Sub-Pixel Optical Satellite Image Registration For Ground Deformation Using Deep Learning

Tristan Montagnon, James Hollingsworth, Erwan Pathier, Mathilde Marchandon, Mauro Dalla Mura, Sophie Giffard-Roisin

SUB-PIXEL OPTICAL SATELLITE IMAGE REGISTRATION FOR GROUND DEFORMATION USING DEEP LEARNING

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

[1] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, Grenoble, France
[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France
Tokyo Tech World Research Hub Initiative (WRHI), School of Computing, Tokyo Institute of Technology, Japan
Institut Universitaire de France (IUF), France

ABSTRACT

Precise estimation of ground displacement maps at regional scales from optical satellite imaging is fundamental for the comprehension of natural disasters, such as earthquakes. Current methods make use of correlation techniques between two acquisitions in order to retrieve a fractional pixel shift. Yet, differences in local lighting conditions between the two acquisitions can lead to high image differences which will bias the estimation of the displacement, and data-driven methods could have the ability to overcome these errors. From the generation of a realistic simulated database based on Landsat-8 satellite pairs of images with added simulated shifts, we developed a Convolutional Neural Network (CNN) able to retrieve a sub-pixel displacement.

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

1. INTRODUCTION

Image registration is a key operation in image processing with applications in different domains, such as computer vision, biomedical imaging and remote sensing [1]. Focusing on the latter, Digital Image Correlation (DIC) has revolutionized satellite geodesy and particularly the study of ground deformation, peculiarly useful for natural hazards such as earthquakes. DIC is used to retrieve displacements fields between two images acquired on two different dates (days, weeks, months, or ever years). The use of optical satellite images is particularly efficient for capturing the ground displacement over large regions in a quick, cheap, and efficient way. However, this task is relatively challenging, as ground displacements are largely in the sub-pixel domain (given the typical range of earthquake displacements, e.g. 0-15 m, and pixel resolution of commonly used satellite imagery, e.g. Landsat-8: 15 m, Sentinel-2: 10 m).

Thanks to IDEX Université Grenoble Alpes for funding, and GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities, for the computations.

Current remote sensing sub-pixel registration methods rely on a sliding window approach, by assuming a uniform shift at a local scale. Image registration is often undertaken in the spatial domain, or the frequency domain [2, 3, 4]. Local spatial cross correlation coupled with SINC-based resampling, non-linear cost-functions [5] and regularization [6] has been implemented in the MicMac software [7]. It allows a robust correlation computation, even with small correlation windows. However, this procedure can suffer from heavy computations. Phase correlation for sub-pixel registration [8] is another efficient technique, based on frequency-domain correlation. The relative displacement between two images is estimated from the phase difference of their Fourier transforms. Splitting the sub-pixel problem in two steps (first, a pixel-wise displacement estimation, second, a sub-pixel displacement estimation, using a minimization algorithm [9]) is an efficient approach that has been developed in the software COSI-Corr [10]. It can reach an accuracy of 0.1 pixel. Phase correlation works well with large correlation windows, but is less satisfactory when using smaller correlation windows (due to increased sensitivity to high frequency noise). Therefore, phase correlation may be less suitable for obtaining spatial detail compared with more robust spatial domain methods.

Image registration and displacement field estimation have been successfully addressed by data-based methods and particularly Deep Learning (DL) models (in the field of medical-imaging [11], for video surveillance, robotics, and self-driving systems [12]). The advantages of DL are that it exploits the large amount of available data as well as its particular structure (such as Convolutional Neural Networks (CNN) for image-like data [13]). Most of the work in DL applied to image registration is based on optical flow and motion estimation, such as the model FlowNet [14], using CNNs to estimate the motion of objects. However, the large majority of registration problems involve estimating large displacements ($>$ 1 pixel), while the estimation of sub-pixel shifts has been less studied. In the field of material science, recent experiments using DL to retrieve sub-pixel displace-

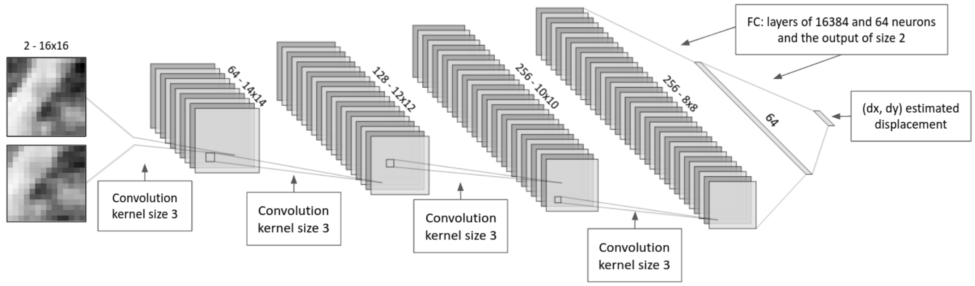

Fig. 1. The proposed CNN model for sub-pixel shift estimation from pairs of remote sensing windows.

