Spatio-Temporal Reasoning
for the Classification of Satellite Image Time Series

François Petitjean[a,*], Camille Kurtz[a,b], Nicolas Passat[a,b], Pierre Gançarski[a,b]

[a]*LSIIT – UMR 7005, Pôle API, Bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France*

[b]*University of Strasbourg, 7 rue René Descartes, 67084 Strasbourg Cedex, France*

Abstract

Satellite Image Time Series (SITS) analysis is an important domain with various applications in land study. In the coming years, both high temporal and high spatial resolution SITS will become available. In the classical methodologies, SITS are studied by analyzing the radiometric evolution of the pixels with time. When dealing with high spatial resolution images, object-based approaches are generally used in order to exploit the spatial relationships of the data. However, these approaches require a segmentation step to provide contextual information about the pixels. Even if the segmentation of single images is widely studied, its generalization to series of images remains an open-issue. This article aims at providing both temporal and spatial analysis of SITS. We propose first segmenting each image of the series, and then using these segmentations in order to characterize each pixel of the data with a spatial dimension (*i.e.*, with contextual information). Providing spatially characterized pixels, pixel-based temporal analysis can be performed. Experiments carried out with this methodology show the relevance of this approach and the significance of the resulting extracted patterns in the context of the analysis of SITS.

Key words: Multi-temporal Analysis, Satellite Image Time Series, Data Mining, Segmentation, Information Extraction

1. Introduction

Satellite Image Time Series (SITS) constitute a major resource for Earth monitoring. For the last decades, these image series have been either sensed with a high temporal resolution (daily coverage at a kilometer spatial resolution) or with a high spatial resolution (weekly coverage at a meter spatial resolution). However, for a few years, satellites such as the Taiwanese Formosat-2 are providing both high temporal and High Spatial Resolution SITS (HSR SITS), but with a limited coverage of the Earth surface and with only four spectral bands. In the coming years, these kinds of data will become widely available thanks to the ESA's Sentinel program. The growing availability of such images, periodically acquired by satellite sensors on the same geographical area, will make it possible to produce and regularly update accurate temporal land-cover maps of a given investigated site.

In order to efficiently handle the huge amount of data that will be produced by these new sensors, adapted methods for SITS analysis have to be developed. Such methods should allow the end-user to obtain satisfactory results, *e.g.*, relevant and accurate temporal evolution behaviors, with minimal time (by automating the tasks which do not require human expertise), and minimal effort (by reducing the parameters).

In the current standard methods, these data are studied by analyzing the radiometric evolution of the pixels through the time series. The underlying idea is to gather sensed areas that undergo similar radiometric evolutions. This structuring of the data makes it possible to extract both abrupt and long-term changes. In this context, there is actually no difference between a "real" change and a gradual one: both are described by evolution behaviors. In this way, if an sensed area (x, y) undergoes an abrupt change (*e.g.*, a clear cut or the building of a house), it will be treated as a particular temporal behavior, *i.e.*, this behavior will emerge in the classification if it is sufficiently represented in the dataset.

Due to the high spatial resolution of the future images, the geometrical information of the scene could also be considered in the classification process by using object-based approaches. To this end, a segmentation process is required to extract segments based on radiometric homogeneity. Once these segments are extracted, it is possible to characterize them using spatial/geometrical properties, to enhance the classification process. However, the integration of a segmentation step in a temporal classification framework remains an open-issue, since neither the mapping between mono-temporal segmentations, nor the temporal segmentation are resolved. A review of the available literature on SITS analysis shows a lack of existing methods responding to this need. This article aims at addressing this issue by characterizing a pixel with spatial properties in order to improve the analysis of SITS.

---

*Corresponding author – LSIIT, Pôle API, Bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France – Tel.: +33 3 68 85 45 78 – Fax.: +33 3 68 85 44 55

Email addresses: `fpetitjean@unistra.fr` (François Petitjean), `ckurtz@unistra.fr` (Camille Kurtz), `passat@unistra.fr` (Nicolas Passat), `gancarski@unistra.fr` (Pierre Gançarski)

This article is organized as follows. Section 2 gives an overview of existing methods for SITS analysis. Section 3 introduces our generic methodology for spatio-temporal analysis of SITS. Section 4 describes the experimental validation carried out with this methodology. Section 5 presents the results obtained using the proposed methodology. Conclusions and perspectives will be found in Section 6.

