FAST AND ACCURATE SUB-PIXEL DISPLACEMENT ESTIMATION FROM OPTICAL SATELLITE IMAGES USING A NEW HYPER-REALISTIC EARTHQUAKE DATABASE AND U-NET ARCHITECTURE

Tristan Montagnon[1], James Hollingsworth[1], Erwan Pathier[1], Mathilde Marchandon[1,2], Mauro Dalla Mura[3,4], Sophie Giffard-Roisin[1]

[1] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France
[2] Department of Earth and Environmental Sciences, Ludwig-Maximillians-Universität München, Germany
[3] CNRS, Inria, Grenoble Institute of Technology (Grenoble INP), Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab), Université Grenoble Alpes, 38000 Grenoble, France
[4] Institut Universitaire de France (IUF), 75231 Paris

ABSTRACT

Estimating the ground displacement from non-rigid registration of two optical satellite images, separated from hours to months, is key in the study of natural disasters such as earthquakes. Compared to standard image registration and flow estimation tasks, a key challenge here lies in resolving very small displacements (typically cm- or m-scale) with sub-pixel accuracy and precision using coarser image resolutions (e.g. 15m for Landsat-8). Traditional block matching/sliding window methods, employing local windowed correlation techniques, are unable to reduce the effects of long-wavelength noise arising from differences in image lightning, vegetation, or acquisition artifacts. By using both local and global scales, fully convolutional deep learning registration models (U-nets) are potentially able to better resolve ground displacements, less affected my multi-scale noise. Yet, no labelled database exists for ground deformation. Here, we develop a new synthetic database of 100,000 realistic satellite image pairs containing simulated earthquake displacements, along with their ground truth displacement maps, which are used to train state-of-the-art fully convolutional deep learning models (U-net).

***Index Terms*—** optical image correlation, image registration, satellite imagery, deep learning, geodesy

1. INTRODUCTION

Precise estimation of ground displacement from the registration of optical satellite images is fundamental for the study of natural disasters. In the case of earthquakes, characterizing near-field displacements around surface ruptures provides valuable constraints on the physics of earthquake slip. Recently, image correlation has been used to investigate the degree of slip localization, and how it may vary as a function of geological parameters (such as fault structural maturity), raising the possibility that slip localization (vs distribution) may be predictable, with important implications for seismic hazard assessment.

Optical satellite geodesy, particularly through Optical Image Correlation (OIC), has transformed the characterization of ground deformation linked to natural hazards like earthquakes [1, 2]. OIC is able to resolve full deformation fields between multiple optical satellite images captured over the same area at different times (hours to years apart). Unlike more precise techniques like InSAR, OIC excels in resolving large, high-strain displacements with fine spatial detail and using extended temporal baselines (up to several decades). It has found extensive application in constraining ground displacements produced by earthquakes, volcanoes, landslides, and glaciers [3]. In earthquake studies, the displacements can often be relatively small compared to the image pixel size, and may also vary spatially in a complex manner: sub-pixel precision and high spatial detail are therefore crucial for accurately capturing the complex variations in displacement values associated with earthquake surface ruptures.

OIC methods for quantifying image displacement have traditionally employed a sliding window approach, in which the displacement is resolved at high frequencies using cross correlation. This may be achieved either in the spatial domain [4, 5, 6, 7] or frequency domains [8, 9, 5, 10]. Spatial cross-correlation compares reference and target images through a sliding window approach [5]. Frequency-based correlation

Thanks to IDEX Université Grenoble Alpes, INSU PNTS, CNES, and CDP Risk for funding. For the computations, thanks to GRICAD infrastructure (gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities, and to the HPC/AI resources of IDRIS under the allocation 2022-AD011012345R1 made by GENCI. ISTerre is part of Labex OSUG@2020 (ANR10 LABX56).

simply takes advantage of the FFT to more efficiently compute the correlation matrix, from which the displacement is estimated, thus giving similar results to spatial-based correlators, albeit with faster run-times [9]. When optimized for sub-pixel performance, traditional correlators can resolve displacements with precision of less than 1/10th of a pixel for typical Earth Observation satellite images. However, in cases where image noise is very low between multiple images, the precision may be even higher.

