

1 2 3

Efficient satellite image time series analysis under time warping

François Petitjean[†], *Member, IEEE*, Jonathan Weber[‡]

Abstract—Earth observation satellites are now providing images with short revisit cycle and high spatial resolution. The amount of produced data requires new methods that will give a sound temporal analysis, while being computationally efficient.

Dynamic Time Warping has proved to be a very sound measure to capture similarities in radiometric evolutions. In this letter, we show that its non-linear distortion behavior is compatible with the use of a spatio-temporal segmentation of the data cube that is formed by a satellite image time series (SITS). While dealing with spatial and temporal dimensions of SITS at the same time had already proven to be very challenging, this article proves that taking advantage of the spatial and temporal connectivities, both the performance and the quality of the analysis can be improved.

Our method is assessed on a SITS of 46 Formosat-2 images sensed in 2006, with an average cloud-cover of one third. We show that our approach induces 1) a sharply reduced memory usage, 2) improved classification results and 3) shorter running time.

Index Terms—Satellite Image Time Series, Dynamic Time Warping, spatio-temporal segmentation.

I. INTRODUCTION

STARTING in 2014, the ESA’s SENTINEL program will provide satellite image time series (SITS, for short) with high temporal and spatial resolutions. This program includes five missions – SENTINEL-1 to SENTINEL-5 – that will support the monitoring of lands, oceans and atmosphere. These missions fulfill revisit and coverage requirements to support the general GMES program. SENTINEL-2 will for example provide a general optical cover of the Earth’s surface every five days with 10 m to 60 m resolution and 13 spectral bands.

This data is obviously very valuable for Earth monitoring: ecosystems modeling, water cycles monitoring, agriculture control, etc. The natural counterpart of the value of this data is its complexity. Considering the coarser spatial resolution of SENTINEL-2, *i.e.*, 60 m, and only considering the emerged part of Earth, dozens of billions of

pixels will be acquired every five days. Tackling this huge amount of data is extremely challenging, especially given its temporal nature.

Understanding evolutions that take place in SITS requires to consider the evolution of every (x, y) area with time. Analyzing and classifying this type of data relies on the ability to understand the similarities and differences between these evolution profiles. Dynamic Time Warping (DTW) has proved to be a sound tool for the analysis of SITS [1], [2]. DTW makes it possible to consistently compare radiometric series with different lengths and sampling, which is a decisive property for the operational analysis of large-scale SITS. In particular, it makes it possible to deal with cloud-covered pixels without interpolating the missing values and/or removing cloudy images.

This letter addresses the computational issues of SITS analysis, in order to support the analysis of the data that will be produced by upcoming satellites.

We start by reducing the vocabulary over which the SITS is described. To this end, we apply a spatio-temporal segmentation to simplify the series of image. The SITS is then made of spatio-temporal regions. SITS classification aims deciding on the class of every (x, y) area, in terms of its evolution over the image series. However, the spatio-temporal segmentation induces that every (x, y) area is described by a series of spatio-temporal regions. This greatly complicates the comparison of the evolution of different areas observed over the SITS. To solve this issue, we take advantage of DTW’s ability to optimally “unpack” these series.

Our experiments show that the proposed method achieves three main results. (1) We reduce the memory requirements by more than a factor five. (2) As a result of the filtering that is induced by the segmentation, the quality of the temporal classification is improved. (3) Dealing with shorter series of values for every (x, y) area, the execution time of the DTW-based classification (including the segmentation step) is reduced.

The method is presented in Section II. Experiments carried out on a series of 46 FORMOSAT-2 images are presented in Section III. Section IV concludes this letter and presents some future work.

II. TEMPORAL CLASSIFICATION OF SPATIO-TEMPORALLY SEGMENTED SATELLITE IMAGE TIME SERIES

The aim of our approach is to reduce the computational requirements of SITS analysis, while maintaining its quality. Our approach combines 1) simplifying the SITS with

[†]Dr Francois Petitjean is with the Clayton School of Information Technology at Monash University, Australia (e-mail: francois.petitjean@monash.edu).