2. State of the art

SITS allow the analysis, through observations of land phenomena, of a broad range of applications such as the study of land-cover or the mapping of damage following a natural disaster. These changes may be of different types, origins and durations. For a detailed survey of these methods, the reader can refer to (Coppin et al., 2004; Lu et al., 2004).

In the literature, we find three main families of methods. *Bi-temporal analysis*, *i.e.*, the study of transitions, can locate and study abrupt changes occurring between two observations. Bi-temporal methods include image differencing (Bruzzone & Prieto, 2000), image ratioing (Jensen, 1981; Wu et al., 2005), image composition (Ouma et al., 2008) or change vector analysis (CVA) (Johnson & Kasischke, 1998; Bovolo, 2009; Bahirat et al., 2012). A second family of *mixed methods*, mainly statistical methods, applies to two or more images. They include methods such as post-classification comparison (Foody, 2001), linear data transformation (PCA and MAF) (Howarth et al., 2006), image regression or interpolation (Kennedy et al., 2007) and frequency analysis (*e.g.*, Fourier, wavelets) (Andres et al., 1994; Tsai & Chiu, 2008). Then, we find methods dedicated to image time series and based on *radiometric trajectory analysis* (Jönsson & Eklundh, 2004; Verbesselt et al., 2010; Petitjean et al., 2011c; Kennedy et al., 2010; Lui & Cai, 2011).

Regardless of methods used in order to analyze satellite image time series, there is a gap between the amount of data composing these time series, and the ability of algorithms to analyze them. Firstly, these algorithms are often dedicated to the study of a change in a scene from bi-temporal representation. Secondly, and this point is even more difficult to deal with, the geometrical/spatial properties of the data are rarely taken into account, except for the use of the pixel coordinates. Finally, High Spatial Resolution SITS have given rise to the need for spatially and temporally dedicated methods.

To improve the analyzing process by using the spatial relationships of the data, object-based methods have been recently proposed (Blaschke, 2010). In a first step, the images are segmented/partitioned into sets of connected regions. Then for each region, geometric features (Carleer & Wolff, 2006) (*e.g.*, area, elongation, smoothness) or even contextual ones (Gaetano et al., 2009; Bruzzone & Carlin, 2006; Kurtz et al., 2010) (*e.g.*, spatial context, multi-scale/multi-resolution attributes) are computed in order to characterize the regions. Finally, the regions are classified using these features (Herold et al., 2003).

Object-based methods have shown promising results in the context of single-image analysis. However, their extension/adaptation to SITS in order to exploit both the spatial and temporal information contained in these data remains an open-issue. Indeed, although several methods have been proposed in order to map segments from one image to another (Gueguen et al., 2006; Bovolo, 2009), to directly build spatio-temporal segments (Fan et al., 1996; Moscheni et al., 1998; Tseng et al., 2009), or even to consider object-based features (Hall & Hay, 2003; Niemeyer et al., 2008; Hofmann et al., 2008; Schopfer et al., 2008; Tiede et al., 2011), their scalability to wide sensed areas and their robustness to local disturbance (temporally and spatially) remain problematic. The use of 3D-dedicated methods indeed requires a high temporal continuity; this constraint is however rarely fulfilled by SITS, where the average time-delay between two images is usually too high. As a consequence, the temporal continuity of the observed phenomena can not be assumed between samples. In addition, the irregular temporal sampling of the image series (due to operational constraints of remote sensing), would create a disparity of the spatio-temporal regions in terms of their informativity. For instance, a region spreading over four months should not have the same importance in the analysis, than a region spreading over a single sample (*i.e.*, built over a single image). Thus, this article focuses on mono-temporal spatial enrichment of the pixels, in order to loosen the constraint on the pseudo-continuity.

We therefore suggest to classify SITS as the radiometric evolution of sensed areas with time. Then, in order to take into account the spatial properties of the data, we propose to characterize each pixel with spatial and geometrical attributes obtained using a pre-segmentation step. This formulation allows the study of spatial characteristics over time while abstracting from the correspondence between segments since the data remains the pixel. Moreover, this formulation is aimed at obtaining accurate and reliable evolution behavior maps both by preserving the geometrical details in the images and by properly considering the spatial context information.