ment and/or strain fields [15, 16] have been conducted on speckle images, demonstrating the feasibility of data-based approaches for sub-pixel measurements. Yet, the data and the particular issues in remote sensing (such as differences in illumination and ground evolution) are substantially different, so these tools can not be used directly.

Based on the promising application of DL in similar domains, we propose, for the first time (to the best of our knowledge), a DL solution for ground deformation estimation using remotely sensed optical satellite images. In this preliminary study, we develop a specific CNN to reproduce a correlator that estimates sub-pixel surface displacements (e.g. for an area subject to a ground motion, such as an earthquake), using satellite imagery. We formulate the problem as a uniform shift between a pair of tiles and estimate the 2D displacement for each window independently, similar to current correlator methods [10, 7]. This proposed method has the potential to overcome issues related to image differences due to different acquisition times of the two images being correlated (differences in illumination conditions can bias the correlation [17], thus leading to registration inaccuracy correlated with topography, etc.). One major contribution of this work is the development of a variety of synthetic databases by combining real acquisitions with simulated sub-pixel shifts. Once trained on our synthetic database, our CNN model is applied on a distinct pair of satellite images using a sliding window process in order to infer the complete displacement map. We compare different models by varying the training database settings and evaluate the precision of the estimated shifts. A complete comparison with a phase cross correlation method [18] and with a state-of-the-art correlation tool optimized for retrieving earthquake ground displacements (COSI-Corr [10]) is performed on a realistic earthquake experimental dataset.

2. METHODOLOGY

2.1. Problem statement

We consider two windows W_1 and W_2 linked by a rigid 2D translation (d_x, d_y). The image registration problem using traditional correlation is solved by identifying a translation vector $(\hat{d_x}, \hat{d_y})$ such that the correspondence between $W_1(x, y)$ and $W_2(x + \hat{d_x}, y + \hat{d_y})$ is maximized (according to some measure of correspondence). The shift between the two images can be addressed by finding

$$(\hat{d_x}, \hat{d_y}) = \text{argmax}_{(d_x,d_y)}\, r\left(I_1(x, y), I_2(x + d_x, y + d_y)\right),$$

in which r refers to cross-correlation. In this work we consider it as a regression problem on a local scale, with a data-driven approach, in which $(\hat{d_x}, \hat{d_y}) = \hat{f}(W_1, W_2)$ where f is a function learned from a set of N reference training data $\{X_n; Y_n\}_{n=1}^{N}$ with $X_n = (W_1, W_2)_n$ and $Y_n = (d_x, d_y)_n$ such that

$$\hat{f} = \text{argmin}_f\, \mathcal{L}\left(f(X_n), Y_n\right),$$

with $\mathcal{L}(\cdot, \cdot)$ being a function computing the loss between the displacement estimated by a regressor f and the reference displacement (here taken as the mean squared error). The proposed approach is to approximate a rigid and non-rotating transformation, by evaluating the displacement (dx, dy) for every image patch of size $k \times k$ (with k the size of the correlation window in pixels).

2.2. Extraction of a rigid 2D shift with a Convolutional Neural Network (CNN)

In this work, a CNN is used to retrieve this bi-directional displacement. Figure 1 gives an overview of the architecture of the network performing the correlation. As input, the model processes a pair of windows of size $k \times k = 16 \times 16$ pixels. This size of window allows to have a sufficient amount of data (pixels) for the model to retrieve the shift, yet small enough to have a relatively good spatial resolution. This pair represents a pre- and post-event window spanning a source of ground deformation (such as an earthquake). For the architecture, convolutional layers are used to extract feature maps, and to reduce the size of the data (as the window is small and no padding was used). Four layers using an increasing number of 3×3 kernels are used (see Figure 1 for the details). The chosen hyper-parameters represent a trade-off between computation time and accuracy. Based on our experiments, increasing the number of feature maps per layer instead of increasing the number of layers is more valuable, and going above four layers does not improve the results. The size of the kernels 3×3 was selected in order to extract small features. A Fully Connected (FC) layer is attached after this structure to reshape the data, and to output a vector (two values) representing the shift in two directions (horizontal and vertical).

