A Global Registration Method for Satellite Image Series

Charles Hessel, Carlo de Franchis, Gabriele Facciolo, Jean-Michel Morel

A GLOBAL REGISTRATION METHOD FOR SATELLITE IMAGE SERIES

Charles Hessel[⋆,†], *Carlo de Franchis*[⋆,†], *Gabriele Facciolo*[⋆] *and Jean-Michel Morel*[⋆]

[⋆] Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, France
[†] Kayrros SAS

ABSTRACT

Image registration is a fundamental tool of remote sensing. The recent proliferation of earth observation satellites has opened the way to the analysis of long image time series with denser temporal repetition. Given this wealth of images, it is crucial to design automatic tools to process them. We thus propose a method for the global registration of satellite image time series, that leverages their redundancy to improve in precision and robustness. By computing the relative displacement for all possible pairs of images, we are able to discard outliers and minimize the number of misaligned images. Experiments on synthetic data show that longer image series are registered with a higher precision.

Index Terms— multi-image alignment, registration, phase correlation

1. INTRODUCTION

Satellite image time series analysis requires precise image registration. However, due to inaccuracies in the orientation of the satellite sensors, pixels coordinates hardly ever exactly match their actual position on the ground. This, in turn, creates small displacements between consecutive images of time series. In this work, we tackle the global registration of such series; our objective is to bring all the images onto a common reference. Our contribution is a method to take full advantage of the redundancy in this multi-image alignment problem, including the detection and handling of outliers or misaligned images. This is achieved by the estimation of the displacement for all possible image pairs, using the phase correlation technique. Verification steps are based on a graph whose nodes are the images and links are the estimated shifts. This global registration algorithm is generic and can be used to register optical and radar image time series. Figure 1 shows the advantage of using several reference images.

Image registration methods are generally divided in two categories: feature-based methods and area-based methods [1].

Work partly financed by IDEX Paris-Saclay IDI 2016, ANR-11-IDEX-0003-02, Office of Naval research grant N00014-17-1-2552 and N00014-20-S-B001, DGA Astrid project « filmer la Terre » n° ANR-17-ASTR-0013-01, MENRT.

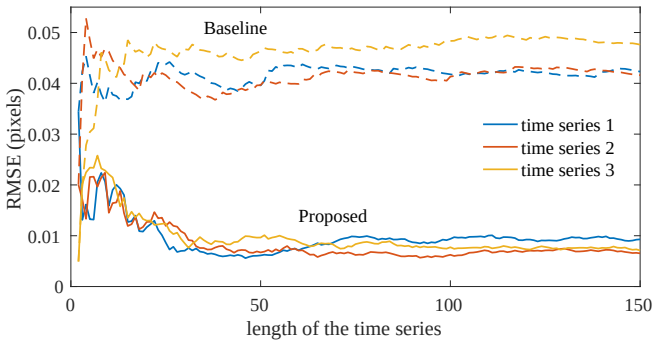


Fig. 1. Evolution of the RMSE on the estimated shifts in function of the number of images in the series, for three synthetic optical time series. Continuous lines are obtained with our method. Dotted lines are obtained using one reference image.

The latter includes correlation methods, among which the normalized cross correlation and other Fourier-based correlation methods, dating back to the seventies [2]. The particularly successful phase correlation method [3] normalizes the correlation image in the Fourier domain by its modulus, thus keeping only the phase information. This makes it invariant to global linear variations in contrast and brightness [4]. The sub-pixel displacement can be estimated from the cross correlation image [5, 6, 7] or by estimating the slope of a plane in the phase of its Fourier transform [8, 9]. Accuracy and robustness of correlation algorithms can be further improved by frequency weighting [8].

These methods only register two images. Moreover, in spite of all these improvements and advances, image registration often fails due to occlusions or extreme contrast changes, which in turn result in severe misalignment. A way around this is to leverage the availability of all images of the time series.

The problem of registering several images at once is called multi-image alignment. It consists in bringing a group of images into a common reference. In addition to being a critical first step in many remote sensing applications, it is also the cornerstone of many image fusion techniques such as high dynamic range imaging, super-resolution, burst deblurring and burst denoising [10]. One approach is to use a single reference and register all images onto it. However, this requires either to have an expected reference image, which, in most practical

situations, cannot be obtained, or to pick from the given series one image, including the degradations it may contain. In brief, there is much to gain with the registration on *several* reference images [11, 12].

Aguerrebere *et al.* [10] report approaches using the maximum likelihood estimator (MLE), the Bayesian MLE, the maximum a posteriori (MAP) estimator, and constrained alignment. Yet, despite the ubiquity of the problem and the number of methods proposed to address it, using all possible pairwise estimations is scarce. Methods in [13, 14] use this overdetermined system of equations by taking advantage of the fact that the shift between two distant images is the sum of the shifts of all adjacent images between them [10]. However, these methods do not include mechanisms to detect outliers and correct misaligned images, although they are critical in the remote sensing context.

