A New Deep-Learning Approach for the Sub-Pixel Registration of Satellite Images Containing Sharp Displacement Discontinuities

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

A NEW DEEP-LEARNING APPROACH FOR THE SUB-PIXEL REGISTRATION OF SATELLITE IMAGES CONTAINING SHARP DISPLACEMENT DISCONTINUITIES

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

Image correlation is a powerful method for remotely constraining ground displacements associated with natural disasters. By employing sub-pixel correlation algorithms, one can obtain a displacement field by correlating satellite images acquired before and after a displacement event. However, this computation may be biased when dealing with sharp discontinuities, typical of earthquake surface ruptures, which are of current interest in the context of quantifying the partitioning of slip between the primary fault core and neighboring damage zone. In this paper, we present an innovative deep learning method to perform sub-pixel correlation of optical satellite images for the retrieval of ground displacement, designed to mitigate bias around fault ruptures. From the generation of a realistic simulated database of images before and after synthetic ground displacement built specifically to deal with fault discontinuities in satellite images (e.g. Landsat-8 in this case), we developed a Convolutional Neural Network (CNN) able to retrieve sub-pixel displacements. Comparison with a state-of-the-art phase correlation method shows our pipeline is able to mitigate the sub-pixel bias in the near-field of earthquake ruptures.

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

1. INTRODUCTION

Precise estimation of ground displacement at regional scales from optical satellite imagery is fundamental for the study of natural disasters, such as earthquakes, volcanoes, landslides, etc. In the case of earthquakes, characterizing the near-field displacement around surface ruptures provides valuable constraints needed to understand the physics of earthquake slip, and to anticipate the seismic hazard posed to neighboring infrastructure and populations. Therefore, a precise and unbiased estimation of ground deformation is crucial to address the location, geometry, and slip distribution of the fault.

Image correlation is a technique used to measure spatial changes (i.e. ground displacements) between two satellite images. This problem is efficiently solved by traditional registration methods, such as Phase Correlation [1], or Spatial Correlation [2], and can attain sub-pixel precision [3], [4]. These methods have been optimized and implemented in various software packages such as COSI-Corr [5] and MicMac [6], which are commonly used to measure earthquake deformation from aerial or satellite images. Current correlation methods rely on the same approach: (1) they work at a local scale, with small sliding windows extracted from a pair of co-registered optical images acquired at different times, and (2) they assume a rigid uniform shift between the two correlation windows. These conditions are appropriate in the majority of cases, yielding maps of 2-D ground displacement for large shallow earthquakes, or other sources of ground motion. However, in the near-field of fault ruptures, where the correlation window spans a sharp fault discontinuity, this hypothesis breaks down, and may bias the displacement measurements, which in turn will impact the use of near-field data in addressing the physics of fault slip.

Retrieval of sub-pixel ground displacements from optical satellite images using a deep learning framework was recently demonstrated for the first time [7]. A Convolutional Neural Network (CNN) was developed to solve the sub-pixel displacement estimation problem. However, their model was not able to address the discontinuity bias, as they imposed a local rigid transformation when generating their training data.

Thanks to IDEX Université Grenoble Alpes, INSU PNTS, CNES, and CDP Risk for funding, and GRICAD infrastructure (gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities, for the computations. ISTerre is part of Labex OSUG@2020 (ANR10 LABX56).

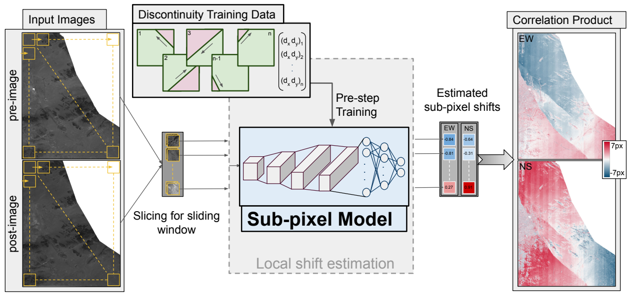


Fig. 1. From left to right: A pair of input images (pre and post) are acquired on two different dates. The local sub-pixel displacement between the two images is retrieved using a sub-pixel model; displacements given in the row (North-South) and column (East-West) direction. Our model is trained with data containing discontinuities.