2.3. Synthetic data generation

As we use a supervised machine learning approach, a reliable training set for learning the image registration operator is required. As too few datasets with real images and reference displacements are available, we create realistic synthetic data for this process. We define two windows W_1 and W_2, taken at two different dates t_1 and t_2. We consider a synthetic displacement field $D : R^2 \to R^2$ with $D(x, y) = (d_x, d_y)$ with $(d_x, d_y) \in [-1, 1]^2$. In order to simulate a uniform shift, a 2D translation on every window is artificially applied, by cropping the window shifted by a fraction of a pixel in row and column (giving a 2D synthetic offset), from the larger satellite image. A re-sampling algorithm is necessary, because the shift is sub-pixel. In this process, the Lanczos 6×6 kernel size [19] is used for the interpolation. This way, we can incorporate a uniform synthetic displacement field $D(x, y)$ in a pair of windows. Three different training datasets are created and standardized (re-scaled with a 0-mean and a unit variance) from real Landsat-8 acquisitions, in order to evaluate which type of data allows the model to reach the highest accuracy (see Table 1):

No noise dataset (NN dataset): pairs of windows are built from the same satellite image. We apply a random shift on half of the sample pairs and the rest is kept with a zero shift. We have: $W_{2,\mathrm{NN}}(x, y) = W_1(x + d_x, y + d_y)$;

Synthetic noise dataset (SN dataset): pairs of windows are also built from the same image, but a synthetic Gaussian noise is added within the second image. Similarly, we applied a random shift on half of the samples. We have: $W_{2,\mathrm{SN}}(x, y) = W_1(x+d_x, y+d_y)+U(x, y)$, where $U(x, y)$ is uniform random noise with a dynamic range consistent with the global image;

Difference in Acquisition Time dataset (DATe dataset): pairs of windows of the same area are taken at a different acquisition times (supposedly without any pre-existing displacement). This allows to create samples containing realistic perturbations in illumination, vegetation, etc. Again, we applied a random sub-pixel shift on half of the samples. We have $W_{2,\mathrm{DATe}}(x, y) = g_{\Delta t}(W_1(x, y))(x + d_x, y + d_y)$ with g the natural evolution of the ground acquisition during Δt.

The CNN model is trained with each dataset separately, and also with a combination of the three datasets. Therefore, four distinctly trained models are evaluated and compared to the phase cross correlation (PCC) method from the scikit-image python library. The evaluation is made on similar Landsat-8 data, albeit imagery acquired on different dates and at different locations, to guarantee that there is no common data used in both training and evaluation.

Dataset	Noise	Δt Acquisition time difference	Nb samples train	eval.
NN	0	0	6000	4000
SN	Gaussian	0	6000	4000
DATe	”Real”	1 to 24 months	9000	4000

Table 1. Description of the datasets used to train and evaluate the model.

3. RESULTS

3.1. Evaluation of the models

We use Mean Average Error (MAE) to compare the precision of the outputs of the models and the PCC method. This criterion evaluates the absolute difference between the prediction and the true value of the shift vector (in pixels).

	SN dataset no shift	SN dataset shift	DATe dataset no shift	DATe dataset shift
PCC	**0.003**	0.183	**0.051**	0.234
CNN-NN	0.052	0.065	0.127	0.157
CNN-SN	0.037	0.063	0.089	0.152
CNN-DATe	0.015	0.068	0.056	**0.125**
CNN-all	0.015	**0.054**	0.058	0.128

Table 2. Mean absolute error (in pixels) on the SN and DATe evaluation datasets, for the PCC and different CNN models.

Table 2 compares the four CNN models and PCC, evaluated on DATe and SN test datasets. On non-shifted data, the PCC method performs better than our trained models, with very similar results between PCC, CNN-DATe and CNN-all on DATe non-shifted. On shifted data, the CNN-all and CNN-DATe have the best score. This shows that our method is able to retrieve accurate displacements vectors. The model that gives the lowest MAE on the most realistic data (DATe dataset) is CNN-DATe. Therefore, this latter model has been selected for the following evaluations on larger images.

3.2. Results on realistic synthetic regional data

The second stage of validation addresses an effective way to evaluate the precision of our CNN by comparing it against PCC and COSI-Corr on a test case using Landsat-8 satellite images re-sampled to include a realistic synthetic offset.

