EXPERIMENTAL COMPARISON OF REGISTRATION METHODS FOR MULTISENSOR SAR-OPTICAL DATA

Béatrice Pinel-Puysségur, Luca Maggiolo, Michel Roux, Nicolas Gasnier, David Solarna, Gabriele Moser, Sebastiano B Serpico, Florence Tupin

EXPERIMENTAL COMPARISON OF REGISTRATION METHODS FOR MULTISENSOR SAR-OPTICAL DATA

Béatrice Pinel-Puysségur[1], *Luca Maggiolo*[2], *Michel Roux*[3], *Nicolas Gasnier*[3], *David Solarna*[2], *Gabriele Moser*[2], *Sebastiano B. Serpico*[2], *Florence Tupin*[3]

[1]CEA, DAM, DIF, F-91297 Arpajon, France
[2]University of Genoa, via Opera Pia 11a, 16145 Genoa, Italy
[3]LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France

ABSTRACT

Synthetic aperture radar (SAR) and optical satellite image registration is a field that developed in the last decades and gave rise to a great number of approaches. The registration process is composed of several steps: feature definition, feature comparison and optimization of a geometric transformation between the images. Feature definition can be done using simple traditional filtering or more complex deep learning (DL) methods. In this paper, two traditional approaches and a DL approach are compared. One can then wonder if the complexity of DL is worth to address the registration task. The aim of this paper is to quantitatively compare approaches rooted in distinct methodological areas on two common datasets with different resolutions. The comparison suggests that, although more complex, the DL approach is more precise than traditional methods.

Index Terms— Multi-modality registration, SAR, optical imagery

1. INTRODUCTION

The currently unprecedented number of missions and sensors leads to an increased interest of the remote sensing community for multi-modal satellite image registration. The present paper focuses on the automatic registration of SAR and optical satellite images. Despite a substantial literature on this subject, this problem can still be considered as only partially solved for several reasons. First, the geometry of SAR images is very particular and completely different from optical images. Secondly, a SAR image looks very different compared to an optical image as it is sensitive to different characteristics of the ground. At last, the on-going increase of resolution and diversification of acquisition modes makes this problem challenging to tackle for a unique algorithm.

The registration process is usually preceded by a pre-processing step so that images are in a common geometry. The aim of registration is to find a geometric transformation so that the input images can be superimposed. The registration process can be divided into three steps [1]. The first one is feature definition, which can be a collection of particular points, windows or even whole images. Feature definition is generally dedicated to SAR or optical images to take into account their different properties. The second step is feature comparison. A comparison function is used in order to match features from the two images. This function is adapted to the nature of the features, for example SIFT-like features [2, 3] or windows matching [4, 5]. The third step is the computation of a geometric transformation between the input images, typically modeled by a simple function, e.g., a global translation with optional homothetic and rotation terms. The third step can be computed by the direct optimization of the comparison function of step 2 or by an optimization relying on a set of previously matched candidates. Steps 2 and 3 are sometimes jointly performed, as in the methods discussed here but also in optical flow approaches [6].

The aim of this paper is to compare the results of two methods using traditional filters versus a deep learning (DL) approach, tested on common datasets with various spatial resolutions and sensors, described in Section 2. These methods mainly differ in the feature definition step (see Figure 1), as described in Section 3. Two methods are used for the feature comparison and optimization steps and are described in Section 4. Different combinations of the last two steps are tested to compare separately the performance of each approach for each step. Section 5 compares the results of the methods on common datasets and Section 6 draws conclusions according to this experimental investigation.

2. DATASETS AND PRE-PROCESSING

Two datasets have been used for the comparisons. The first one is composed of a Pleiades panchromatic image and a COSMO-SkyMed SAR image acquired in Spotlight mode on an ascending orbit over a countryside area near Bussac-Forêt (France). The study zone is composed of woods, fields, roads and a few buildings. This dataset has very high spatial resolution (50 cm). The SAR image is in full resolution and the optical image has been projected into the radar geometry using

ALOS DEM. The COSMO-SkyMed image was despeckled using two techniques (see Figure 1): a Wiener filter applied with a homomorphic filtering strategy (logarithmic scale) and the wavelet-based method in [7].

The second dataset has been derived from Sentinel-1 (S1) SAR and Sentinel-2 (S2) optical data, acquired in 2018 over Amazon, north of Pozo Colorado (Paraguay). The area covers crops, forests and rivers. The S2 image is made of the Sentinel-2 channels with 10 m resolution (*i.e.* RGB-NIR). The S1 data come from multitemporally despeckling a time series of 7 Sentinel-1 acquisitions using the method in [8]. Before registering, the S1 data were resampled on the same pixel lattice as the S2 image.

