Performance Analysis of Optical Flow Techniques for Airborne SAR Image Registration

Prithvi Laguduvan Thyagarajan, *Graduate Student Member, IEEE*, Holger Nies, Patrick Berens, and Ivo Ihrke

***Abstract*— Image registration is a crucial step in interferometric synthetic aperature radar (InSAR) and in synthetic aperture radar (SAR) tomography. Generally, a displacement of the sensors, e.g., due to different flight tracks, causes an image distortion that is dependent on the terrain height that is being observed. While ground truth digital elevation map (DEM) data are often available to roughly compensate for this distortion, there are application scenarios where the acquisition paths are too irregular or DEMs are not available. In such cases, image registration via image processing is a suitable choice. While the SAR community prefers patch-based correlation techniques, the computer vision community has investigated the same issue from technically different points of view. In this article, we study the performance of correlation- and computer vision-based image registration techniques for the image registration problem in SAR, in particular in airborne SAR without ground truth. We show that computer vision algorithms can outperform correlation-based techniques.**

***Index Terms*— Airborne synthetic aperture radar (SAR), computer vision, image registration, optical flow.**

I. Introduction

IMAGE registration is the process of matching pairs of images of the same scene obtained with different imaging parameters. Often, the image pair is taken from different positions of a sensor, possibly acquired at different times or even with different sensors [1], [2]. We focus on synthetic aperture radar (SAR) image coregistration that is mandatory in the fields of SAR interferometry (InSAR) [3] and SAR tomography [4], but that is also useful in other applications such as change detection, comparison of data from different acquisitions, etc. Our goal in this article is the comparison of classical correlation-based approaches that are commonly used in the literature [1] with algorithms developed over the last decades in computer vision community that faces similar challenges.

In both, InSAR and SAR tomography, SAR data is captured from different positions and different incident angles of the sensor which results in an apparent shift of pixels in every image with respect to a reference position. In practice, the shift may not be uniform in all areas of the image due to factors such as the influence of the flight track, the height of the targets and the atmospheric conditions while capturing the data.

In spaceborne InSAR constellations (like TerraSAR-X and TanDEM-X) the position and the rough height model of the scene are known a priori, so that the mutual shifts of the pixels can be compensated using this geometry [5]. The remaining mutual shifts are usually detected using cross correlation [6], even though the techniques evaluated in this article may prove useful for this task as well.

Instead, we investigate the more challenging scenario of airborne SAR interferometry, i.e., SAR with data obtained from airplane-mounted SAR sensors because it includes a set of additional challenges. Supposedly, an algorithm that is robust in the more challenging scenario also yields more robust results on simpler problems. For this scenario, the additional challenges include imperfect, nonparallel airplane trajectories as well as poor coherence between images due to a wide variety of incidence angles.

In the computer vision context, the image registration problem, when considered on a per-pixel level, is known as the optical flow problem. Optical flow refers to the visual perception of the movement of objects in a scene, which arises from the motion of either the sensor or an object, or both. Since the literature on the subject is extensive, we refer to the Middlebury benchmark [7] for a comprehensive reference and comparison of different algorithms developed over the course of about four decades of research on the problem in this community. In the field of SAR, some researchers have drawn inspiration from optical flow techniques to analyze high-resolution 3-D scene data by monitoring how the movement of circular SAR affects the energy patterns generated by targets [8], [9].

Even though the benchmark is extensive and has been methodologically improved several times, it is relevant only to photographic visible light images, the conclusions of which may not carry over to other application scenarios [10]. In particular, the benchmark is, at the time of writing, dominated by neural network-based algorithms. These algorithms achieve success by having access to a very large database of ground truth results. This is possible because sensor variation between different photographic cameras and lenses is less pronounced than that found in different SAR sensors. It is therefore

Manuscript received 13 June 2023; revised 14 December 2023; accepted 25 January 2024. Date of publication 23 February 2024; date of current version 8 March 2024. This work was supported by the European Union's Horizon 2020 Research and Innovation Program through the Marie Skłodowska-Curie under Grant 860370. *(Corresponding author: Prithvi Laguduvan Thyagarajan.)*

Prithvi Laguduvan Thyagarajan and Patrick Berens are with the Fraunhofer Institute for High Frequency Physics and Radar Techniques (Fraunhofer FHR), 533343 Wachtberg, Germany (e-mail: prithvi.laguduvan.thyagarajan@fhr.fraunhofer.de).

Holger Nies and Ivo Ihrke are with the Department of Computational Sensorics/Communication Engineering, University of Siegen, 57076 Siegen, Germany.

Digital Object Identifier 10.1109/TGRS.2024.3369402

difficult to apply network-based approaches in the SAR image registration context. We demonstrate this claim by including recurrent all-pairs field transform (RAFT) [11], one of the top-performing network-based methods in our evaluation.

Before the machine learning revolution in computer vision, energy-based optimization approaches were dominating the benchmark. They were modifications of two classical approaches, both published in 1981: the Lucas and Kanade [12] approach is a local window-based gradient descent approach that is lacking regularization, whereas the Horn and Schunck [13] method is a global variational approach including a regularization term, which enables its application to images containing homogeneous regions. Both approaches rely on local linearizations which makes them applicable to small shifts on the order of 1–2 pixels only. Over the years different improvements have been employed, the most important being multiscale processing for handling larger displacements and the avoidance of approximations that were necessary for computational efficiency in the original papers. Given modern implementations, Sun et al. [14] concluded, classical energy-based optimization techniques are competitive in settings where large-scale datasets such as the KITTI dataset [15] are not available. We, therefore, include modern implementations of these algorithms in our evaluation: eFolki [1] is a variant of the Lukas-Kanade algorithm developed for SAR data, whereas a multiscale Horn and Schunck method [16] (H&S) and one with total variation (TV) regularization [17] (TV-L1) are representatives of variational techniques. The latter allows for sharp edges and corners in the estimated displacement fields, which the original method smoothes over. All evaluated algorithms have open source implementations which we have used for evaluation.

The article is structured as follows: for the reader's convenience, we first review the major ideas of the algorithms participating in our evaluation, Section II. We then describe the characteristics of the acquisition scenario, Section III, and the testing methodology, Section IV, which includes both synthetic and real data. Finally, we discuss the results of our extensive tests and derive an outlook for future work in the context of SAR image registration, Section V.

A. Notations

We adopt the notation of using boldface lower case for vectors $\mathbf{a}$ and boldface upper case for matrices $\mathbf{A}$. The transpose and complex conjugate operators are denoted by the symbols $(.)^T$ and $(.)^*$, respectively. $\mathbb{R}^{M\times N}$ is the set of $M \times N$ real matrices. $\mathbb{C}^{M\times N}$ is the set of $M \times N$ complex valued matrices. Where applicable the notation follows the original articles to facilitate an easy comparison. In addition, for a vector $\mathbf{a}$, $E[\mathbf{a}]$ denotes the expectation of $\mathbf{a}$.

II. Coregistration Techniques

In the following, we are reviewing the basic principles behind the algorithms that participate in our evaluation. We mention their principal characteristics which inform the interpretation of the results provided in the experimental Section IV.