We shall describe the proposed method in Section 2, then present in Section 3 an evaluation of its results.

2. METHOD DESCRIPTION

This section describes a global registration method for optical and SAR image time series. Our aim is to bring a series of unregistered images into a common reference, which is inferred from the input images themselves. Thus, we do not require a reference image. We start by presenting the phase correlation method, used to estimate translations in pairs of images, with two particular cases: the registration of two images, and the registration of one image onto several, already registred images. Then, we shall describe our global scheme to register several images without references, along with strategies to detect and remove outliers.

We shall denote by $\{u_0, ..., u_n, ..., u_{N-1}\}$ the input time series, composed of N images. The estimated registration shifts are denoted by $\{\hat{\tau}_0, \ldots, \hat{\tau}_n, \ldots, \hat{\tau}_{N-1}\}$. They translate the images to the reference point, that we define as the centroid of the estimated relative positions of the input images. Hence the reference is not known in advance. The shifts are finally applied to the images using spline interpolation of order five.

2.1. Phase correlation

Let us denote $U = \mathcal{F}\{u\}$ the discrete Fourier transform of image u. We call c the phase-correlation image between u and a second image v. Its discrete Fourier transform is given by

$$C = \frac{U \cdot V^*}{|U \cdot V^*|}, \tag{1}$$

where * denotes the complex conjugate and $|\cdot|$ the norm. The shift is given by the position of the maximum in $c = \mathcal{F}^{-1}\{C\}$. Some weights can be applied to C to reduce the effect of the borders and of the noise [8]. Since this estimation only gives integer translations, a second refinement step is needed to obtain a shift at sub-pixel precision. Assuming the maximum position at position $(x^{\max}, y^{\max})$, we fit a one-dimensional sinc function for the x-axis using the two neighboring points $x^{\max} - 1$ and $x^{\max} + 1$:

$$\underset{a_x, b_x, c_x}{\arg\min} \sum_{x=x^{\max}-1}^{x^{\max}+1} \|c(x, y^{\max}) - c_x \cdot \mathrm{sinc}(b_x \cdot (x - a_x))\|^2, \tag{2}$$

and similarly for the y-axis. Finally, the position of the maximum is $\hat{\tau} = (a_x, a_y)$. In this paper, we assume that $\hat{\tau} = \tau + \nu$, with τ the exact (and unknown) translation that registers image u onto the reference. The random perturbation ν models the method's inaccuracy; we assume its expected value to be null.

Often, we need to register a new image onto several already registered images. We denote these M reference images by $\{v_0, ..., v_m, ..., v_{M-1}\}$. The estimated translation for input u_n with the m^{th} reference image v_m is denoted $\hat{\tau}_{n,m}$. To limit the impact of outliers, the final shift is estimated as the the median of all shifts with the reference images, *i.e.*

$$\hat{\tau}_n = \mathrm{median}\{\hat{\tau}_{n,m}\}. \tag{3}$$

References are, however, not always available.

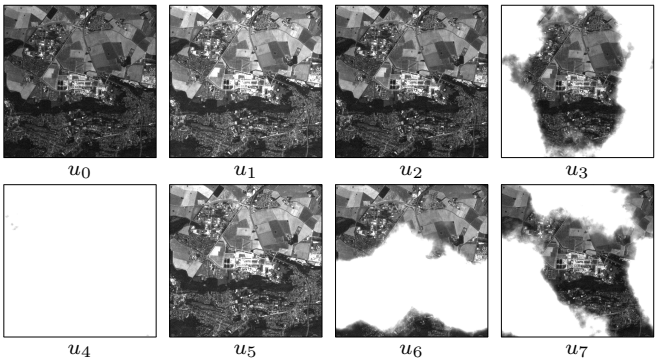


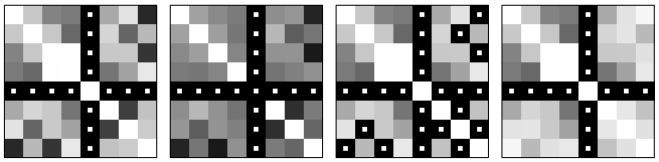
Fig. 2. Synthetic optical time series with eight images, and including opaque clouds.

2.2. Global registration

An example of candidate optical time series is given in Figure 2. It contains eight images, including one totally covered by a cloud. We synthesized these images following the procedure described in Section 3, hence we know the ground truth.