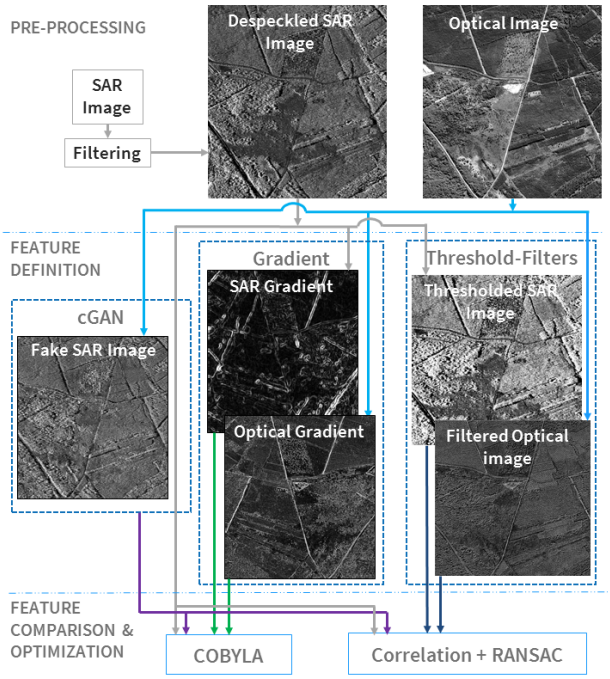

Fig. 1. Block diagram of the different steps of the registration process for the methods compared here.

3. FEATURE DEFINITION

This section presents the traditional and the DL methods tested here for feature definition.

3.1. Traditional methods

The first proposed traditional approach is based on gradient matching. Indeed, most edges are common features shared by optical and SAR images. This is especially true in countryside landscapes, with various kind of fields. In dense urban areas with buildings signatures, it can be more appropriate to use linear features for SAR data and edges in the optical data as proposed in [9]. Thus, the features are gradient magnitude maps (see Figure 1). The gradient computation on optical images is done using Deriche's approach [10], derived from Canny's criteria [11], while the SAR gradient can be efficiently computed using Gradient by Ratio (GR) in [2].

The second method is a multi-scale technique called OSCAR (Optical to SAR Correlation-based Automatic Registration). At each scale, the SAR image is automatically thresholded in order to avoid very high values to mislead the following correlation step. The optical image is filtered so that it looks like a SAR image (see Figure 1). As a rule of thumb, bright SAR pixels often correspond to areas with local variations on the optical image whereas homogeneous parts of the optical image are often quite dark in the SAR image (e.g. flat areas as roads or water). Five filters are used to reproduce this first-order correspondence: the first one is the standard deviation on a square window and the second is the minimum of the standard deviations already computed on the same window. The three other filters are the classical Sobel filter, morphological gradient and absolute value of the Laplacian. These five filtered optical images can be considered as handcrafted fake SAR images.

3.2. Image-to-image translation through DL

Image-to-image translation based on DL has attracted considerable attention for registration lately. It aims to transform one or both multisensor images to a common domain. That way, it is easier to compare their inherently different characteristics and get features useful to register. Generative adversarial networks (GANs) proved to be particularly efficient. A GAN is made of two networks trained in competition [12]. In particular, a conditional GAN (cGAN) is aimed to generate output data whose distribution matches that of a target source from a non-noise input source [13]. Here, the desired output is an image with SAR-like distribution ("fake SAR"), while the non-noise input is the optical image. The adopted DL method is the one in [14], which consists of two steps: first, a translation stage by means of the `pix2pix` cGAN [13], and then, an area-based registration stage using an ℓ^2 similarity (see Section 4.2). Details can be found in [14]. The application of the cGAN to the Amazon dataset is also described in [14]. Regarding the Bussac dataset, the Pléiades panchromatic image was manually warped to the COSMO-SkyMed image grid (which was around 5500×5500 pixels) to have paired patches to be used for training. Around 75% of the scene was used to train the cGAN. The training set was made of 101 patches (512×512 pixels each) drawn from the whole scene except the South-West corner, which was used for testing. This training strategy was used here for experimental purposes, but in a general application of the method, the training would make use of pre-registered patches coming from the same data sources but not from the specific images to be registered [14]. `pix2pix`, which is aimed at RGB im-

ages, was modified to map from the single-channel panchromatic to the single-channel SAR domains. The filter size was enlarged to 5×5 to account for possible residual misregistrations in the training patches. The number of training epochs was experimentally fixed at 200. The training set size is rather limited, so it is important to minimize the risk of overfitting.

4. FEATURE COMPARISON AND TRANSFORMATION COMPUTATION

This section describes the two tested approaches for feature comparison and optimization of the transformation.