A. Correlation Based Registration

The standard method for coregistration in the SAR community is a 2-D normalized cross correlation (NCC) between a reference and one or multiple secondary images [18]. NCC is often applied to complete images or using a block partitioning strategy for images that contain localized distortions. We denote the available reference image as $\mathbf{I}_{\text{ref}} \in \mathbb{C}^{M\times N}$ and the available secondary image as $\mathbf{I}_{\text{sec}} \in \mathbb{C}^{M\times N}$. For our evaluation, we use the magnitude of the reference, $\mathbf{I}_{\text{ref}} \in \mathbb{R}^{M\times N}$, and secondary images, $\mathbf{I}_{\text{sec}} \in \mathbb{R}^{M\times N}$. The SAR image undergoes double oversampling, effectively addressing concerns related to offset estimation issues arising from cross correlation. We implement the algorithm to perform a per-pixel window search for the maximal NCC score. Denoting the window by W and enumerating the pixels (x_i, y_i) in the reference image using the index i, we aim to determine a most likely displacement (u_i, v_i) for each pixel that aligns the window W around $(x_i + u_i, y_i + v_i)$ in the secondary image with that around (x_i, y_i) in the primary image. Each candidate displacement (u_i, v_i) is evaluated using the NCC score

$$\rho(u_i, v_i) = \frac{\frac{1}{N}\sum_{(\Delta x,\Delta y)\in W} \begin{array}{r}[\mathbf{I}_{\text{ref}}(x_i+\Delta x, y_i+\Delta y)-\mu^W_{\text{ref}}]\cdot\\ [\mathbf{I}_{\text{sec}}(x_i+u_i+\Delta x, y_i+v_i+\Delta y)-\mu^W_{\text{sec}}]^*\end{array}}{\sigma^W_{\text{ref}}\sigma^W_{\text{sec}}} \tag{1}$$

where μ^W represents the mean and σ^W the standard deviation of the windows in $\mathbf{I}_{\text{ref}}$ and $\mathbf{I}_{\text{sec}}$; N is the number of pixels inside the window. The highest correlation score indicates the most likely displacement between the two images, i.e., the output is determined by the following equation:

$$\text{argmax}_{(u_i,v_i)}\, \rho(u_i, v_i) \tag{2}$$

for every pixel i. The tunable parameter for correlation-based registration are the window size and the displacement search range.

NCC can be applied to complex-valued SAR images or to real-valued amplitude images [18]. In the case of good SNR, the phase information can be helpful for registration, however, in the case of low SNR it is usually leading to more noisy results [18], [19]. Since our application scenario, Section III, involves large displacements and low SNR, magnitude-only-based correlation is preferred.

We include the technique in our evaluation as the baseline technique for SAR coregistration. In practice, we use the fast implementation of Lewis [20] that is based on real-valued images. We use sub-sampling to achieve sub-pixel accuracy.

B. Interlude: Optical Flow Background

Before discussing the evaluated techniques, we make a few general remarks that apply to all evaluated optical flow methods.

Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image [13]. It can arise from the relative motion of the object and the viewer, and, as a result, can also provide information about the spatial arrangement of the objects viewed and the rate of change of the arrangement [21]. The point of view is slightly shifted in that the reference and the secondary images are considered to be part of a temporal sequence of images: $I(x, y, t) := \mathbf{I}_{\text{ref}}$ and $I(x, y, t + \Delta t) := \mathbf{I}_{\text{sec}}$. Optical flow computation is based on the brightness constancy assumption

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{3}$$

where a continuous motion of image points from (x, y, t) to $(x + \Delta x, y + \Delta y, t + \Delta t)$ is assumed. Similarly, the image I is considered to be a continuous space-time function.

Considering a Taylor expansion of the temporal image sequence to first-order

$$\begin{aligned} I(x + \Delta x, y + \Delta y, t + \Delta t) &= I(x, y, t) \\ &+ \frac{\partial I}{\partial x}\bigg|_{(x,y,t)} \Delta x + \frac{\partial I}{\partial y}\bigg|_{(x,y,t)} \Delta y + \frac{\partial I}{\partial t}\bigg|_{(x,y,t)} \Delta t \end{aligned} \tag{4}$$

and enforcing the brightness constancy constraint, (3), the optical flow constraint is obtained

$$\frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t} = 0. \tag{5}$$

The optical flow is then considered as the vector field $(u(x, y), v(x, y)) := ((\Delta x/\Delta t), (\Delta y/\Delta t))$ that pixel-wise aligns the two images. With these conventions, the optical flow constraint is often written as $I_x u + I_y v + I_t = 0$, which is a linear equation in the unknown displacement vectors (u, v). The linearization is due to the Taylor expansion and restricts the application of the linearized optical flow constraint to small displacements on the order of 1–2 pixels. Typical extensions are therefore multiscale search schemes [14].

C. Lucas-Kanade/eFolki

Lucas and Kanade [12] tracking is a differential technique to compute optical flow locally by applying the optical flow constraint, (5), to every pixel in a window W surrounding a pixel (x_i, y_i). A single displacement vector (u_i, v_i) corresponding to the center of the window is computed by solving a linear system for the N equations resulting from the constraint. In some variants, the window is combined with a weighting function and a weighted least squares problem is solved per pixel. The process can be iterated and/or applied in a multiscale fashion to compute larger displacements either by pre-warping the image $\mathbf{I}_{\text{sec}}$ or by shifting the window in the secondary image. Lucas-Kanade tracking is a purely local method and cannot calculate flow in the interior of uniform regions of the image [22].

A fast and robust approach derived from the Lucas-Kanade method is the eFolki algorithm [1] which is a derivative of the FOLKI optical flow estimator [23]. eFolki includes hierarchical estimation, a computational trick that enables a faster evaluation and a number of improvements regarding matching scores and filtering that have been established over the years [14]. We use the eFolki algorithm as a modern variant of Lucas-Kanade tracking in our evaluation. eFolki has been developed and tested for SAR coregistration [1]. We have used this implementation for our evaluation.

The variable parameters that can be adapted for the SAR case include the window size ω which is varied from coarse to fine for every iterative step, the number of iterations per level K and the pyramid levels J and a rank constraint r.

We include the technique in our evaluation as an advanced processing technique that has been tested for SAR coregistration.

D. Horn and Schunck

Local techniques as discussed above are not able to compute flow in homogeneous image regions. The Horn and Schunck (H&S) [13] method is a global energy minimization technique that includes regularization to mitigate this problem

$$\operatorname{argmin}_{u,v} \int_{\Omega} \|I_x u + I_y v + I_t\|_2^2 + \alpha^2 \left(\|\nabla u\|_2^2 + \|\nabla v\|_2^2\right) dx dy. \tag{6}$$

The first term encodes the optical flow constraint and is known as the data term, whereas the second term emphasizes a smooth displacement field and is known as prior term. The support of the image is denoted as Ω. The user parameter α^2 allows selecting a trade-off between a good data fit and a smooth solution. The parameter is squared to enable an interpretation as the standard deviation of the expected Gaussian noise [16]. The method is typically implemented in a multiscale fashion to allow for the computation of large displacements. In our evaluation, we use the implementation [16]. In practice, the number of scales and the scaling ratio are additional parameters to be tuned.

The technique provides a smooth flow and allows the possibility of using more than two frames of data. The disadvantage of the technique is that it can result in smoothed boundaries and is sensitive to outliers, like salt and pepper noise, in the data. Both disadvantages are due to the use of the square norm in the data and the prior terms, respectively.

We include the method as a representative of classical energy-based optimization techniques. Despite its age, it was found to be performing well if implemented correctly [14]. In some applications such as fluid mechanics, it can outperform more advanced techniques [10].

E. TV-L1 Algorithm

The TV-L1 algorithm [17] addresses the aforementioned problems by replacing the data and the prior term norms with the L1-norm. The L1-norm of the Euclidean norm of a gradient field is also known as TV

$$\operatorname{argmin}_{u,v} \int_{\Omega} \lambda \|I_x u + I_y v + I_t\|_1 + (\|\nabla u\|_2 + \|\nabla v\|_2) dx dy. \tag{7}$$

The TV-L1 algorithm is robust to outliers in the data and enforces step edges in the flow field due to its use of the TV. In usual implementations, again, multiscale strategies are employed. For our evaluation, we use the implemenation of [17]. The user parameters are λ to trade-off prior enforcement and data fit, as well as the number of scales and their ratio as in the H&S case. It should be noted that λ and α^2 behave inversely since λ can be factored out of the integral without changing the optimal u, v. However, note that the parameters are still not strictly comparable because of the different norms involved in the H&S case and TV-L1.

We include the technique as it was the top-performing class of algorithms in computer vision applications before the advent of machine learning techniques.