This paradigm, using spatially characterized pixels, was previously introduced and studied for contextual analysis (Melgani & Serpico, 2002), multi-level segmentation of a single image (Bruzzone & Carlin, 2006) and change detection in bi-temporal images (Bovolo, 2009). In all these application domains, such a paradigm has shown promising results. We propose, in this article, to extend it to the analysis of large SITS.

3. Spatio-temporal analysis methodology

In this section, we present the proposed approach, which is composed of five main steps that are sequentially applied:

A. Segmentation of the images;
B. Characterization of the regions;
C. Construction of the vector images;
D. Construction of the time series;
E. Classification of the time series.

These steps are fully described in the remainder of this section. The reader may also refer to Figure 1 for a visual outline of the workflow of the proposed approach. Let us first establish the terminology used in the remainder of this article. The
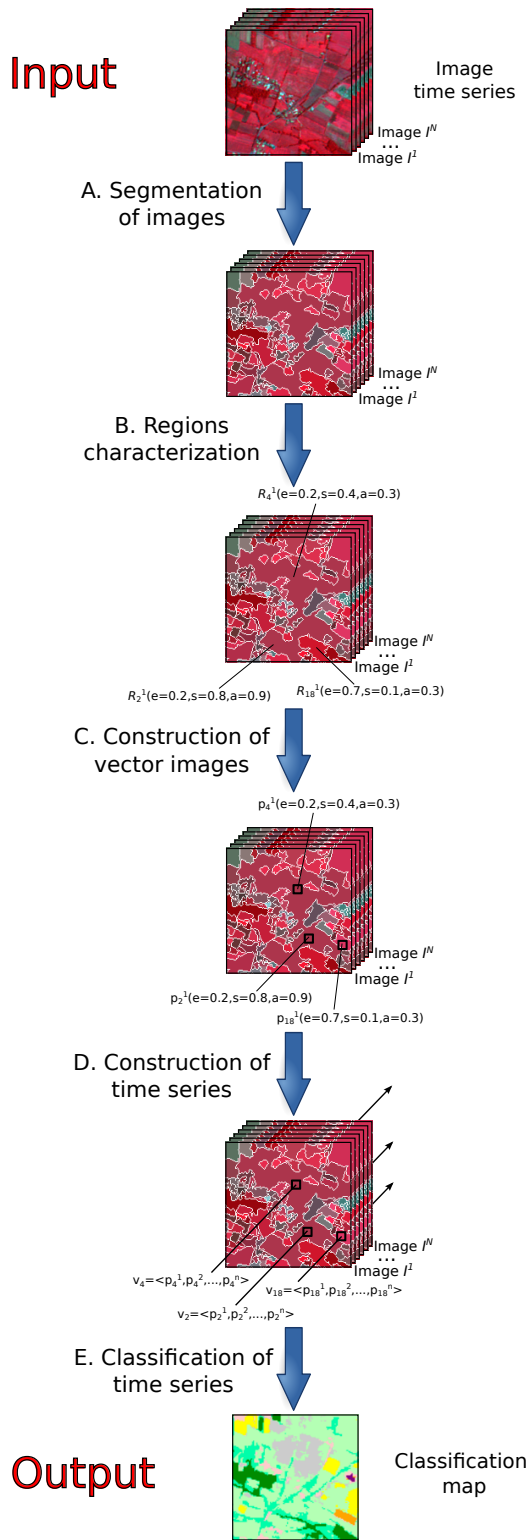
Input

Image time series
Image I^N
...
Image I^1

A. Segmentation of images

Image I^N
...
Image I^1

B. Regions characterization

$R_4^1(e=0.2, s=0.4, a=0.3)$

Image I^N
...
Image I^1

$R_2^1(e=0.2, s=0.8, a=0.9)$ $R_{18}^1(e=0.7, s=0.1, a=0.3)$

C. Construction of vector images

$p_4^1(e=0.2, s=0.4, a=0.3)$

Image I^N
...
Image I^1

$p_2^1(e=0.2, s=0.8, a=0.9)$
$p_{18}^1(e=0.7, s=0.1, a=0.3)$

D. Construction of time series

Image I^N
...
Image I^1

$v_4=<p_4^1, p_4^2, ..., p_4^n>$ $v_{18}=<p_{18}^1, p_{18}^2, ..., p_{18}^n>$
$v_2=<p_2^1, p_2^2, ..., p_2^n>$