Image registration and displacement field estimation from optical images has been successfully addressed by recent data-driven approaches, and in particular CNNs, e.g. in medical imaging [11], and remote sensing [12]. Displacement field estimation between two images can be efficiently solved by deep learning, e.g. treating optical flow estimation as a learning task [13]. The large majority of deep learning image registration and flow estimation approaches are derived from fully-convolutional U-net architectures [14]. However, they focus on the estimation of large displacements ($>$ 1 pixel) from temporally dense datasets (e.g. typical of video feeds), while the estimation of sub-pixel shifts from temporally limited and distant acquisitions (e.g. typical of remotely sensed images) has been little studied. Several recent studies have demonstrated the potential of neural network architectures to also retrieve sub-pixel displacements [15, 16, 17]; although, no applications in remote sensing were proposed.

In [18], we developed an innovative deep learning method to estimate ground displacement maps with sub-pixel precision from optical satellite images. Our proposed approach relied on the same principle as state-of-the-art OIC approaches, by working at the local scale with small windows (typically 16x16 or 32x32 pixels), while making the assumption of a locally rigid and non-rotating transformation; the translational displacement between the two windows is evaluated using a CNN. To minimize the displacement bias in the vicinity of sharp discontinuities, we further developed a second training dataset which included non-rigid displacements (i.e. simulating sharp discontinuities). Training a CNN with this second archive, allowed us to significantly reduce the bias in the near-field of earthquake surface ruptures, which is critical both for the accurate documentation of near-field displacement in earthquakes, as well as correctly investigating the underlying physics of fault slip.

While this method has improved the accuracy with respect to traditional OIC methods in discontinuity areas, thanks to our specific discontinuity simulated training dataset, it does not solve all their limitations; as such, we are still sensitive to image noise arising from differences in lightning, vegetation, anthropic changes, acquisition artifacts, etc.

In this new work, we propose a U-net-based method to solve the sub-pixel displacement estimation problem at a global scale. Such architecture incorporates contracting and expansive paths, and is able to retrieve full scale surface displacement maps, making use of both global ad local features, which allow to further reduce the displacement noise resulting from differences in the input images. To do so, we trained our model with a newly generated synthetic database: real satellite acquisitions, warped with 100,000 different ultra-realistic synthetic displacement maps, representing realistic faults.

2. METHODS

2.1. SYNTHETIC DATASET GENERATION

In the Earth Science community, no suitable archive of synthetic earthquake displacements currently exists for the purpose of training such a Convolutional Neural Network. Moreover, the absence of accurate and spatially comprehensive ground displacement measurements in real earthquake scenarios poses a challenge for establishing a pertinent database of applicable ground truth information. As a result, we create synthetic satellite images with known displacements to train our network.

We generate realistic synthetic earthquake image pairs by re-sampling Landsat-8 satellite acquisitions, incorporating realistic synthetic displacement fields on one of the two acquisition dates. The key is to make the displacement fields as realistic as possible. To do so, we create a pipeline to generate realistic fault discontinuities with rough geometries and slip distributions in a homogeneous elastic half-space. Surface displacement maps are then computed using analytical expressions linking slip on triangular fault patches to surface displacement, based on prescribed fault geometry and earthquake slip distribution. The faults follow natural earthquake fault scaling, rupture only the seismogenic crust, and exhibit geometric roughness and fractal slip distributions. The models focus on strike-slip faults, discretized with an unstructured meshing approach (Mesh2D, [19]) using triangular displacement elements (TDEs). The high resolution near the model surface enables us to generate realistic displacements at the resolution of the satellite imagery. Observation points are generated using an unstructured mesh, densifying points near the surface rupture. We then extract 1024×1024 patches from Landsat-8 satellite images acquired on two different dates (called pre and post, separated by weeks to months) in a stable region. An initial global co-registration step ensures accurate image alignment over the entire region. The displacement field warps the second image using a quintic-order spline re-sampling algorithm [20] with high precision ($\sim$ 1/100th to 1/1000th pixel) compared to state-of-the-art sub-pixel registration methods ($<$1/10th pixel).

We build two training dataset (one with pre and pre-warped, and the other with pre and post-warped), each made of 100,000 samples for training of size 1024x1024 pixels (80% for training; 20% for validation).