F. Recurrent All-Pairs Field Transform

RAFT is a modern machine learning technique used in optical flow and achieves state of the art performance on optical data. The RAFT extracts per pixel features, builds multiscale 4-D correlation volumes for all pairs of pixels, and iteratively updates a flow field through a recurrent unit that performs look-ups on the correlation volumes [11].

The working of the method can be divided into three main categories.

1) Feature encoder: this step extracts per pixel features from both images and also includes a context encoder which extracts features from the reference image.
2) Correlation layer: the inner product of the feature vectors are computed to result in a 4-D correlation volume. The last two columns of the volume are pooled at multiple scales resulting in a multiscale volume [11].
3) Update operator: performs an update on the current estimate of the optical flow values to the look up values obtained from the correlation layer.

The method and its variants are, at the time of writing, dominating the Middlebury benchmark and the KITTI benchmark [24].

We include RAFT [11] in our evaluation to assess the potential and the challenges associated with machine learning techniques in the field of SAR coregistration. Note that we do not perform retraining on SAR data due to a lack of sufficient amounts of training data. The results are therefore only indicative of the potential of RAFT for SAR, but not exhaustive. A discussion of the findings is included in Sections IV and V.

III. Evaluation Methodology

We evaluate the candidate algorithms under increasingly realistic conditions. For this, we first generate synthetic data from the real dataset described in Section III-A, that features realistic distortions, first without and subsequently with realistic noise levels. We also use different pairs of images to directly evaluate the algorithms on real data. The test data preparation is described in Section III-B. Finally, in Section III-C we describe the evaluation methodology, i.e., testing conditions and performance scores that will be used in Section IV.

A. Dataset: Multitrack Airborne SAR Images

1) Description: The SAR raw data is obtained from a flight campaign using the wachtberg imaging radar (WIR-10) operated at X-band at the Fraunhofer Institute for High Frequency Physics and Radar Techniques (FHR). The sensor was mounted on an aircraft and recorded the radar echoes of the scene while it was passing eight different tracks in succession. The incidence angles of the tracks toward the scene center for the different images over flight time are shown in Fig. 1. The scene is centered at 50.913N and 8.047E and is of the region Dreis-Tiefenbach, Siegen, Germany, as seen in Fig. 2 (left). The signal of the sensor has a bandwidth of 500 MHz and a center frequency of 9.8 GHz. The scene to be investigated has an extent of 400 m in the range and 300 m in the azimuth direction, discretized at 4000 × 3000 pixels.

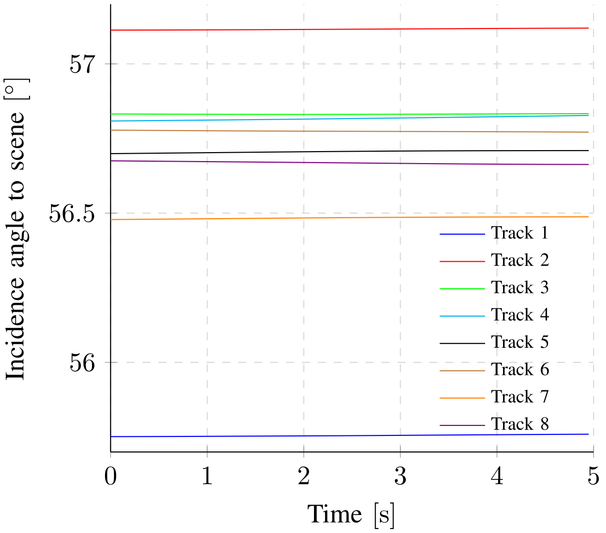


Fig. 1. Incidence angle over time to the scene center for all tracks of the airborne dataset.

Complex valued SAR images are processed from the radar raw data for a fixed reference height taken from scene center using a backprojection processor. The aircraft flew eight times in an almost straight parallel course whereby the altitude varied in a range of approximately 48 m. In our specific case, the SAR images obtained from the experimental radar system were not subjected to radiometric calibration. Nevertheless, all radar images were captured from a consistent distance and under identical radar parameter settings. In order to equalize the intensity values, the histograms of the images were first adjusted to each other; however, this adjustment had no discernible influence on the final results. The main aim of the SAR campaign was to obtain a stack of SAR images which can be used for SAR tomography. In this article, we use it to test coregistration algorithms in challenging scenarios.

2) Data Characteristics: The different heights of the tracks of the acquired dataset lead to different incidence angles toward the individual targets which results in a corresponding shift in pixels for every acquisition geometry. According to this, different local shifts in the image result for different target heights. The expected shifts due to a ground truth digital elevation map (DEM) and the relative positioning of Tracks 1 and 2 are depicted in Fig. 2 (middle).

The instability of the airborne sensor (variation of altitude, heading and lateral direction) lead to additional variations of the shifts between the individual SAR images. However, the main cause of the locally different pixel-wise shift is resulting from the foreshortening effect as depicted in Fig. 3, which is scaled differently in each image pair due to the different angle of incidence of the individual images. Large differences of incidence angles lead to large decorrelation effects, as is the case for example between Track 1 and Track 2. The coherence map of the two resulting SAR images is shown in Fig. 2 (right). This in turn makes the coregistration process difficult and complex, providing a challenging test environment for the algorithms in our evaluation.

B. Evaluation Datasets

In order to quantitatively evaluate the different coregistration methods, we need a known ground truth modality, e.g., expected pixel shifts as in Fig. 2 (middle). Since this is difficult to obtain for real data, we proceed in three steps of increasing realism in Sections III-B1–III-B3. We generate log-scale normalized intensity data in all cases since first, the computer vision techniques require this input format, and second, the decorrelation between different tracks is so large as to make the phase information unreliable.

1) Synthetic Case: For the used test site in the region of Dreis-Tiefenbach, a high-quality DEM is available, which can be used as ground truth. From the DGM1 data of Geobasis NRW with a grid size of 1 m and a height accuracy of $+/-2$ dm, an exact height value can be determined for each pixel position. In conjunction with the flight track coordinates, Fig. 1, 3-D positions on the ground can be transformed into the coordinate systems of the individual SAR images. We use this information to compute ground truth displacements, Fig. 2 (middle), between any pair of SAR images in our dataset. One of the images, Track 1, is used as a reference image throughout the synthetic tests. This image is warped using the ground truth displacements using inverse lookup to produce a synthetic secondary SAR image. Note that this procedure ignores occlusion and disocclusion effects at height discontinuities of the DEM. For the current purpose, we consider this to be acceptable since, e.g., wall heights of buildings are generally small in our data. The LIDAR-based data contains large changes in areas with vegetation in the range of several meters within a few pixels, so we smoothed this reference dataset for the evaluation of the synthetic data.

With reference to Fig. 3, the transformation is computed as follows: Given the height (H_i), and the ranges to the height of the object (r_{ih}) and to the reference height (r_{i0}) the pixel displacement on ground (Δr_{gi}) within the different images i depending on the height model can be calculated using trigonometry

$$\Delta r_{gi} = \sqrt{H_i^2 - r_{i0}^2} - \sqrt{H_i^2 - r_{ih}^2}. \tag{8}$$

The vector field $\mathrm{VF}_{\mathrm{ref}}$ used as ground truth is given by the difference of the displacements of images i and j

$$\mathrm{VF}_{\mathrm{ref}} = \Delta r_{gi} - \Delta r_{gj}. \tag{9}$$

We refer to this test case as "synthetic noiseless." It is important to note that the noise and speckle in the original image essentially become features in this setting: they need to be considered as artificial surface texture for proper interpretation. This case serves as a baseline result.

2) Synthetic Case With Gaussian Noise: In real situations, noise is a complex issue due to the combined actions of random processes associated with both the emitter, the detector, and the scene. Overall, these lead to locally different contrast and noise properties as well as speckle.

Analyzing the noise properties of the SAR images used, we observed that the real and imaginary parts of the images have an approximate Gaussian-shaped distribution (cf. [25]). We measured the statistics of this Gaussian noise in homogeneous and low amplitude areas of the reference image.