[‡]Dr Jonathan Weber is with the Université de Lorraine, LORIA, UMR 7503, Vandœuvre-lès-Nancy, F-54506, France (e-mail: jonathan.weber@univ-lorraine.fr).

©2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works. doi:http://dx.doi.org/10.1109/LGRS.2013.2288358

a segmentation algorithm and 2) using Dynamic Time Warping to “unpack” the series of spatio-temporal regions, in order to make this simplified data analyzable.

1) Simplifying the SITS: We first reduce the vocabulary over which the data is described, in order to sharply reduce the memory consumption for the analysis. To this end, we use a spatio-temporal segmentation, namely *spatio-temporal quasi-flat zones* [3]. We chose this segmentation method because it only slightly reduces the informativity of images, and because it is dedicated to spatio-temporal data. We can then map every pixel of the SITS to the spatio-temporal region it belongs to. This allows us to sharply reduce the amount of memory that is required for the analysis.

2) Comparing series of spatio-temporal regions: The spatio-temporal segmentation is very interesting from the computational perspective, because it makes the most of the redundancy of the data. Dealing with spatio-temporal regions for SITS analysis is however very challenging. Two different (x, y) areas may belong to the same regions at time t_1 , not anymore at t_2 , but again at t_3 . Even more challenging, the region \mathcal{R}_1 they belong to at t_1 can be the same as the one they belong to at t_3 . The length and sampling of two series can moreover be very different. This leads us to the second idea of this letter. We look at the series of spatio-temporal regions in which every geographic area (x, y) successively belongs to. We take advantage of DTW’s ability to optimally “unpack” and assess the similarity of these series.

A. Input

The method takes as input a series $\mathcal{S}_{image} = \langle I^1, \dots, I^N \rangle$ of N ortho-rectified $W \times H$ images. Let $E = \llbracket 1, W \rrbracket \times \llbracket 1, H \rrbracket$ and \mathcal{B} be the number of spectral bands for each image. Each multivalued image I^n ($n \in \llbracket 1, N \rrbracket$) describes function:

$$I^n : \begin{array}{l} E \rightarrow \mathbb{Z}^{\mathcal{B}} \\ (x, y) \mapsto I_1^n(x, y), \dots, I_{\mathcal{B}}^n(x, y) \end{array} \quad (1)$$

B. Segmentation of the SITS

A spatio-temporal segmentation of the SITS \mathcal{S}_{image} is a partition $\mathfrak{S} = \{R_i\}_{i=1}^{\mathcal{R}}$ of $\llbracket 1, N \rrbracket \times E$. Broadly speaking, the cube that is formed by \mathcal{S}_{image} is “decomposed” into \mathcal{R} distinct regions R_i . We associate to every original image I^n of \mathcal{S}_{image} a region image I_R^n :

$$I_R^n : \begin{array}{l} E \rightarrow \llbracket 1, \mathcal{R} \rrbracket \\ (x, y) \mapsto I_R^n(x, y) \end{array} \quad (2)$$

Note that the \mathcal{R} regions are spatio-temporal; as a consequence, they are shared among all the N I_R^n images. In addition, for every spatio-temporal region R , we compute and store the \mathcal{B} average radiometric level. The corresponding map is defined by:

$$A : \begin{array}{l} \llbracket 1, \mathcal{R} \rrbracket \rightarrow \mathbb{Z}^{\mathcal{B}} \\ R \mapsto A(R) \end{array} \quad (3)$$

In this step, any spatio-temporal segmentation algorithm can be used. For remote sensing, the segmentation

step has two main aims. First, it aims at constructing objects of interest for high resolution images, in order to perform an object-based image analysis. Second, it reduces the amount of data to analyze by focusing the analysis on regions. In our case, the use of segmentation is slightly different. The classification will still be performed for every (x, y) area and not for every spatio-temporal region. Our objective is actually to reduce the vocabulary of description of the data.