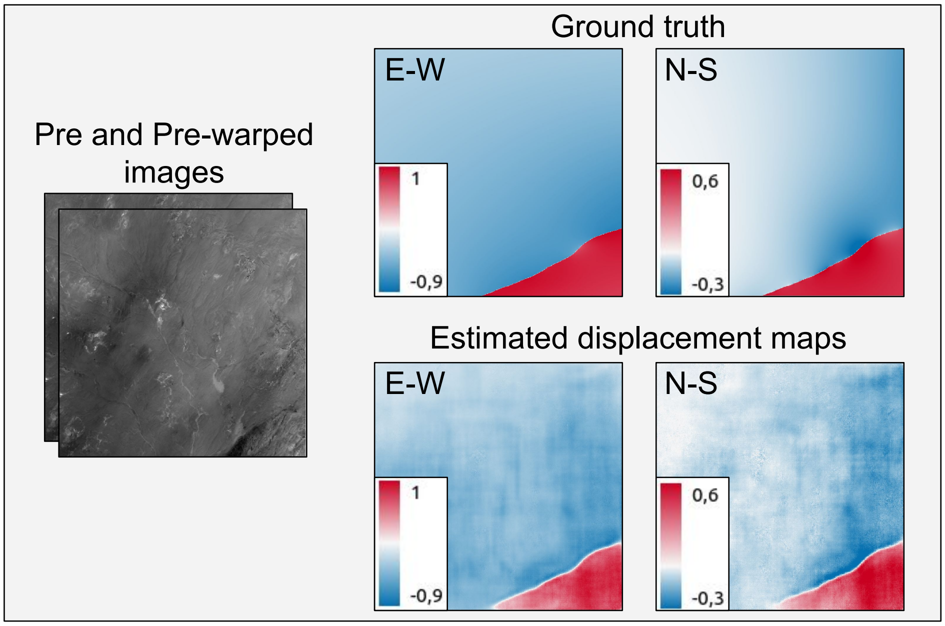

Fig. 1. E-W and N-S displacement maps for the trained StrainNet on one of the 20k validation samples. On the left, the pre and pre-warped images, and on the right the target ground truth and the estimated displacement maps. Results are expressed in pixels (mean error: 0.063 px).

2.2. SUB-PIXEL REGISTRATION U-NET MODELS

The input of the U-net registration model is composed of two 512x512 black&white satellite images (a 'pre-earthquake' and a 'post-earthquake'), and outputs displacement maps of size 512x512 (two channels, the East-West displacement and the North-South displacement), where both inputs and targets are cropped from the initial 1024x1024 samples.

In this study, three neural network architectures, FlownetS, FlownetSD (both are bricks of the Flownet-2.0 [15], and FlownetSD being optimized for small displacements) and StrainNet [16], were trained with our custom dataset, each designed with specific parameters and hyperparameters. FlownetSD and StrainNet are able to retrieve full scale displacement fields, while FlownetS only retrieves a 4x smaller output. The three models have around 38 million parameters. The key hyperparameters include 10 convolutional layers in the contracting part and 4 in the upscaling part, with an increasing and decreasing number of kernels per layer in the respective sections. The kernel sizes differ, with successively the first layer of 7x7, the second of 5x5, and the remaining layers of 3x3 for FlownetS and StrainNet, and only 3x3 for FlownetSD. A window size of 512x512, a batch size of 8, and a multiscale loss function coming from [16] were employed during training, during 100 epochs. In terms of computation time per epoch, FlownetS required 4.6 minutes, FlownetSD took 41 minutes, and StrainNet took 52 minutes, all three using four GPUs V100.

3. PRELIMINARY RESULTS AND CONCLUSION

We show on Figure 1 the outputs for the trained StrainNet architecture, using an image pair from the pre/pre-warped validation dataset. The main features of the ground truth displacement maps are well-retrieved by the model.

In total, the mean accuracies (absolute value of ground truth minus estimation) on the whole validation dataset are respectively 0.078, 0.068, and 0.069 pixels, for trained FlownetS, FlownetSD and StrainNet.