We then generate different additive white Gaussian noise for the reference and synthesized secondary image, Section III-B1, via the estimated mean and standard deviation of the low amplitude level patches for the real and imaginary parts.

Our experiments are performed for the calculated shifts using all interferometric image pairs with respect to image 1, and on different noise levels: the same noise values are added to the SAR data in differently scaled versions reducing the SNR of the whole image: 0× (noiseless), 2× (-3 dB), 4× (-6 dB), 6× (-8 dB) and 8× (-9 dB). The purpose of the test is to explore the degradation in algorithm performance.

3) Real Case: Finally, our airborne dataset, Section III-A, contains eight different SAR images. We use them in a pairwise fashion as reference and secondary images. In addition to the synthetic tests, for these image pairs all real-world effects such as different brightness distribution due to the different angles of incidence, different speckle noise and inconsistent occlusion/dis-occlusion regions play a role in the performance of the algorithms. We evaluate $N/2(N-1)$ pairs of images with $N = 8$ to obtain statistics, Figs. 8 and 9, across expected algorithm performance.

C. Testing Methodology

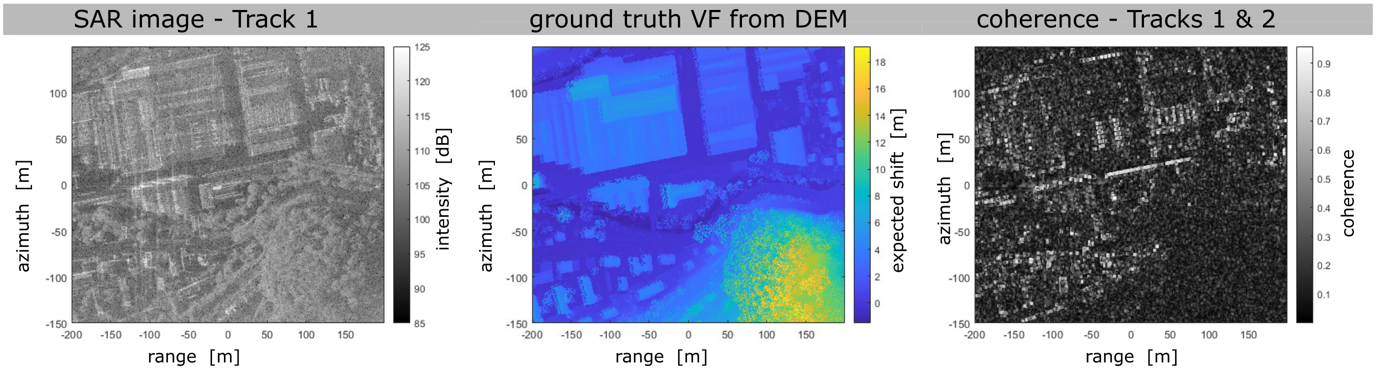
Our main test scenarios are the synthetic data cases Sections III-B1 and III-B2 since exact ground truth is available in these settings. We analyze the performance of the algorithms using different measures. The measures can be classified into two main categories: 1) an evaluation of the intensity image after coregistration and 2) an evaluation of the actual flow vector fields underlying the coregistration.

1) Intensity Image Evaluation: The measures evaluated for this case are on the resampled intensity images obtained after compensation of the shift, except for the measure of coherence, where complex data is used.

a) Coherence: Measures the preservation of phase relationships in the received signal and thus a measure for the similarity of two images. The coherence of two SAR images $\mathbf{a}$ and $\mathbf{b}$ is given by the cross correlation coefficient given by the following equation [26]:

$$\rho = \frac{E\left[\mathbf{a}\ \mathbf{b}^*\right]}{\sqrt{E\left[|\mathbf{a}|^2\right] E\left[|\mathbf{b}|^2\right]}}. \tag{10}$$

Coherence is a standard measure in SAR to evaluate the quality of registration [19]. It is evaluated on complex resampled images.


SAR image - Track 1

ground truth VF from DEM

coherence - Tracks 1 & 2

Fig. 2. (Left) One of eight airborne SAR images of the area Dreis-Tiefenbach captured by the WIR. (Middle) Magnitude of expected displacement between tracks 1 and 2 derived from a DEM of the region. (Right) Coherence map between SAR images 1 and 2 before coregistration.

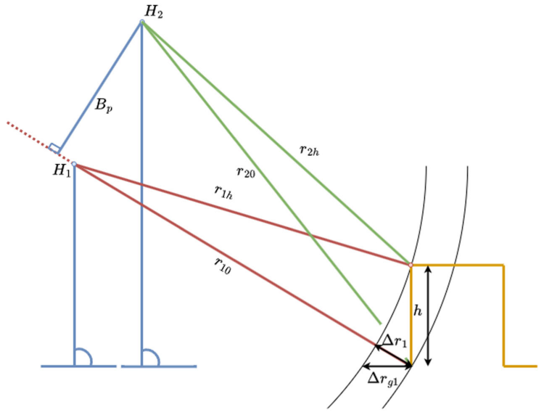


Fig. 3. Geometry of the foreshortening effect in SAR for two different incident angles. The corner of the object will be imaged at different ground range positions r_{gi} due to foreshortening. In addition, layover of object and ground plane may occur. In black the iso range lines of position 1 are shown, explaining this effect in slant range (Δr_1) and its projection onto ground range (Δr_{g1}).

b) Root Mean Squared Error: Root mean squared error (RMSE) measures the average difference between the intensity values, $\hat{I}$ produced by the technique and the reference intensity values, I. RMSE is given by the following equation:

$$\text{RMSE} = \sqrt{\sum\nolimits_{i=1}^{N} \frac{\left(I_i - \hat{I}_i\right)^2}{N}}. \tag{11}$$

We evaluate RMSE on log-intensity images.

c) SSIM: The structural similarity index (SSIM) [27] is a perceptual quality metric in computer vision. It is designed to be robust against local contrast changes and other common variations in visual images. SSIM is a full reference metric, i.e., a noise-free image is necessary for its computation. We use the noiseless reference image for that purpose. The SSIM for images is given by the following equation:

$$\text{SSIM} = \frac{(2\mu_I \mu_{\hat{I}} + c_1)(2\sigma_{I\hat{I}} + c_2)}{\left(\mu_I^2 + \mu_{\hat{I}}^2 + c_1\right)\left(\sigma_I^2 + \sigma_{\hat{I}}^2 + c_2\right)} \tag{12}$$

where, μ_I and $\mu_{\hat{I}}$ is the pixel mean of images I and $\hat{I}$ respectively, σ_I and $\sigma_{\hat{I}}$ is their standard deviation, respectively, and $\sigma_{I\hat{I}}$ is the covariance of I and $\hat{I}$. The constants $c_1 = 10^{-4}$, $c_2 = 9 \cdot 10{-}4$ are user variables with standard values. We evaluate SSIM on log-intensity images. The values of SSIM range between 0 and 1, where larger is better.