2. METHODOLOGY

In our study, we propose to extend the method of [7] to address the specific challenge faced when discontinuities are present within a correlation window, e.g. in the near-field of fault ruptures, leading to biased estimates of displacement in zones adjacent to the discontinuity. We achieve this by training an end-to-end model on a more-realistic synthetic dataset that includes samples mimicking real fault discontinuities. This paper contributes to the study of sub-pixel registration in three major aspects:

1) Creation of a synthetic training dataset that allows data-based techniques to learn how to retrieve sub-pixel surface displacements in the presence of sharp discontinuities (i.e. where the wavelength of the displacement gradient is smaller than the correlation window);

2) Development of a CNN able to precisely estimate the displacement field near discontinuities;

3) Comparisons with COSI-Corr [5] to evaluate our method quantitatively against synthetic realistic images.

The initial statement of the problem relies on the same principle of state-of-the-art solutions: we work at the local scale with two small windows W_1 and W_2, of size $k \times k$ (with k the size of the correlation window in pixels, and we will take by default $k = 16$). We also make the same assumption: the model tries to estimate a rigid displacement between the two windows, represented by a translation (d_x, d_y).

2.1. Sub-pixel Convolutional Neural Network

A Convolutional Neural Network (CNN) is a type of neural network architecture that is designed for image recognition and other tasks that involve processing pixel data. It generally has three main types of layers, which are convolutional layer, pooling layer, and fully-connected (FC) layer that work together to learn features in the input data and make predictions based on those extracted patterns.

To perform the sub-pixel correlation, we developed a CNN architecture that takes as input two 16-by-16 pixel patches, given as $2 \times 16 \times 16$ tensors. The output is a vector of two values, representing the estimated shift between the two input frames. The architecture of our network can be summarized as follows: The input is passed through four convolutional layers, with an increasing number of small kernels (64, 128, 256 and 256). The size of the kernels (3×3) was selected in order to extract small features in already small (16×16) windows, and is well-suited to work on small displacements [8]. The output of each layer is processed by the ReLU activation function, mainly to overcome vanishing gradient problems. This method is effectively the default activation function for such a large network. After convolutions, 2 fully connected layers reduce the size of the data, from 16384 to 64 to 2, and outputs the estimated shift.

2.2. Generation of the training database

In the Earth Science community, no large synthetic dataset has yet been created to help solve this ground displacement estimation problem. Furthermore, real data (i.e. spanning real earthquakes) cannot be used to build a relevant database, due to a lack of ground truth. Therefore, we generate our own dataset to train our network.

The purpose of our model is to retrieve deformation between two windows, given as input. Thus, we need a training dataset that contains pairs of pixel patches (inputs) linked by ground truth distortions (targets). Note that in order to create a dataset as close as possible from the ground truth, we use real Landsat-8 satellite images, to generate samples containing realistic perturbations in illumination, vegetation, topographically-correlated noise, etc. For this, we need well-registered satellite images with as few artifacts as possible. We generally focus on images covering arid regions, which feature more limited changes in surface reflectance with time.

We define two windows: W_1 and W_2, extracted from two large satellite images: I_1 and I_2 acquired on two different dates: t_1 and t_2, over the same location. We consider a synthetic displacement field $D(x, y)$, which is used to warp W_2 to obtain W_{2s}, the distorted version of W_2. One unit of the training dataset is the standardized pair (W_1, W_{2s}) (re-scaled with a zero-mean and a unit variance), with the associated deformation $D(x, y)$. A re-sampling algorithm is necessary during the warping process, because the shift applied is sub-pixel: a Lanczos kernel (6×6) is used for the interpolation, to minimize resampling artifacts. This procedure can be summarized as follows:

$$W_{2s}(x, y) = e_{\Delta t}(f_D(W_1(x, y)))(x, y)$$

with e the natural evolution of the ground acquisition during Δt, and f the distortion operation associated to D. This process is repeated randomly (random $D(x, y)$, random window extraction location, random pair (I_1, I_2), following a uniform distribution), to create a large and relevant number of unique samples.

To overcome the assumption of uniform shifts in the training data, and thus to learn how to retrieve displacements in the presence of discontinuities, we built the discontinuity dataset (DIS), where $D(x, y)$ used to warp W_2 is not uniform, but contains a discontinuity. Formally,

$$D(x, y) = \begin{cases} D_a(x, y) = (d_{x_a}, d_{y_a}) & \text{if } (x, y) \in A \\ D_b(x, y) = (d_{x_b}, d_{y_b}) & \text{if } (x, y) \in B \end{cases} \quad (1)$$

where A and B are the two areas created by intersecting a random line with the correlation window (See the Discontinuity Training Data square in Figure 1). The idea behind this approach is that we want the model to handle discontinuities, by identifying the area of interest in a given pair of windows, and use only this area to retrieve the corresponding shift. The desired output is still a vector of two values, but the input window W_2 is warped with a discontinuity to mimic a real fault. 125k samples are created for training and 25k for test.