We developed an algorithm that randomly generates realistic synthetic displacement fields. This physics based algorithm is able to mimic an earthquake realistic rough surface discontinuity and associated displacement field (with surface displacement fields computed considering an homogeneous elastic half-space [20]). These displacement fields $D(x, y)$

are used to warp satellite images using a quintic spline re-sampling algorithm [21]. Here, $D(x, y)$ is now not uniform, as it describes a realistic fault, and the warped satellite images are larger than before ($k = 128$). The pair of images generated and the synthetic displacement map are given in Figure 2. We apply the different models (CNN, PCC and COSI-Corr) as a sliding 16×16 window to obtain the displacement maps.

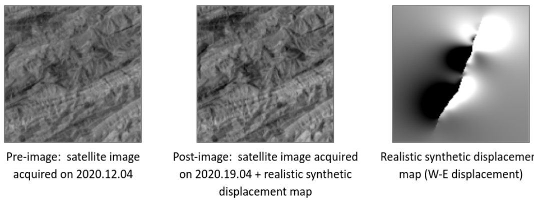

Fig. 2. Pre-image (left) and post-image (center) warped with the synthetic displacement map (right).

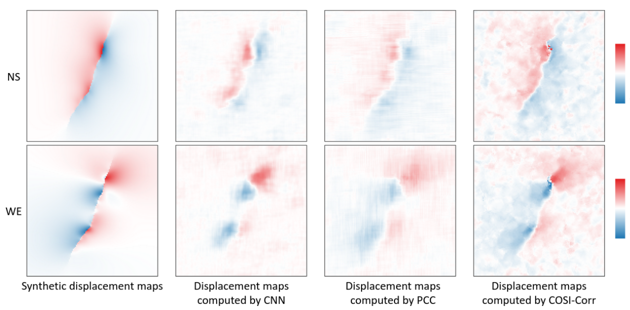

Fig. 3. (top) North(red)-South(blue) displacement maps; (bottom) West(red)-East(blue) displacement maps for the CNN, PCC, and COSI-Corr approaches.

From Figure 3, all three methods recover the first-order features of the input synthetic displacement maps. However, because the displacement maps are constructed using a sliding window operation, the resulting displacement field is subject to a spatial smoothing controlled by the correlation window size. Furthermore, whenever the correlation window straddles the discontinuity (representing the earthquake ground rupture), the assumption of a simple 2D translation breaks down, and the retrieved values will be biased. N-S residual displacement maps (correlation minus synthetic) are shown in Figure 4. The large residual close to the synthetic discontinuity reflects the inability of the correlation window to reliably capture the displacement when the window crosses the discontinuity. Table 3 gives the mean, the standard deviation (std), and the maximum error around the fault artificially created. PCC has high mean and std error for the N-S component, while CNN is closer to the state-of-the-art COSI-Corr. The latter contains the best mean and std results, yet with some outliers (larger maximal errors).

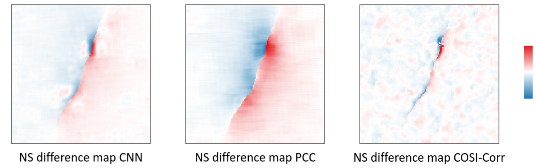

Fig. 4. North(red)-South(blue) residual displacement maps

	Mean		Std		Max	
	WE	NS	WE	NS	WE	NS
CNN	0.066	0.059	0.065	0.077	0.97	0.98
PCC	0.047	0.12	0.069	0.14	0.97	1.14
C-Corr	0.041	0.046	0.059	0.071	1.74	1.81

Table 3. Displacement errors (mean, standard deviation and maximum) of our CNN, PCC and C-Corr correlators.

4. DISCUSSION AND CONCLUSIONS

This paper presented a technique to perform estimation of ground deformation using satellite images (medium-resolution Landsat-8), based on CNN. 4 models were evaluated, and the one with most accurate results was selected, with a precision below 0.1 pixel. The main purpose of the study was to demonstrate, with an evaluation on realistic synthetic displacement maps, the feasibility of a machine learning approach for accurate sub-pixel measurement in the context of surface deformation estimation from remotely sensed satellite imagery. We showed that we already have competitive results with respect to state-of-the-art methods.