2) Flow Vector Evaluation: The metrics used in this case are to evaluate the flow vector quality.

a) Root Mean Squared Error: Measures the accuracy of the flow vectors. We compute the RMSE of the Euclidean distance between the predicted flow vectors $\hat{\mathbf{v}}$ and the reference flow vectors $\mathbf{v}^{gt}$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N} \frac{\left(\hat{\mathbf{v}}_i - \mathbf{v}_i^{gt}\right)^2}{N}}. \tag{13}$$

The RMSE is sensitive to outliers, i.e., a large difference between predicted value and ground truth [28].

b) Average Endpoint Error: EPE is used in computer vision to estimate the absolute accuracy of the estimated flow. It is defined as the average Euclidean distance between the estimated $\hat{\mathbf{v}}$ and the ground truth $\mathbf{v}^{gt}$ vector fields

$$\text{EPE} = \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\mathbf{v}}_i - \mathbf{v}_i^{gt} \right\|. \tag{14}$$

N is the total number of samples in the SAR image. Since the norm is not squared, EPE is relatively insensitive to outliers and large errors but can be affected by the density of the flow field. EPE is the widely used performance metric in the Middlebury benchmark [7] for evaluating optical flow estimation methods.

c) Average Angular Error: Average angular error (AAE) is defined as the mean angular distance between the estimated flow vector and the ground truth flow. This is represented as follows:

$$\text{AAE} = \frac{1}{N} \sum_{i=1}^{N} \left\| \arccos\left(\frac{\hat{\mathbf{v}}_i \cdot \mathbf{v}_i^{gt}}{||\hat{\mathbf{v}}_i|| \cdot ||\mathbf{v}_i^{gt}||} \right) \right\|. \tag{15}$$

AAE is used in combination with EPE to provide a more comprehensive evaluation of the accuracy of the optical flow techniques. AAE is commonly reported in the Middlebury benchmark along with EPE.

IV. Results

We first verify the coregistration ability of the different algorithms on synthetically generated data. As described in Sections III-A and III-B1, the dataset consists of seven image pairs containing the synthetically generated pixel shift that arises from the acquisition geometry of the different flight

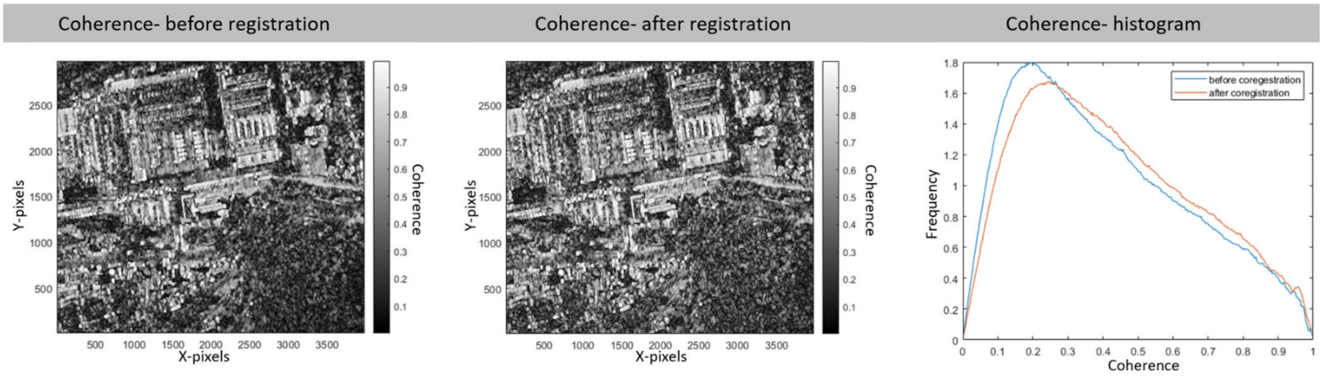


Fig. 4. (Left) Coherence of SAR image 5 and 7 before registration. (Middle) Coherence of SAR image 5 and 7 after registration using TV-L1. (Right) Histogram of the coherence obtained before and after registration.

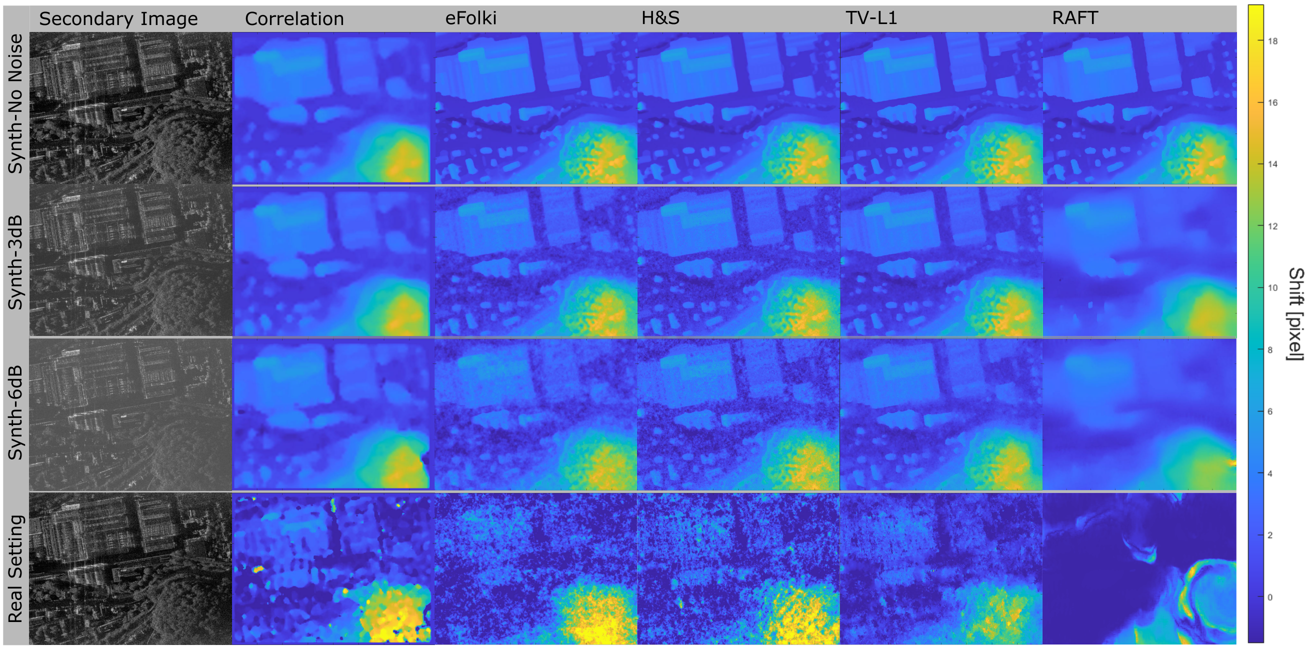


Fig. 5. (Top row) Registration results on noise-free data. (Middle two rows) Registration results with added noise. (Bottom row) Performance on real data. Zoom into the digital version for full detail.

trajectories. The reference trajectory and reference image are in each case Track 1 since it represents an extreme case in this measurement campaign. In addition, noise is added as described in Section III-B.

As an example, Fig. 5 shows the magnitude of the vector fields computed by the different algorithms. The dynamic range is the same for all images and is normalized to the maximum and minimum of the displacement magnitude of the reference vector field (varying from −1.5 to 19.2 pixels, using the data of Tracks 1 and 2). In the noiseless case, Fig. 5 (top row), all algorithms perform satisfactorily even though the machine learning technique RAFT produces overly smooth structures.

The noisy synthetic cases depicted in Fig. 5 (middle two rows), show results obtained for the weakest levels of reduction of the SNR in the test set (−3, −6 dB). The overall performance of the correlation method is mainly hampered by outliers and the window size of the patches. For outlier elimination, a median filter of size 4 × 4 pixels is used. In contrast, the optical flow methods are able to detect displacements without outliers. For the low noise setting illustrated here, all optical flow methods perform reasonably well, while small variations can be discerned: Horn & Schunck preserves the features of the structures in comparison to eFolki. TV-L1 shows the best results in both noise settings by suppressing noise and preserving the sharp edges in the image. RAFT again over-smooths the vector field.

To further meaningfully compare the correlation method with the optical flow-based ones, we increase the correlation window size of the patch to produce less outlier regions. These measures lead to a smoothing of the vector field and to longer computation times. We also balance oversampling for sub-pixel precision (≥8) against computation time. The resulting larger spacing of the control points creates larger gaps at the edges, which have to be extrapolated.