To conclude, the models exhibits promising preliminary results, showcasing their capabilities to retrieve full-scale surface displacement maps with high accuracy. While more complete validations and direct comparisons with other state-of-the-art approaches are ongoing, our first findings suggest that the proposed database and U-net-based approaches have the potential to significantly reduce the different noises that are currently present in the displacement maps estimated using traditional state-of-the-art sliding window correlators. For the moment, only pre/pre-warped dataset has been used to train the model. With pre/post-warped samples for training, we have good expectations that our U-nets will manage the displacement noise resulting from differences in the input images. Also, our two datasets are a valuable addition for the Earth Science community, providing realistic synthetic ground truth for various applications.

4. REFERENCES

[1] J. Avouac and S. Leprince, "Geodetic imaging using optical systems," in *Geodesy*. Elsevier Inc., 2015, pp. 387–424.

[2] M. Marchandon, T. J. Wright, and J. Hollingsworth, "Remote sensing of the earthquake deformation cycle," *Surface Displacement Measurement from Remote Sensing Images*, pp. 191–246, 2022.

[3] S. Leprince, E. Berthier, F. Ayoub, C. Delacourt, and J.-P. Avouac, "Monitoring earth surface dynamics with optical imagery," *Eos, Transactions American Geophysical Union*, vol. 89, no. 1, pp. 1–2, 2008.

[4] P. A. Rosen, S. Hensley, G. Peltzer, and M. Simons, "Updated repeat orbit interferometry package released," *Eos, Transactions American Geophysical Union*, vol. 85, no. 5, pp. 47–47, 2004.

[5] A.-M. Rosu, M. Pierrot-Deseilligny, A. Delorme, R. Binet, and Y. Klinger, "Measurement of ground displacement from optical satellite image correlation using the free open-source software micmac," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 100, pp. 48–59, 2015.

[6] E. Rupnik, M. Daakir, and M. Pierrot Deseilligny, "Micmac–a free, open-source solution for photogrammetry," *Open geospatial data, software and standards*, vol. 2, no. 1, pp. 1–9, 2017.

[7] C. Milliner and A. Donnellan, "Using daily observations from planet labs satellite imagery to separate the surface deformation between the 4 july m w 6.4 foreshock and 5 july m w 7.1 mainshock during the 2019 ridgecrest earthquake sequence," *Seismological Research Letters*, vol. 91, no. 4, pp. 1986–1997, 2020.

[8] N. Van Puymbroeck, R. Michel, R. Binet, J.-P. Avouac, and J. Taboury, "Measuring earthquakes from optical satellite images," *Applied Optics*, vol. 39, no. 20, pp. 3486–3494, 2000.

[9] S. Leprince, S. Barbot, F. Ayoub, and J.-P. Avouac, "Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1529–1558, 2007.

[10] X. Tong, Z. Ye, Y. Xu, S. Liu, L. Li, H. Xie, and T. Li, "A novel subpixel phase correlation method using singular value decomposition and unified random sample consensus," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4143–4156, 2015.

[11] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration–a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.

[12] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 148–164, 2018.

[13] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2758–2766.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[15] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2462–2470.

[16] S. Boukhtache, K. Abdelouahab, F. Berry, B. Blaysat, M. Grediac, and F. Sur, "When deep learning meets digital image correlation," *Optics and Lasers in Engineering*, vol. 136, p. 106308, 2021.

[17] R. Yang, Y. Li, D. Zeng, and P. Guo, "Deep dic: Deep learning-based digital image correlation for end-to-end displacement and strain measurement," *Journal of Materials Processing Technology*, vol. 302, p. 117474, 2022.

[18] T. Montagnon, J. Hollingsworth, E. Pathier, M. Marchandon, M. Dalla Mura, and S. Giffard-Roisin, "A new deep-learning approach for the sub-pixel registration of satellite images containing sharp displacement discontinuities," in *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2023, pp. 5123–5126.

[19] D. Engwirda, "Locally optimal delaunay-refinement and optimisation-based mesh generation," PhD thesis, Faculty of Science, School of Mathematics and Statistics, University of Sydney, 11 2014.

[20] R. Wiemker, "Registration of airborne scanner imagery using akima local quintic polynomial interpolation," in *Presented at the Second International Airborne Remote Sensing Conference and Exhibition*, vol. 24, 1996, p. 27.