We thus advise to use a conservative segmentation method that will construct radiometrically homogeneous spatio-temporal regions, while keeping the informativity of the data. We propose to use (ω) -*quasi-flat zones* (also named (ω) -*constrained connectivity*) based segmentation [4]. Quasi-flat zones have been successfully extended to spatio-temporal data [3]. Quasi-flat zones algorithms require to set one parameter ω that sets the maximum difference between two pixels of the same region. Experiments will study the influence of this parameter on the results.

C. Classification of the time series

Let \mathcal{S} be the dataset built from the segmented \mathcal{S}_{image} . For every (x, y) area, \mathcal{S} contains the sequence of regions R_i that the area (x, y) goes through over time. Note that the number of regions can be different from one (x, y) area to another.

The classification of the sensed area over time corresponds to a label image I_C that maps each sensed area (x, y) to a class index $I_C(x, y)$ among the \mathcal{C} possible ones:

$$I_C : \begin{array}{l} E \rightarrow \llbracket 1, \mathcal{C} \rrbracket \\ (x, y) \mapsto I_C(x, y) \end{array} \quad (4)$$

To build such a classification map from a SITS, the (x, y) areas have to be classified in terms of their evolution over time. In the heart of almost any analysis process, there is a tool capturing similarities in data. In our case, every (x, y) area is described by a series of regions and their average radiometric values. “How similar have these two areas evolved?” is the question to which the analysis process often has to answer to. Disposing of a similarity measure makes it possible to answer to this question. Different similarity measures can be of different relevance. *Dynamic Time Warping* (DTW) [5] has proved to be a sound similarity measure for radiometric series [1], [2]. The founding idea was to define a measure that would be able to capture local distortions on the time axis. Simplifying its concept, DTW can be seen as a two steps process that (1) finds the optimal way to align the two compared series and (2) returns how different these aligned series are. DTW can deal with series that are irregularly sampled, because its alignment is optimal. Note that this implies that DTW can compare series of different lengths as well.

DTW warps one sequence into the other so that the total cost of the transformation is optimal. The cost of the optimal alignment of two sequences A and B respectively

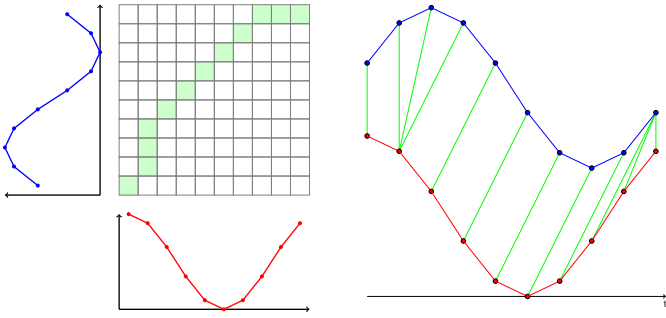


Fig. 1. Two slightly shifted series aligned with DTW. On the left: matrix in which DTW is computed. The warping path is depicted in green. On the right: resulting alignment of the series.

aligned up to the i^{th} element of A and j^{th} is:

$$\text{DTW}(A_i, B_j) = d(a_i, b_j) + \min \begin{cases} \text{DTW}(A_{i-1}, B_{j-1}), \\ \text{DTW}(A_i, B_{j-1}), \\ \text{DTW}(A_{i-1}, B_j) \end{cases}$$

where A_i is the sub-sequence $\langle a_1, \dots, a_i \rangle$. For two sequences of respective lengths $|A|$ and $|B|$, the overall similarity is then given by $\text{DTW}(A_{|A|}, B_{|B|})$.

In our case, as the (x, y) areas are described by sequences of regions, the elements a_i and b_j are regions. To use DTW, we have to be able to compare a_i with b_j . To this end, we define d as Euclidean distance (δ) between the corresponding average intensities of the regions:

$$d(R_i, R_j) = \delta(A(R_i), A(R_j)) \quad (5)$$

In practice, all the regions are numbered between 0 and $\mathcal{R} - 1$ and their multi-spectral average values are stored in a $\mathcal{R} \times \mathcal{B}$ matrix. This makes it possible to access the values in $\Theta(1)$ and thus to keep the same computational complexity.