We start by estimating shifts for all possible pairs of images. We denote by $\hat{\tau}$ this matrix of size $N \times N$, whose elements $\hat{\tau}_{n,n'}$ are the estimated translations between images u_n and $u_{n'}$. This matrix is antisymmetric, and its main diagonal is always zero. The number of shifts to be estimated is therefore $N(N-1)/2$. Assuming an input time series without outliers and all shifts correctly estimated, the final shift $\hat{\tau}_n$ that registers image u_n onto the reference is obtained by averaging the shifts estimated with all other images, *i.e.*

$$\hat{\tau}_n = \tfrac{1}{N} \textstyle\sum_{n'=0}^{N-1} \hat{\tau}_{n,n'}. \tag{4}$$



(a) $\|\hat{\tau}^{\text{ini}} - \tau^{\text{GT}}\|$ (b) R (c) $\|\hat{\tau}^{\text{tri}} - \tau^{\text{GT}}\|$ (d) $\|\hat{\tau} - \tau^{\text{GT}}\|$

Fig. 3. Triconsistency and shift error matrices. Values go from 0 (white) to 0.09 pixels (black). Brighter is better. Dotted cells indicate discarded shifts. Matrix (b) is the triconsistency matrix R. Matrices (a,c,d) are the difference with the ground truth shifts at different steps of the method: initial ($\hat{\tau}^{\text{ini}}$), after the triconsistency test ($\hat{\tau}^{\text{tri}}$), and final ($\hat{\tau}$). Values are computed with the synthetic time series in Figure 2.

But this is not robust to outliers. Our practical approach therefore introduces two steps to detect, remove, and even correct the wrongly estimated shifts. The first is based on the correlation values. The second on the estimated shifts global consistency.

2.2.1. *Correlation values*

The phase correlation image has values in $[-1, 1]$, and its maximum $c^{\text{max}} = \max(c)$ is a good indicator of the number of pixels in agreement. It will be 1 if the compared images are identical (-1 if they have reversed contrast), and degrades towards 0 when they differ. We thus set a threshold, under which estimated shifts will be considered as erroneous

$$c^{\text{max}} < t^{\text{maxval}}, \tag{5}$$

where we set $t^{\text{maxval}} = 0$. Indeed, we do not need a very restrictive threshold since we perform a complementary validation test. This second test uses the ratio between c^{max} and second maximal value $c^{\text{second max}}$ found after discarding the maximal value and its eight neighbors in the image c.

$$\frac{c^{\text{max}}}{c^{\text{second max}}} < t^{\text{maxratio}}, \tag{6}$$

with t^{maxratio} empirically set to $10/6$. All shifts that do not pass these two tests are discarded.

In Figure 3 (a) we show in matrix form the norm of the differences between all $\hat{\tau}^{\text{ini}}$ and the corresponding ground truths for the series in Figure 2. Dotted cells represent discarded shifts. The image u_4, fully covered by clouds, did not pass the tests on the correlation values in all pairs.

The matrix $\hat{\tau}^{\text{ini}}$ is then interpreted as a graph, where two images (the nodes) are connected if their estimated shift (link) is not discarded, and disconnected otherwise. This graph is illustrated in Figure 4. We then keep only the largest connected component. In general, the isolated components removed at this step are due to cloudy images.

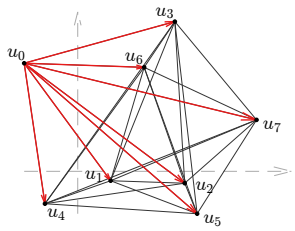


Fig. 4. Actual positions of eigth images with the same, noisy coordinates. So as to bring them all onto the same reference point, we compute shifts for all possible image pairs. They are displayed as black lines, or red arrows for image u_0.

2.2.2. *Global shifts consistency*

To further improve the precision of the registration, we detect and replace badly estimated shifts. Indeed, for any triplet of images $(u_k, u_{k'}, u_{k''})$, we have $\tau_{k,k'} + \tau_{k',k''} = \tau_{k,k''}$. This allows us to measure the consistency of a shift in this connected graph. We call this measure r the *triconsistency*, defined as

$$r_{k,k',k''} = \|\tau^{\text{ini}}_{k,k'} + \tau^{\text{ini}}_{k',k''} - \tau^{\text{ini}}_{k,k''}\|, \tag{7}$$

for any three different $k \in 0..(N-1)$. We compute the triconsistency matrix R of size $N \times N$ by averaging all triconsistency values for a particular pair of images:

$$R_{n,n'} = \frac{1}{N-2} \sum_{n'' \in \{0\ldots(N-1)\} \setminus n,n'\}} r_{n,n',n''}. \tag{8}$$

An example triconsistency matrix is given in Figure 3 (b). Triconsistency values are not computed for discarded shifts.

Using this triconsistency matrix R, we then look for the smallest threshold t^{tri} such that all connected images in the graph remain connected. All shifts whose triconsistency value is below this threshold are removed. The cleaned shift matrix is denoted $\hat{\tau}^{\text{tri}}$. Figure 3 (c) shows the norm of the difference of this matrix with the ground truth.