4.1. Correlation and RANSAC based transformation computation

The comparison and optimization steps for the OSCAR algorithm is a multi-scale approach. Phase correlation between the thresholded SAR image and each of the five filtered optical images is performed on windows of 256x256 pixels with an overlap of 32 pixels. Five disparities maps are then obtained and filtered altogether by the RANSAC filter to obtain a rotation scale translation (RST) transform. Finally the SAR image is resampled to be superimposed to the optical image. As the optimization step described is based on correlation, it is relatively similar to that of the next section.

4.2. ℓ^2 area-based similarity metric

Area-based registration operates on the whole image directly, with no focus on specific points or linear features. It generally provides low registration error at the cost of high computational burden, often relying on information-theoretic comparison functions [15, 1]. However, if the feature definition stage is effective at mapping the multisource data to a common domain, then a simpler and faster correlation-type similarity may become feasible. Here, the area-based function in [14] is used, that is the ℓ^2 similarity between the outputs of the feature definition stage, expressed as a function of an RST transformation. Since it is generally a nondifferentiable function of the transformation parameters, a derivative-free minimization method is necessary. Among this family of algorithms, an accurate and efficient constrained maximization technique is identified in COBYLA. It iteratively approximates the optimization problem by a suitable sequence of linear programming subproblems. Details can be found in [14].

5. COMPARISON OF THE PRESENTED METHODS

In order to quantitatively compare the different methods, the SAR and optical images were first registered using manual tie points. Then, four RST transformations were applied with rotations ranging from 1.4° to 2.5°, a scale factor of 1.01 and translations from 25 to 45 pixels in each dimension. As the ground truth transformation was known, it was possible to estimate the Root Mean Square Error (RMSE) as in [16]. On the Amazon test area, the GR and DL methods were both tested with the optimization step in Section 4.2 and compared with OSCAR. Results are shown in Table 1. It appears that the DL approach outperformed the OSCAR approach as it attained sub-pixel RMSE while RMSE varied between 1 and 8 pixels for OSCAR. However, even being not very precise, the final RMSE of OSCAR was significantly reduced compared to the initial one. Using GR and the ℓ^2 function, small RMSEs (down to $\sim$0.2 pixels) were obtained in the case of some transformations, whereas the method did not converge to accurate solutions in other cases. This might be ascribed to the presence of local maxima in the ℓ^2 similarity between the GR features. When the same functional was applied to the DL features, this variance was not observed and low RMSEs were consistently achieved in the case of all transformations.

This first comparison suggests that the whole DL processing chain yielded smaller RMSE than OSCAR, at least on the considered datasets. However, this analysis cannot separate the effects of the feature definition and comparison/optimization steps. To focus on the role of the latter step, the fake SAR and despeckled SAR images were fed to both the ℓ^2 function and the correlation-RANSAC approach on the Bussac dataset (see Figure 1). The gradient maps were combined with the ℓ^2 function. The results are shown

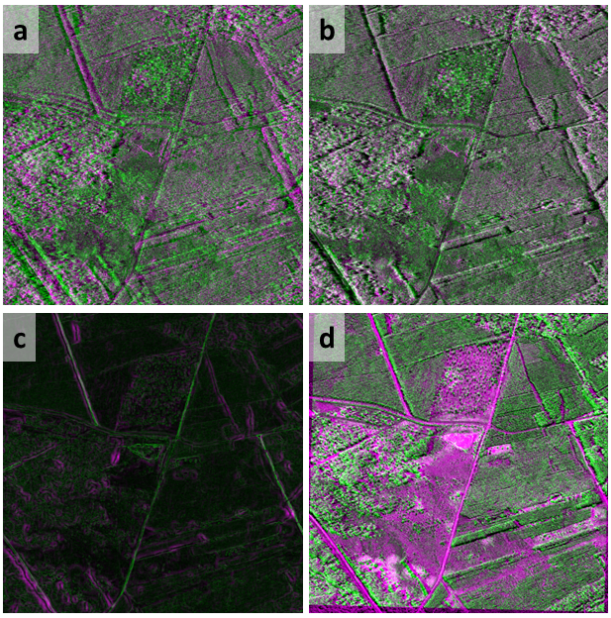

Fig. 2. Results for transformation 4 on Bussac. (a) Overlay of fake SAR and SAR image before registration. (b) Same as (a) after registration. (c) Optical and SAR gradients after registration by gradient and ℓ^2/COBYLA. (d) Optical and SAR image after registration by OSCAR. SAR products are in green and optical ones in magenta.