A numerical evaluation of all our synthetic tests is presented in Fig. 6. In the left column of the figure, the image-dependent measures are plotted, whereas on the right side, the measures

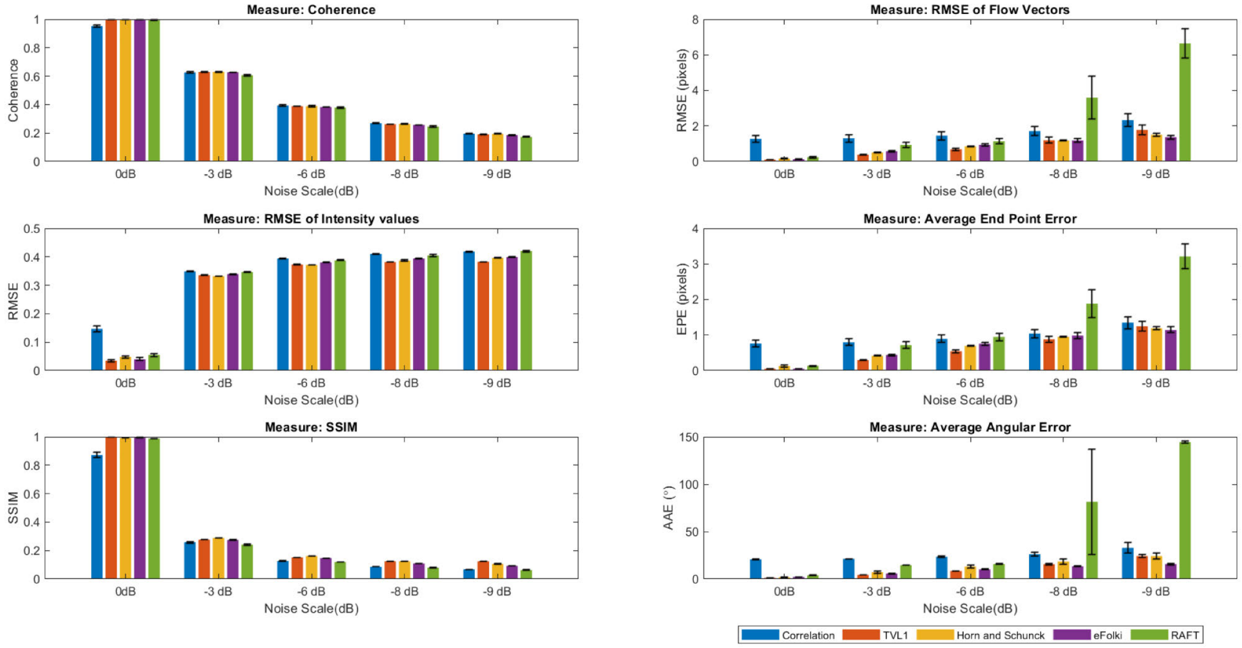


Fig. 6. Evaluation of optical flow algorithms for the synthetic case for different performance measures. For coherence and SSIM a larger score is better, for the other measures, a lower score indicates better results.

evaluating the vector field quality are shown. The parameters for the different techniques in our test are provided in Table I.

For the image-measures, it is notable that the optical flow-based algorithms perform similarly as a group and yield better results than correlation-based and RAFT coregistration. The optimization-based approaches degrade continuously with increasing noise levels. The errors witnessed in the vector field of optical flow techniques at 0, −3, and −6 dB exhibit similarities, whereas the errors are slightly smaller in the case of TV-L1 and H&S for −8 and −9 dB in contrast to eFolki. The vector field quality measures clearly show the problems of RAFT for the −8 and −9 dB SNR reduction settings. The RAFT algorithm rapidly drops in performance once the noise level exceeds the training noise level of the network. In particular, the angular error at −8 dB includes partial success and partial failure cases explaining the large variance. The −9 dB RAFT case consists of almost random results. If you look at the results with regard to the vector field quality measures of the correlation-based method, it is noticeable that with this method large errors occur even at 0 dB, which certainly deteriorate less with worse SNR. This is obviously due to the large window size. Overall, TV-L1 is the best algorithm in terms of the quality of reconstruction of the original vector field across noise levels.

We further apply the algorithms to the real data, Fig. 5 (bottom row). We note that RAFT completely fails in this setting, which hints at different noise sources than Gaussian noise in the real data. The other algorithms can handle the non-Gaussianity, however, the neural network-based technique has not been trained on these statistics and fails catastrophically. In terms of robustness in real-world settings, TV-L1 is the clear winner. If we look at Fig. 9, the TV-L1 algorithm provides the best coherence values of all algorithms, regardless of the length of the baseline.

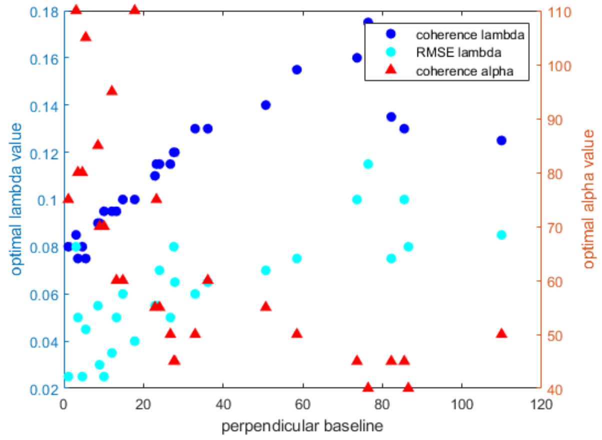


Fig. 7. Optimal λ values regarding the smallest distance of the vector field with respect to ground truth (blue) and the best coherence values (light blue). The optimal α values with respect to the best coherence values (red). All are plotted over the absolute perpendicular baseline.

The results obtained by TV-L1 or the Horn-Schunck method depend on the choice of the corresponding parameters. With poorer SNR (or coherence), for example, when using the TV-L1 algorithm, a larger λ value should be used in the calculation; with the H&S algorithm, α should be chosen smaller. In the case of real data, where a larger perpendicular baseline value causes a worse coherence, the values must be adjusted accordingly. For these two algorithms, using all possible image pairs, Section III-B3, and their baselines, we determined the optimal parameters, α and λ, respectively, by parameter scanning for maximum coherence or minimum deviation from the reference vector fields. The results are shown in Fig. 7. Here, a linear increase of the optimal λ value (TV-L1, blue circles) can be observed. The difference in the two curves is due to the imperfect ground truth, which, for example, does not take into account the azimuth

TABLE I

USER-ADJUSTABLE PARAMETERS FOR DIFFERENT LEVELS OF SNR REDUCTION OF THE SYNTHETICALLY GENERATED DATA

Algorithm	**Parameter**	**0 dB**	**-3 dB**	**-6 dB**	**-8 dB**	**-9 dB**
Correlation	Patch size	100^2	100^2	100^2	100^2	100^2
eFolki	N	181	181	181	181	181
HS	α	60	38	31	26	24
TV-L1	λ	0.09	0.1	0.12	0.14	0.16

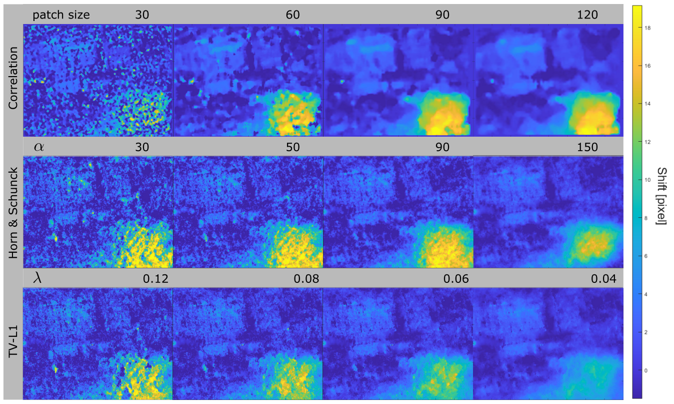


Fig. 8. Influence of algorithm parameters on estimated vector field quality (real datasets 1 and 2).