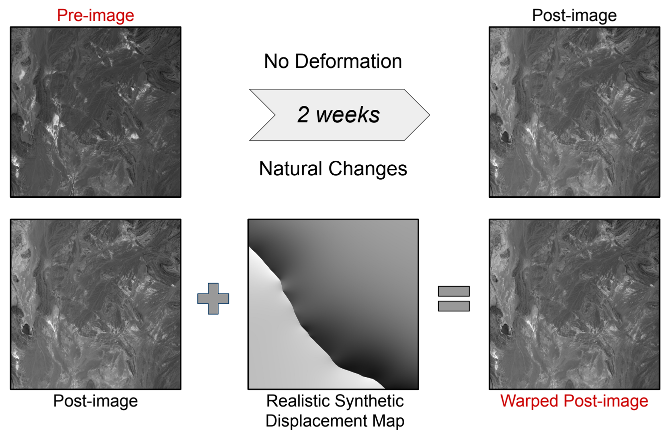


Fig. 2. Creation of a pair of realistic synthetic earthquake images. Here, the Warped Post-image (bottom right) contains natural changes due to the different acquisition time of Pre-image (top left), and carries the synthetic displacement map (bottom center), that our model should retrieve.

3. RESULTS

3.1. Generation of realistic synthetic earthquake images

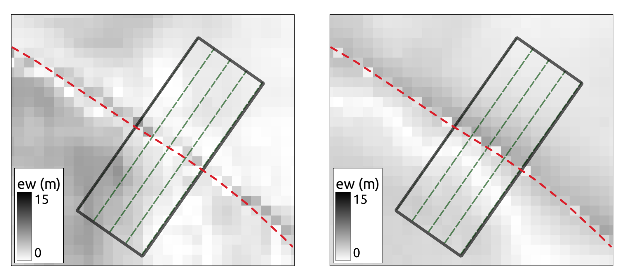
We developed an algorithm that computes synthetic surface displacements for randomly generated realistic fault discontinuities with rough (fractal) geometries and slip distributions embedded in a homogeneous elastic halfspace [9]. These displacement fields $D(x, y)$ are used to warp satellite images using a quintic spline re-sampling algorithm [10]. Here, $D(x, y)$ is now not uniform, as it describes a realistic fault displacement, and the warped satellite images are much larger ($k = 1024$). The creation of input images is summarized in Figure 2. This dataset is only used for evaluating our model. We compare two satellite images, one with a specific acquisition time, and a second taken after the first one, and warped with $D(x, y)$. We apply the procedure 3 times, with 3 different satellite images, and 3 different displacement fields, to evaluate the precision on different cases.

Error between the different models is assessed using the residuals (absolute difference) computed between the output correlation maps and the known synthetic displacement maps. The mean of each statistical measurement is also estimated to obtain average evaluations.

3.2. Comparisons with COSI-Corr

We apply the two models, CNN and COSI-Corr, as a sliding ($k \times k$) window to obtain the large-scale displacement maps. COSI-Corr was applied with a 32×32 window, although the effective width is reduced almost by half due to the windowing function which is applied to mitigate spectral leakage when computing the Fast Fourier Transform of the two images [11]. COSI-Corr makes use of a frequency masking scheme to mitigate the impact of noisy high frequencies on the correlation estimation. We compare our model trained with the DIS dataset with COSI-Corr (which is commonly used to study near-field deformation in earthquakes), with a particular focus on the sub-pixel performance close to the fault discontinuity.

We compute the residual maps to assess the bias of the different methods; see the example Figure 3 close to fault.



(a) CNN (b) COSI-Corr

Fig. 3. East-West close ups of residual maps (absolute displacement error), in meters. Red line shows the discontinuity trace. Dark green dashed lines show fault-perpendicular profile locations from Figure 4.