Test	Initial	GR + ℓ^2	DL + ℓ^2	OSCAR
1	76.45	0.19 / 87	0.34 / 0.35	4.35 / 7.82
2	72.23	0.16 / 85	0.24 / 0.22	2.95 / 5.46
3	60.35	0.19 / 0.11	0.22 / 0.42	2.63 / 3.64
4	36.22	30 / 40	0.22 / 0.27	1.89 / 2.48
avg.	61.31	7.63 / 53.02	0.25 / 0.31	2.96 / 4.85

Table 1. RMSE in pixels (area1 / area2) for the Amazon dataset and four test transformations.

Test	Initial	GR + ℓ^2	DL + ℓ^2	OSCAR	DL + RANSAC
1	111.02	3.15	0.87	11.04	2.96
2	96.38	23	0.81	13.30	2.69
3	88.92	47	0.85	11.14	3.65
4	34.37	2.83	0.89	14.67	1.04
avg.	82.67	18.99	0.85	12.53	2.59

Table 2. RMSE in pixels for the Bussac dataset and four test transformations.

in Table 2. Again, the full DL approach provided the best results among the considered techniques. Comparison of columns 3 and 5 shows that the ℓ^2 similarity yielded smaller RMSE than RANSAC by a factor of about 3. Comparison of columns 4 and 5 shows that the first step had a greater impact as there was a factor of about 4 between OSCAR and the DL/RANSAC combination. This led to a gain of 12 between the DL and OSCAR methods. The GR + ℓ^2 combination yielded RMSE $\sim$ 3 pixels for two of the considered transformations, while it did not achieve accurate solutions in the remaining cases, thus confirming the variance observed with the Amazon dataset. The DL + ℓ^2 method consistently led again to subpixel RMSE.

6. CONCLUSION

The experiments indicate that all considered methods improved the registration of the optical and SAR images as compared to the initial input multisensor data. This confirms their potential in the framework of this challenging multimodal fusion task. The comparison shows that the considered DL approach is especially effective on the considered datasets. Yet, an intrinsic drawback is that the network should be trained on a pre-registered dataset, contrarily to traditional methods, which do not require training processes. This suggests a tradeoff between DL and traditional approaches in terms of registration accuracy vs training requirement/complexity. DL approach complexity is justified when post-processing requires very high registration accuracy. While DL training generally takes a considerable time and a dedicated computing infrastructure (e.g., GPU), the prediction of the fake SAR image from the input optical scene through the trained network is quite fast. Although less precise, the GR and OSCAR methods are both fast and easy to run and give reasonable results that significantly reduce the initial offset between the images.

7. ACKNOWLEDGEMENTS

We thank D. Hateau and C. Guérin (CEA) for their help. The DL method was developed within the ESA CCI+ HRLC project: the support is gratefully acknowledged.

8. REFERENCES

[1] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image Vision Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[2] F. Dellinger et al., "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE Trans Geosci Remote Sens*, vol. 53, no. 1, pp. 453–466, Jan. 2015.

[3] Y. Xiang et al., "Automatic Registration of Optical and SAR Images Via Improved Phase Congruency Model," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 13, pp. 5847–5861.

[4] J. Inglada and A. Giros, "On the Possibility of Automatic Multisensor Image Registration," *IEEE Trans Geosci Remote Sens*, vol. 42, no. 10, pp. 2104–2120, 2004.

[5] S. Suri and P. Reinartz, "Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 2, pp. 939–949, 2009.

[6] G. Brigot et al., "Adaptation and evaluation of an optical flow method applied to coregistration of forest remote sensing images," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 9, no. 7, pp. 2923–2939, 2016.

[7] S. Parrilli et al., "A nonlocal sar image denoising algorithm based on llmmse wavelet shrinkage," *IEEE Trans Geosci Remote Sens*, vol. 50, no. 2, pp. 606–616, 2012.

[8] W. Zhao et al., "Ratio-based multitemporal sar images denoising: Rabasar," *IEEE Trans Geosci Remote Sens*, vol. 57, no. 6, pp. 3552–3565, 2019.

[9] G. Lehureau et al., "Registration of metric resolution SAR and Optical images in urban areas," *EUSAR 2008*, 2008.

[10] R. Deriche, "Using canny's criteria to derive a recursively implemented optimal edge detector," *Int J Comput Vis*, vol. 1, no. 2, pp. 167–187, Dec 1987.

[11] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, , no. 6, pp. 679–698, 1986.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT Press, 2016.

[13] P. Isola et al., "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017.

[14] L. Maggiolo et al., "Automatic Area-Based Registration of Optical and SAR Images Through Generative Adversarial Networks and a Correlation-Type Metric," in *IGARSS 2020*.

[15] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, Eds., *Image Registration for Remote Sensing*, Cambridge University Press, 2011.

[16] I. Zavorin and J. Le Moigne, "Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery," *IEEE Trans Image Process*, vol. 14, pp. 770–782, 2005.