DTW makes use of dynamic programming to find the optimal alignment in polynomial time [5]. Optimal solution to sub-problems are stored in a matrix, and the optimal alignment is given by the *warping path*, *i.e.*, the path of minimal cost in the matrix. This process is briefly illustrated in Figure 1.

III. EXPERIMENTS

A. Material and method

The area of study for this work is located near the town of Toulouse in the South West of France. 46 FORMOSAT-2 images sensed over one cultural years is used. The temporal distribution of this SITS as well as its cloud covering is given in Figure 2. We use the multi-spectral product at a spatial resolution of 8 m and keep the three bands Near-Infrared, Red and Green.

A number of corrections are applied on the FORMOSAT-2 products before being used in the analysis process. Firstly, images are orthorectified to guarantee that a pixel (x, y) covers the same geographic area throughout the SITS. Secondly, every image is radiometrically corrected from the effects of the sensing conditions, in order to make the pixel values comparable from one image to another. Digital counts are first converted into reflectances

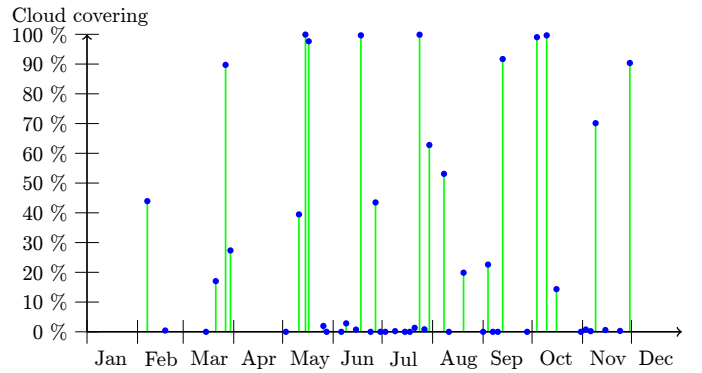


Fig. 2. Temporal distribution and cloud covering of the images from the studied SITS. Each spot represents an acquired image.

(normalized physical quantity of solar irradiance) by using the instrument radiometric model. The monitoring of the FORMOSAT-2 sensor, conducted by the French Space Agency, provides the absolute calibration coefficients. Comparing the measured reflectance in simulations at the top of the atmosphere – carried out for atmospheric and geometric conditions of measurement – makes it possible to process the inversion of the surface reflectance. Simulations carried out for various altitudes – including a weighting of the atmospheric pressure and the amounts of aerosols and water vapor – allows us to take the elevation into account. The condition of the atmosphere at the sensing time is characterized using meteorological sources (NCEP for the pressure and the humidity), using ozone data sources (TOMS or TOAST) and using aerosol data (SEAWIFS, AERONET). Otherwise, climatological values are used.

Figure 3(b) depicts the reference land-cover map that is associated to this dataset [6]. Cloud-contaminated pixels are screened automatically [7] and removed from the initial SITS.

Our experiments aim at studying the influence of the spatio-temporal segmentation step on the classification. Our objective is to demonstrate the interest of our approach in general, and not to highlight the best result only. We want to show that our results are consistent, not only for a particular set of parameters, but for their own sake. We will show how the segmentation parameter (ω) influence the classification in terms of (1) the memory footprint, (2) the quality of the classification and (3) the execution time of the global process (segmentation+classification).

Classification problems are usually addressed using supervised or unsupervised algorithms. Supervised classification algorithms require training examples to learn the classification model. In our case, we want to study the relevance of the proposed data representation. The choice and the suitability of the examples would create a bias, which would make difficult to assess our method. Unsupervised classification allows us to highlight the consistency of the proposed approach, without being influenced by several issues linked to the evaluation of supervised approaches (choice of the algorithm, cross-validation, building and sampling of the training set). We use the standard

K-MEANS algorithm and then automatically label every cluster with regard to the most similar class [1].

