SAR and LIDAR images to estimate structural parameters of forest.

He has already completed two internships at JPL in the field of remote sensing of forests: A first internship during summer 2014 about POLINSAR Tandem-X data analysis on forests of Quebec, with Maxim Neumann, and a second internship on LI-

DAR data analysis with Marc Simard in 2015. He is currently directed by Elise Koeniguer in Onera and Marc Simard in JPL.



**Elise Colin Koeniguer** received jointly the Dipl.Ing. degree in electrical engineering from École Supérieure d'électricité (Suplec), Gif-Sur-Yvette, France, and the M.Sc. degree in theoretical physics from the University of Orsay (Paris XI), Orsay, France, in 2002, and the Ph.D. degree from the University of Paris VI, Paris, France, in 2005. She received the Habilitation from the University of Paris Saclay, France, in 2014.

Then, she joined the Electromagnetism and Radar Department, French Aerospace Laboratory (ONERA), Palaiseau, France. Since 2013, she has been with the Modeling and Information Processing Department, ONERA, where her activities include now computer vision. Her research interests include various subjects in image processing for remote sensing, comparison between radar polarimetry and optical polarimetry, big data for Earth Observation.



**Aurélien Plyer** received the Ph.D. degree in signal processing and image from the University Paris 13, Villetaneuse, France, in 2013.

He is interested in issues of computer vision or how geometrical information can be obtained from a signal registered by camera systems. Among other he contributed to the creation of the eFolki algorithm, a solution of measurement of the optical flow to reach a compromise between accuracy/robustness/cost calculation unmatched in the literature. This algorithmic work has a large number of applications whether for measuring fluid flow in the wind tunnel on tightly controlled measures or for perception of autonomous robotics in uncontrolled environment. It also included the super-resolution video application which allows via the analysis of movement to reverse the sampling pattern of a digital camera to estimate a higher resolution than the signal provided by the sensor.



**Fabrice Janez** received the Ph.D. degree in science from the University of Angers, Angers, France, in 1996 on the subject of data fusion.

Since 1997, he has joined the Department of Information Processing and Modeling, French National Aerospace Research Center, ONERA. His last research subjects concern the processing of large remote sensing image and the contextualization of the applied algorithms.