TABLE II

PARAMETERS USED FOR THE DIFFERENT METHODS FOR THE REAL DATASET

Method	**Used Parameter**
Correlation	Patch size = 100^2, Oversampling factor = 4 distance of control points = 50 pixels
eFolki	Number max. iterations N = 181
HS	Regularization weight α = 40-110
TV-L1	Data attachment weight λ = 0.1 Stopping criterion threshold ϵ = 0.017

dependencies due to nonlinear, nonconstant trajectories, but only the displacements due to different heights in the scene. In practice, a value that lies between the two curves has proven to be a robust solution. For α (H&S, red triangle) for our dataset the values vary in the opposite way from 40 to 110 having the highest value for the smallest baseline. For an impression of the visual impact of the parameter changes, the outputs of the single algorithms are shown in Fig. 8. We observe that the choice of the parameters has a great influence on the smoothness of the result and that different numerical measures of quality correspond to different parameter settings. Depending on the task, it can thus be helpful to adjust the parameters. For example, for the highest possible coherence between the co-registered images, a different value may be optimal as compared to the value for optimal vector field quality. To evaluate the possible advantage of using optical flow methods for coregistration, we consider the coherence of all image pairs before and after coregistration. This is also illustrated in Fig. 4, where the buildings exhibit a higher correlation index as compared to the hills. In Fig. 9 the coherence of the outputs of all methods for all image pairs is shown. The coherence value of the original image pairs is given in black as a baseline for comparison.

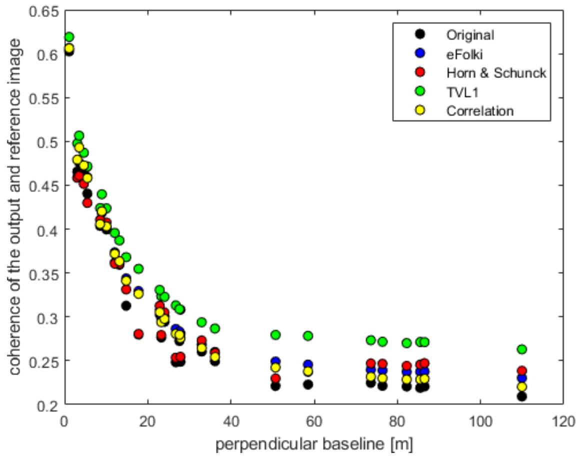


Fig. 9. Output coherence over perpendicular baseline of all images pairs is depicted for all algorithms for the real case.

TABLE III

PEAK MEMORY AND CPU TIME FOR EACH ALGORITHM (IMAGE SIZE 4000 × 3000 PIXELS)

Algorithm	**Peak Memory**	**CPU time (sec)**
Correlation	281.25 MB	1275.87
TV-L1	469.68 MB	209.21
Horn & Schunck	471.44 MB	70.34
eFolki	469.68 MB	945.15
RAFT	256 GB	17596.20

All methods can improve the coherence between the images, TV-L1 outperforms the others in our experiment, whereby the traditional correlation-based method makes only a small improvement. As a comment, it should be noted here that the correlation-based method can achieve better results if, for example, more, or only suitable, sampling points are used and the oversampling factor is increased. Likewise, a smarter sorting of the outliers can also lead to a better result. However, the calculation time increases enormously in this case and we settled for a compromise (see Table III). In addition, when evaluating the results, we noticed that with the Horn–Schunck algorithm it was necessary to adjust the parameter α to the dataset in order to obtain improved results in each case. For all other algorithms a constant parameter was used (see Table II). The TV-L1 algorithm produced the best result for each image pair. The CPU time and the peak memory usage for each technique are given in Table III. From this table, we observe that the correlation-based method with a reasonably comparable number of control points, depending on the resolution of the image, needs considerably longer time to calculate a vector field than the variational optical flow techniques H&S and TV-L1, which have about the same requirement of computational time and memory. eFolki is close to our compromise correlation technique. RAFT on the other hand has much higher requirements on the used PC than all other techniques. To demonstrate an application, we illustrate the improvement of an interferogram after coregistration. Fig. 10 shows the interferometric phase from a pair of images as an example. Using the TV-L1 algorithm to register two images results in a significant improvement in phase. The noise is generally reduced, but the phase information is preserved. The visual quality of the

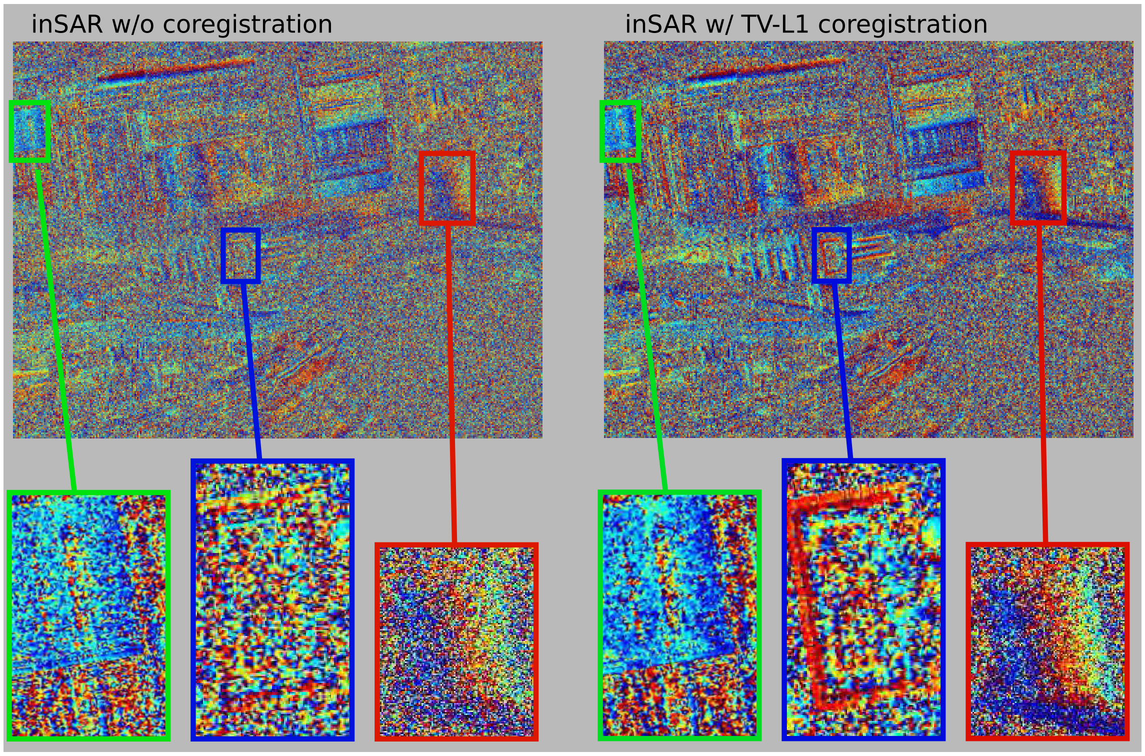


Fig. 10. (Left) Interferometric phase before coregistration and (Right) after coregistration using TV-L1 for tracks 5 and 7.

result is similar to the other optical flow techniques except RAFT.

V. Conclusion

We have compared optical flow techniques from computer vision for the task of co-registering a set of SAR images under challenging conditions such as large baseline and high noise level. Our experiments have shown that computer vision techniques outperform classical correlation-based techniques both, in terms of accuracy and in terms of computational performance. An added advantage is that dense registration maps are computed that enable a pixel-wise coregistration without tuning-heavy post-processing as in correlation techniques.

Among the computer vision techniques, variational methods are, at the time of writing, the best-performing techniques. The TV-L1 technique, in particular, offers excellent robustness and parameter stability. From our tests, we recommend its use as a baseline technique for future applications. The machine learning technique RAFT, while showing that ML can work on SAR data in principle, is not specialized enough to be competitive without retraining. To also harvest the promises of ML-based processing, as demonstrated in the computer vision Middlebury benchmark, in the SAR context, large datasets of real data with ground truth from a variety of acquisition platforms needs to be generated. In principle, methods like style transfer and domain adaptation could be utilized to synthetically generate data with SAR characteristics from other image sources. However, we also see that significantly higher computational resources will be required for progress in this direction.