In this article, the data objects are time series describing the evolution of the (x, y) geographic areas. To use the K-MEANS algorithm to classify this type of data, it is necessary to provide a way to evaluate the similarity between these time series, as well as a way to construct an average time series from a set of time series. We use DTW to measure how the two optimally aligned series differ from each other (Section II-C). We use DBA [2] to compute an average time series from a set of time series.

We report the memory footprint and execution time of the proposed approach, as well as the quality of the classification. The latter is assessed with the overall accuracy ($\mathcal{O}\mathcal{A}$) and with the F-measure. The overall accuracy gives the percentage correctly classified (x, y) areas over the total number of (x, y) areas. The F-measure corresponds to the harmonic mean between precision (\mathcal{P}) and recall (\mathcal{R}): $\mathcal{F} = 2 \cdot \frac{\mathcal{P} \cdot \mathcal{R}}{\mathcal{P} + \mathcal{R}}$.

B. Results

We present quantitative and qualitative results. The quantitative results are presented in Table I, and compared to the “reference” pixel-based SITS classification (first line of the table).

First, it can be noticed that for almost all parameters from 50 to 180, the quality of the classification improved. We also sharply reduced the memory consumption and optimized the running time. For all cases, we reduce the memory footprint by at least a factor five.

The increased quality of the classification is due to the segmentation step tends to smooth out the radiometric heterogeneity and helps capturing spatial similarities. We obtain the best result in terms of quality for $\omega = 100$. The F-measure and the overall accuracy are both increased by two points, while the memory footprint is decreased by more than a factor seven.

Even if we use low values for ω (e.g., 50 or 60) – which produces an over-segmentation of the SITS – the memory consumption is sharply reduced. This is due to the reduced vocabulary for the description of the data, which is induced by the spatio-temporal segmentation. At the same time, the quality of the classification is improved compared to the results obtained by the standard DTW method. This is a result of the segmentation step which ensures the spatial consistency of the results. The computation time (segmentation + classification) is also reduced by about 20 % for all the parameters of the segmentation.

To qualitatively assess our results, we propose to depict in Figure 3(a) the classification results for the smallest ω . This parameter leads to an over-segmentation of the image series and thus do not make the most of the spatial consistency. What we want to highlight is that, even if no information is available about how to set the segmentation parameters, the classification can be obtained with less computing resources and with very good quality.

We can remark that the classification map is spatially regular, which gives a first positive quality assessment of

ω	\mathcal{F} %	$\mathcal{O}\mathcal{A}$ %	Memory consumption (%)	Run-time in hh:mm
<i>#reference</i>	77.5	77.0	100	1:41
50	78.8	79.0	18.7	1:26
60	79.1	78.9	16.9	1:24
70	78.0	77.4	15.4	1:23
80	77.9	78.0	14.4	1:22
90	79.0	79.0	13.6	1:21
100	79.4	79.4	12.9	1:21
110	78.0	78.2	12.5	1:21
120	79.3	79.6	12.1	1:21
130	78.8	79.3	11.7	1:21
140	78.4	78.8	11.5	1:21
150	78.2	78.8	11.3	1:20
160	78.2	77.9	11.2	1:20
170	78.1	78.4	11.1	1:20
180	77.7	77.1	11.1	1:20
190	77.3	76.4	11.1	1:20
200	76.7	76.0	11.1	1:21

TABLE I
INFLUENCE OF THE SEGMENTATION PARAMETER ON THE PERFORMANCE AND QUALITY OF THE RESULTS. THE *#reference* RESULTS CORRESPOND TO THE PIXEL-BASED RESULTS, WITHOUT THE SEGMENTATION STEP. NOTE THAT THE RUN-TIMES INCLUDE SEGMENTATION AND CLASSIFICATION.

the results. The classification map globally separates the different classes very well. This is the case for corn crop class in orange, for the wheat crop class in yellow or for the hardwood class in dark green. If we compare these results to the ones obtained by the reference pixel-based, we observe a significant improvement F-measure for these different classes: +5 % for corn, +3 % for wheat and +9 % for hardwood.