Using the same consistency idea, we finally correct the discarded shifts of the connected graph by the composition of valid ones. That is, we estimate the discarded $\hat{\tau}_{n,n'}$ between images u_n and $u_{n'}$ by using a third image u_k as bridge:

$$\hat{\tau}_{n,n'} = \frac{1}{K} \sum_k \hat{\tau}_{n,k} + \hat{\tau}_{k,n'} \tag{9}$$

for all k such that both $\hat{\tau}_{n,k}$ and $\hat{\tau}_{k,n'}$ are valid, K being the number of such valid couples. Figure 3 (d) shows the norm of the difference between this final matrix $\hat{\tau}$ and the ground truth.

3. EVALUATION AND RESULTS

To evaluate the precision attained by the proposed method we needed a ground truth. For that, we generated three synthetic image time series, each composed of 150 images, following

this procedure: first, we computed a high signal-to-noise ratio image by registering, then averaging eleven images between March and August 2020 in the same weather conditions. Then, we applied 150 randomly generated shifts to this image, using spline interpolation with order eleven. These shifts follow a Gaussian distribution with zero mean and a standard deviation of two pixels. We then simulated changes in the weather conditions by applying an affinity to the intensities in the images, *i.e.* the value $u(\mathbf{x})$ is transformed to $a \cdot u(\mathbf{x}) + b$, where the parameters a and b are constant in the image, but different between images. They both follow a uniform distribution in the range $[1, 2]$ for a and $[-50, 50]$ for b. Finally, Gaussian noise was added to the image, with zero mean, and standard deviation 100. In the case of the eight-images long series in Figure 2, we added synthetic opaque clouds.

We present in Figure 1 precision results for these three time series. The registration precision is computed as the root mean squared error (RMSE) of the estimated shifts. The graph shows the evolution of the average error as we increase the number of images. We observe a quick improvement in the registration accuracy between 3 and 50 images, then the curves stabilize. Moreover, the variance between the series also decreases with the number of images, which suggests a more stable result for long time series. Experiments conducted on SAR images showed the same behavior.

4. CONCLUSION

We presented a global registration method for time series of optical and radar images. This method detects and discards unregistrable images (*e.g.* fully covered by clouds), and detects and corrects misaligned images. As a result, the global registration is more precise and reliable. These features are obtained by exploiting the redundancy brought by the multiple images. A graph is constructed with the relative shift of the images; unreliable links are trimmed and replaced thanks to a measure of global consistency. Validation with synthetic data showed that the precision is improved by increasing the number of images in the time series.

An online demo is available at `https://ipolcore.ipol.im/demo/clientApp/demo.html?id=77777000117`.

5. REFERENCES

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

[2] P.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, oct 1970.

[3] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *IEEE International Conference on Cybernetics and Society*, New York, NY, 1975, vol. 6, pp. 163–165.

[4] Tong X. et al., "Image Registration with Fourier-Based Image Correlation: A Comprehensive Review of Developments and Applications," *IEEE JSTAR*, vol. 12, no. 10, pp. 4062–4081, oct 2019.

[5] H.S. Stone et al., "A fast direct Fourier-based algorithm for subpixel registration of images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2235–2243, 2001.

[6] H. Foroosh, J.B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–200, 2002.

[7] Z. Ye et al., "Robust Fine Registration of Multisensor Remote Sensing Images Based on Enhanced Subpixel Phase Correlation," *Sensors*, vol. 20, no. 15, pp. 4338, aug 2020.

[8] 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, jun 2007.

[9] Y. Dong, T. Long, W. Jiao, G. He, and Z. Zhang, "A novel image registration method based on phase correlation using low-rank matrix factorization with mixture of Gaussian," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 446–460, jan 2018.

[10] C. Aguerrebere, M. Delbracio, A. Bartesaghi, and G. Sapiro, "A Practical Guide to Multi-Image Alignment," in *ICASSP*. 2018, vol. 2018-April, pp. 1927–1931, IEEE.

[11] C. Aguerrebere, M. Delbracio, A. Bartesaghi, and G. Sapiro, "Fundamental limits in multi-image alignment," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5707–5722, nov 2016.

[12] M. Rais, C. Thiebaut, J.M. Delvit, and J.M. Morel, "A tight multiframe registration problem with application to Earth observation satellite design," in *2014 IEEE IST Proceedings*. oct 2014, pp. 6–10, IEEE.

[13] V.M. Govindu, "Lie-algebraic averaging for globally consistent motion estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2004, vol. 1, pp. 684–691, IEEE.

[14] S. Farsiu, M. Elad, and P. Milanfar, "Constrained, globally optimal, multi-frame motion estimation," in *IEEE Workshop on Statistical Signal Processing Proceedings*. 2005, vol. 2005, pp. 1396–1401, IEEE.