We also compute statistics in the vicinity of the discontinuity. The precision of our model (0.12 px) exceeds that of COSI-Corr (0.15 px). To compute these means, we set all absolute displacement residuals larger than 1 pixel (15 m for Landsat-8) to 1 pixel, to prevent unrealistically large values having a large impact on the statistics (note: this only affects the COSI-Corr results, because the error in our model never exceeds 1 pixel, while COSI-Corr can reach 8+ pixels or 130

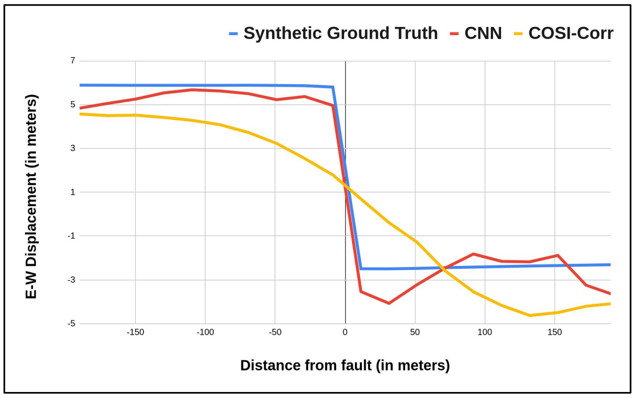


Fig. 4. Profile across fault of the synthetic ground truth, CNN and COSI-Corr correlation maps. This profile is computed by averaging three neighbouring profiles. See Figure 3 for profile locations.

meters).

Additionally, the spatial distance perpendicular to the fault-discontinuity over which the retrieved displacements are biased is narrower (1-2 px / 15-30 m) in our model compare with COSI-Corr (8 px / 120 m, i.e. half the correlation window) - see Figure 4.

4. CONCLUSIONS

In this study, we address the problem of sub-pixel bias close to sharp discontinuities, when retrieving displacement fields from correlation of two optical images using traditional phase or spatial cross correlation techniques. This near-field bias is the result of spatial smoothing of the displacement field by the correlation window, coupled with a break-down in the underlying assumption of a rigid-body translation during correlation. We approach this problem by training a sub-pixel registration CNN using realistic synthetic displacement data which includes discontinuities within the correlation window. This approach gives a significant improvement in the accuracy of displacements retrieved close to sharp discontinuities, while reducing the spatial length-scale of the bias, thereby allowing for a more accurate description of the near-field displacement. If left unaccounted for, this bias can complicate our view of how earthquake ruptures break the Earth's surface, thus giving an incorrect description of how displacement is partitioned between the primary fault plane and the surrounding damage zone. Better characterization of these processes will help to address the physics of fault slip, and the expression of past earthquakes in the landscape, which forms an important component of Seismic Hazard Assessment.

5. REFERENCES

[1] Charles D Kuglin, "The phase correlation image alignment methed," in *Proc. Int. Conference Cybernetics Society*, 1975, pp. 163–165.

[2] J Lewis, "Fast template matching. 1995," *URL http://www. scribblethink. org/Work/nvisionInterface/vi95_lewis. pdf. Cited on*, p. 6, 1995.

[3] Hassan Foroosh, Josiane B Zerubia, and Marc Berthod, "Extension of phase correlation to subpixel registration," *IEEE transactions on image processing*, vol. 11, no. 3, pp. 188–200, 2002.

[4] Irene G Karybali, Emmanouil Z Psarakis, Kostas Berberidis, and Georgios D Evangelidis, "An efficient spatial domain technique for subpixel image registration," *Signal Processing: Image Communication*, vol. 23, no. 9, pp. 711–724, 2008.

[5] Sébastien Leprince, François Ayoub, Yann Klinger, and Jean-Philippe Avouac, "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*. IEEE, 2007, pp. 1943–1946.

[6] Ana-Maria Rosu, Marc Pierrot-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, pp. 48–59, 2015.

[7] Tristan Montagnon, James Hollingsworth, Erwan Pathier, Mathilde Marchandon, Mauro Dalla Mura, and Sophie Giffard-Roisin, "Sub-pixel optical satellite image registration for ground deformation using deep learning," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 2716–2720.

[8] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *Proceedings of the IEEE CVPR conference*, 2017, pp. 2462–2470.

[9] Brendan J Meade, "Algorithms for the calculation of exact displacements, strains, and stresses for triangular dislocation elements in a uniform elastic half space," *Computers & geosciences*, vol. 33, no. 8, pp. 1064–1075, 2007.

[10] 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.

[11] T Heid and A Kääb, "Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery," *Remote Sensing of Environment*, vol. 118, pp. 339–355, 2012.