The grassland class (light green) and the fallow land class (gray) have however been mixed up for a few fields, mainly because they have a very close radiometrical evolution over time (fallow lands often present new growth of grass). Note however that this is also the case in the results for the reference pixel-based approach.

Similarly, the sunflower class (purple) is not completely well separated from the soybean class, because the sunflower temporal behavior is very variable. Sunflower crops can actually grow between the end of the summer crop season and the start of the winter crop season. This makes it a difficult class to separate from the other temporal ones. Note that the F-measure is improved by 14 % for the sunflower class and by 31 % for the soybean class, when compared to the results of the reference pixel-based approach.

IV. CONCLUSION

In this letter, we showed that combining Dynamic Time Warping with a conservative spatio-temporal segmentation achieves three main objectives. First, the segmentation of the data makes it possible to take advantage of the spatio-temporal redundancy, in order to improve the quality of the classification. Second, the reduction of the vocabulary allows us to index the values of the spatio-temporal regions, and thus to sharply reduce the

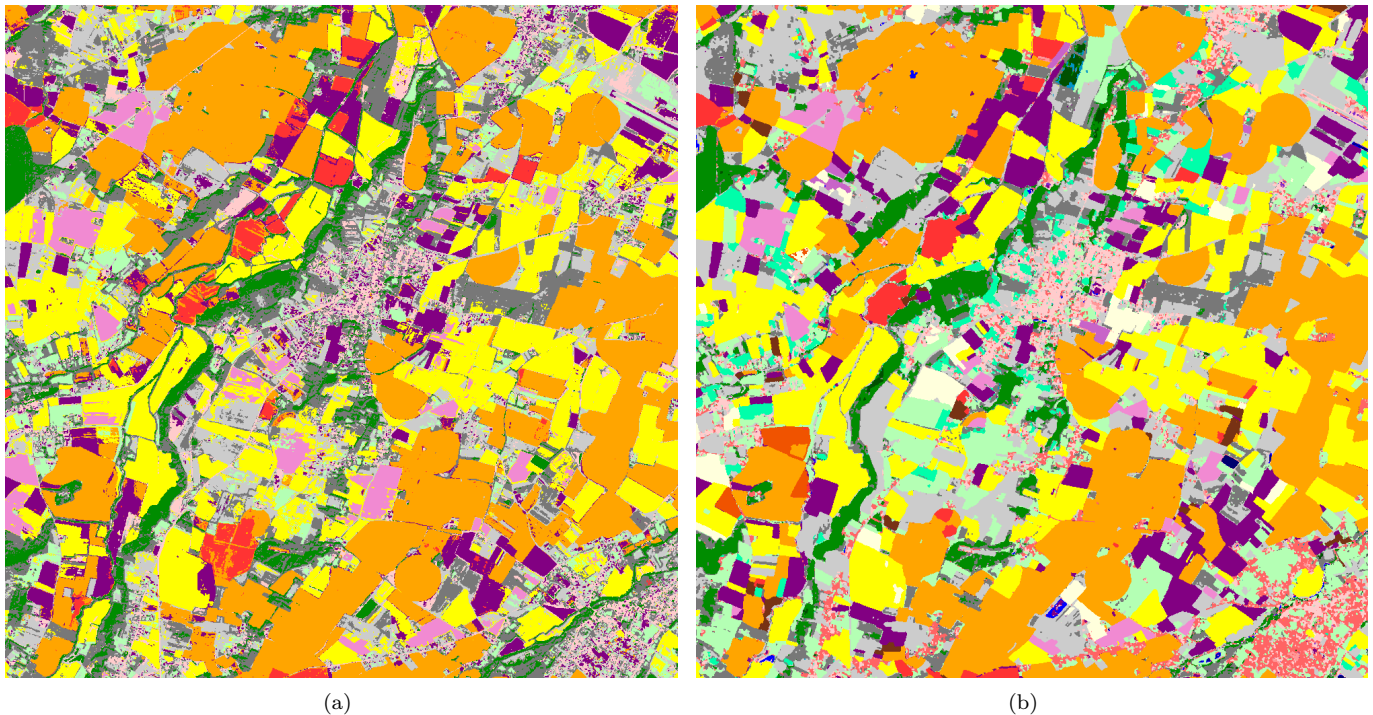


Fig. 3. (a) Classification map obtained with $\omega = 50$. (b) Land-cover reference map.