Yet, various issues remain to be further explored to improve the results. One significant limitation is the calculation of the sub-pixel shifted image when creating synthetic data: fractional pixel re-sampling can introduce bias and error in the training data which will ultimately limit the precision we can currently achieve. To some extent, this bias can be learned, which likely explains why we achieve higher accuracy than traditional phase correlation methods (which do not learn). However this bias does not exist in real images. In a future development, we will also add non-uniform shifts when training in order to be more robust to real displacements. The architecture of our model could also be improved in order to operate with different sizes of windows, or even to output a full displacement field instead of a 2D local vector (e.g. using U-net architectures [15]). Finally, the advantage of a data-driven approach is that additional data could be added in the training step to give the model more information in order to learn the subtle relationships which contribute to perceived noise in the final correlation maps. For example, adding illumination conditions associated with a particular image may allow the illumination bias to be learned. With this very encouraging preliminary study, we hope that future developments of data-driven sub-pixel registration techniques will be able to improve the current resolution of ground displacement maps.

5. REFERENCES

[1] Barbara Zitova and Jan Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[2] R. Manduchi and G.A. Mian, "Accuracy analysis for correlation-based image registration algorithms," in *1993 IEEE International Symposium on Circuits and Systems*, 1993, pp. 834–837 vol.1.

[3] Paul E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques," *IEEE Transactions on Geoscience Electronics*, vol. 8, no. 4, pp. 353–368, 1970.

[4] John A. Leese, Charles S. Novak, and Bruce B. Clark, "An Automated Technique for Obtaining Cloud Motion from Geosynchronous Satellite Data Using Cross Correlation.," *Journal of Applied Meteorology*, vol. 10, no. 1, pp. 118–132, Feb. 1971.

[5] P. Thevenaz, T. Blu, and M. Unser, "Interpolation revisited [medical images application]," *IEEE Transactions on Medical Imaging*, vol. 19, no. 7, pp. 739–758, 2000.

[6] M. Deseilligny and N. Paparoditis, "A multiresolution and optimization-based image matching approach: An application to surface reconstruction from spot5-hrs stereo imagery," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36-1, 05 2012.

[7] Ana-Maria Rosu, Marc Deseilligny, Arthur Delorme, Renaud Binet, and Yann 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, 04 2014.

[8] Hassan Foroosh, Josiane Zerubia, and Marc Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11 3, pp. 188–200, 2002.

[9] N. Van Puymbroeck, R. Michel, R. Binet, and J. Taboury, "Measuring earthquakes from optical satellite images," *Appl. Opt*, pp. 3486–3494, 2000.

[10] Leprince S., Ayoub F., Klinger Y., and Avouac J.-P., "Co-registration of optically sensed images and correlation (cosi-corr): an operational methodology for ground deformation measurements," in *2007 IEEE International Geoscience and Remote Sensing Symposium*, 2007, pp. 1943–1946.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, 09 2014.

[12] Ajeet Ram Pathak, Manjusha Pandey, and Siddharth Rautaray, "Application of deep learning for object detection," *Procedia Computer Science*, vol. 132, pp. 1706–1717, 2018, International Conference on Computational Intelligence and Data Science.

[13] A. Krizhevsky, I. Sutskever, and G. E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. 2012, vol. 25, Curran Associates, Inc.

[14] P. Fischer, A. Dosovitskiy, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," *CoRR*, vol. abs/1504.06852, 2015.

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

[16] Ru Yang, Yang Li, Danielle Zeng, and Ping Guo, "Deep dic: Deep learning-based digital image correlation for end-to-end displacement and strain measurement," *Journal of Materials Processing Technology*, vol. 302, pp. 117474, Apr 2022.

[17] P. Lacroix, G. Araujo, J. Hollingsworth, and E. Taipe, "Self-entrainment motion of a slow-moving Landslide inferred from Landsat-8 time series," *Journal of Geophysical Research : Earth Surface*, vol. 124, pp. 1201–1216, 2019.

[18] Manuel Guizar-Sicairos, Samuel Thurman, and James Fienup, "Efficient subpixel image registration algorithms," *Optics letters*, vol. 33, pp. 156–8, 02 2008.

[19] B. N. Madhukar and R. Narendra, "Lanczos resampling for the digital processing of remotely sensed images," in *Proceedings of International Conference on VLSI, Communication, Advanced Devices, Signals & Systems and Networking (VCASAN-2013)*, Veena S. Chakravarthi, Yasha Jyothi M. Shirur, and Rekha Prasad, Eds., India, 2013, pp. 403–411, Springer India.

[20] Brendan Meade, "Algorithms for the calculation of exact displacements, strains, and stresses for triangular dislocation elements in a uniform elastic half space," *Computers Geosciences*, 08 2007.

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