For future work, the most important development is the inclusion of the full complex image information into the estimation procedures, possibly taking care of the cyclicity of the phase term, rather than relying on intensity-only information as done in this article. Explicit prior models for speckle noise could further improve the ability of the algorithms to differentiate image structure that should be registered from noise that is to be ignored.

References

[1] A. Plyer, E. Colin-Koeniguer, and F. Weissgerber, “A new coregistration algorithm for recent applications on urban SAR images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2198–2202, Nov. 2015.

[2] W. Zhang, “Robust registration of SAR and optical images based on deep learning and improved Harris algorithm,” *Sci. Rep.*, vol. 12, no. 1, p. 5901, Apr. 2022.

[3] D. Massonnet and K. L. Feigl, “Radar interferometry and its application to changes in the Earth’s surface,” *Rev. Geophys.*, vol. 36, no. 4, pp. 441–500, Nov. 1998.

[4] A. Reigber and A. Moreira, “First demonstration of airborne SAR tomography using multibaseline L-band data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.

[5] E. Sansosti, P. Berardino, M. Manunta, F. Serafino, and G. Fornaro, “Geometrical SAR image registration,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2861–2870, Oct. 2006.

[6] N. Yague-Martinez, M. Eineder, R. Brcic, H. Breit, and T. Fritz, “TanDEM-X mission: SAR image coregistration aspects,” in *Proc. 8th Eur. Conf. Synth. Aperture Radar*, Jun. 2010, pp. 1–4.

[7] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, “A database and evaluation methodology for optical flow,” *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, Nov. 2011.

[8] J. Zhang, Z. Suo, Z. Li, and Q. Zhang, “DEM generation using circular SAR data based on low-rank and sparse matrix decomposition,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 724–728, May 2018.

[9] S. Palm and U. Stilla, “3-D point cloud generation from airborne single-pass and single-channel circular SAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8398–8417, Oct. 2021.

[10] B. Atcheson, W. Heidrich, and I. Ihrke, "An evaluation of optical flow algorithms for background oriented Schlieren imaging," *Experim. Fluids*, vol. 46, no. 3, pp. 467–476, Mar. 2009.

[11] Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 1–8.

[12] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, vol. 2. San Francisco, CA, USA: Morgan Kaufmann, 1981, pp. 674–679.

[13] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[14] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. Comput. Vis. Pattern Recognit. Conf. (CVPR)*, 2010, pp. 2432–2439.

[15] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.

[16] E. Meinhardt-Llopis, J. Sánchez Pérez, and D. Kondermann, "Horn-schunck optical flow with a multi-scale strategy," *Image Process. Line*, vol. 3, pp. 151–172, Jul. 2013.

[17] J. Sánchez Pérez, E. Meinhardt-Llopis, and G. Facciolo, "TV-L1 optical flow estimation," *Image Process. Line*, vol. 3, pp. 137–150, Jul. 2013.

[18] Y. Wang, Q. Yu, and W. Yu, "An improved normalized cross correlation algorithm for SAR image registration," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Munich, Germany, Jul. 2012, pp. 2086–2089.

[19] Z. Li and J. Bethel, "Image coregistration in SAR interferometry," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 37, pp. 433–438, Jul. 2008.

[20] J. P. Lewis, "Fast template matching," in *Proc. Vis. Interface, Can. Image Process. Pattern Recognit. Soc.*, Quebec City, QC, Canada, May 1995, pp. 120–123.

[21] J. J. Gibson, *The Perception of the Visual World*, 1st ed. Boston, MA, USA: Houghton Mifflin, 1950.

[22] D. Patel and S. Upadhyay, "Optical flow measurement using Lucas Kanade method," *Int. J. Comput. Appl.*, vol. 61, no. 10, pp. 6–10, Jan. 2013.

[23] G. L. Besnerais and F. Champagnat, "Dense optical flow by iterative local window registration," in *Proc. IEEE ICIP*, vol. 1, Sep. 2005, pp. 137–140.

[24] M. Zhai, X. Xiang, N. Lv, and X. Kong, "Optical flow and scene flow estimation: A survey," *Pattern Recognit.*, vol. 114, Jun. 2021, Art. no. 107861.

[25] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images* (EngineeringPro Collection). SciTech Publ., 2004. [Online]. Available: https://books.google.de/books?id=IeGKe40S77AC

[26] R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," *Inverse Problems*, vol. 14, no. 4, pp. 1–54, Aug. 1998.

[27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[28] X. Liu, H. Liu, and Y. Lin, "Video frame interpolation via optical flow estimation with image inpainting," *Int. J. Intell. Syst.*, vol. 35, no. 12, pp. 2087–2102, Dec. 2020.

Prithvi Laguduvan Thyagarajan (Graduate Student Member, IEEE) received the bachelor's degree in telecommunications engineering from Visvesvaraya Technological University, Belagavi, India, in 2016, and the master's degree in electrical engineering from the Technical University of Delft, Delft, The Netherlands, in 2019, with a focus on Signals and Systems.

She started working with waveform design for airborne SAR during her bachelor's thesis and continued to work as a Research Assistant with the Center for Airborne Systems, Bengaluru, India. In 2020 to 2023, she worked as an early stage researcher funded by the EU H2020 project at the University of Siegen. She currently works with the Fraunhofer Institute for High frequency physics and radar techniques (FHR), Wachtberg, Germany. Her research interests include radar/SAR imaging and SAR tomography.

Holger Nies received the Diploma degree in electrical engineering and the Dr.-Ing. degree from the University of Siegen, Siegen, Germany, in 1999 and 2006, respectively.

Since 1999, he has been a member of the Center for Sensor Systems (ZESS) and a Lecturer with the Department of Electrical Engineering, University of Siegen, where he is currently working as the Chair of the Computational Sensorics, University of Siegen, and is responsible for radar-based signal and image processing. He was involved in different project works for the Mercedes-Benz Group AG (formerly Daimler AG, Stuttgart, Germany) in the field of engine modeling and parameter estimation with Kalman filters. He was working in the area of SAR interferometry for the German TerraSAR-X/TanDEM-X Mission. His research interests include mono- and bistatic SAR analysis and mission design, trajectory estimation, SAR interferometry and passive radar systems, and image analysis and optimization.

Patrick Berens received the Diploma degree in electrical engineering and the Ph.D. degree from Ruhr University Bochum, Bochum, Germany, in 1997 and 2003, respectively.

Since 1997, he has been working at the Fraunhofer Institute for High Frequency Physics and Radar Techniques (FHR), Wachtberg, Germany. Since 2011, he leads a Team that is engaged in radar imaging methods. His research interests include all kinds of imaging radar: SAR imaging for single- and multichannel sensors and multipass operation and ISAR imaging of moving objects.

Ivo Ihrke received the M.S. degree in scientific computing from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2002, and the Ph.D. degree (summa cum laude) in computer science from Saarland University, Saarbrücken, Germany, in 2007.

He is a Professor of computational sensing at the University of Siegen, Siegen, Germany. Prior to joining Siegen, he was a Staff Scientist at the Carl Zeiss Research Department, which he joined on-leave from Inria Bordeaux Sud-Ouest, Talence, France, where he was a Permanent Researcher. At Inria, he lead the Research Project "Generalized Image Acquisition and Analysis" which was supported by an Emmy-Noether fellowship of the German Research Foundation (DFG). Prior to that, he was heading a research group within the Cluster of Excellence "Multimodal Computing and Interaction," Saarland University. He was an Associate Senior Researcher at MPI Informatik, Saarbrücken, and an Associated with the Max-Planck Center for Visual Computing and Communications. Before joining Saarland University, he was a Post-Doctoral Research Fellow at the University of British Columbia, Vancouver, BC, Canada, supported by the Alexander von Humboldt-Foundation.