memory consumption of the analysis. Third, DTW makes the most of the temporal dimension of regions created by the segmentation step, and performs in a noteworthy reduced computation time.

We believe this work opens up a number of research directions. Some work could be carried out in order to improve the segmentation step. Quasi-flat zones tend to generate small regions in transition regions [8]. Filtering techniques as in [9] should be studied for SITS analysis to reduce this phenomenon. Moreover, not only the quality of the regions could be improved, but also the information that is extracted from them. The use of *spatial* characteristics for SITS analysis has recently been introduced [10], [11]. To the best of our knowledge, spatio-temporal features haven't been studied for remote sensing images, and thus raises an important area of investigation.

More generally, the temporal dimension raises numerous issues for remote sensing images analysis. Time induces a sequencing in the description of every (x, y) area, and can thus not be tackled as any other dimension of the data. In addition, observations are *not* continuous over time: there is indeed no information about the (x, y) areas between two images of the series. We believe that studies addressing the impact of time-related assumptions on the quality of SITS results, would greatly benefit the remote sensing community.

ACKNOWLEDGMENT

This research was supported by the Australian Research Council under grant DP120100553. We would like to thank the colleagues from CESBIO (J. Inglada, D. Ducrot, C. Marais-Sicre, O. Hagolle and M. Huc) for providing the land-cover maps and the corrected and cloud-screened FORMOSAT-2 images.

REFERENCES

- [1] F. Petitjean, J. Inglada, and P. Gañçarski, "Satellite Image Time Series Analysis under Time Warping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 8, pp. 3081–3095, 2012.
- [2] F. Petitjean, A. Ketterlin, and P. Gañçarski, "A global averaging method for Dynamic Time Warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.
- [3] J. Weber, S. Lefevre, and P. Gañçarski, "Spatio-temporal quasi-flat zones for morphological video segmentation," in *International Symposium on Mathematical Morphology*, ser. Lecture Notes in Computer Science, vol. 6671. Springer, Jul. 2011, pp. 178–189.
- [4] P. Soille, "Constrained connectivity for hierarchical image partitioning and simplification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1132–1145, Jul. 2008.
- [5] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.
- [6] S. Idraïm, D. Ducrot, D. Mammass, and D. Aboutajdine, "An unsupervised classification using a novel ICM method with constraints for land cover mapping from remote sensing imagery," *International Review on Computers and Software*, vol. 4, no. 2, pp. 165–176, March 2009.
- [7] O. Hagolle, M. Huc, D. V. Pascual, and G. Dedieu, "A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENUS, LANDSAT and SENTINEL-2 images," *Remote Sensing of Environment*, vol. 114, no. 8, pp. 1747–1755, 2010.
- [8] P. Soille and J. Grazzini, "Constrained connectivity and transition regions," in *Mathematical Morphology and Its Application to Signal and Image Processing*, ser. Lecture Notes in Computer Science, M. Wilkinson and J. Roerdink, Eds. Springer Berlin Heidelberg, 2009, vol. 5720, pp. 59–69.
- [9] J. Weber and S. Lefèvre, "Fast quasi-flat zones filtering using area threshold and region merging," *Journal of Visual Communication and Image Representation*, vol. 24, no. 3, pp. 397–409, 2013.
- [10] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 1, pp. 33–37, 2009.

- [11] F. Petitjean, C. Kurtz, N. Passat, and P. Gañçarski, “Spatio-Temporal Reasoning for the Classification of Satellite Image Time Series,” *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1805–1815, 2012.