E. Classification of time series

Output

Classification map

Figure 1: Workflow of the proposed approach that takes as input N images and provides as output a classification map of the sensed scene. Different symbols have been used to exemplify the steps of the proposed approach (R refers to regions, p refers to pixels, v refers to vectors of pixels built through time). For instance, the symbol $R_4^1(e = 0.2, s = 0.4, a = 0.3)$ means that the region R_4^1 is characterized by 3 feature values (e = elongation, s = smoothness, a = area).

term "sensed area" will be rather used than the term "pixel", since the notion of "pixel" usually refers to a mono-temporal value, while we use in this work the (x, y) coordinates to locate a geographic area. Contrary to the mono-image case, these notions are not mixed up in the temporal case. Consequently, the term "sensed area" will be used to designate the evolution of the (x, y) geographic area with time, while the term "pixel" will be used to designate a sensed value in a particular image.

Input/Output

Let us briefly define the input and the output of the proposed method.

Input. The method takes as input a series $\mathcal{S}_{image} = \langle I^1, \ldots, I^N \rangle$ of N ortho-rectified images of width W and height H. Let $E = [\![1, W]\!] \times [\![1, H]\!]$ where $[\![a, b]\!]$ denote the interval on $\mathbb{Z}$, bounded by a, b. The set E corresponds to the discretization of the continuous space (*i.e.*, the part of $\mathbb{R}^2$) which will be visualized in the images. Let $\mathcal{B}$ be the number of bands of the images composing the series. Each multivalued (*i.e.*, with multiple bands) image I^n ($n \in [\![1, N]\!]$) can be seen as a function:

$$I^n : \quad E \quad \rightarrow \quad \mathbb{Z}^{\mathcal{B}}$$
$$(x, y) \quad \mapsto \quad I_1^n(x, y), \ \cdots, \ I_{\mathcal{B}}^n(x, y) \tag{1}$$

Note that the radiometric levels of the images do not have to be comparable from one image to another. Thus, images can be acquired by different sensors but must be of the same spatial resolution.

Output. The method provides as output a classification of the sensed scene, where areas that have evolved in a similar way are clustered. Such classification can be modeled by a label image $I_C : E \rightarrow [\![1, C]\!]$, which associates to each sensed area (x, y) a class value $I_C(x, y)$ among the C possible ones.

Each class of the classification is also modeled by a centroid sequence, which provides a concise representation of the underlying evolution behavior. This extra information is however not studied in this article.

3.1. Segmentation of images

A segmentation of a multivalued image I^n is a partition $\mathfrak{S}^n = \{R_i^n\}_{i=1}^{\mathcal{R}^n}$ of $[\![1, W]\!] \times [\![1, H]\!]$; broadly speaking, the scene visualized in I^n is "decomposed" into $\mathcal{R}^n$ distinct parts R_i^n, which are supposed to present specific radiometric properties. We will denote R_i^n as a *region* of the image I^n. To any segmented image I^n, we then associate a region image

$$I_R^n : \quad E \quad \rightarrow \quad [\![1, \mathcal{R}^n]\!]$$
$$(x, y) \quad \mapsto \quad I_R^n(x, y) \tag{2}$$

Such region image is a function that associates to each sensed area (x, y) a region label $I_R^n(x, y)$ among the $\mathcal{R}^n$ possible ones.

Once the N images have been segmented (producing N region images $I_R^n, (n \in [\![1, N]\!])$), it is then possible to characterize each region of each segmentation by following the next step.

3.2. Regions characterization

Numerous features (spectral, geometrical, topological, *etc.*) can be computed for the regions of a segmentation in order to characterize them. Each feature can be seen as a function F associating to each region R_i^n ($i \in [\![1, \mathcal{R}^n]\!]$) of a segmentation $\mathfrak{S}^n$ a corresponding feature value $F(R_i^n) \in \mathbb{R}^\alpha$. Although the classical case corresponds to mono-dimensional features in $\mathbb{R}$, certain features can be seen as multi-dimensional ones in $\mathbb{R}^\alpha$ (*e.g.*, correlated textural features, multi-scale features).

$$F : \begin{array}{ccc} [\![1, \mathcal{R}]\!] & \to & \mathbb{R}^\alpha \\ R_i^n & \mapsto & F(R_i^n) \end{array} \tag{3}$$

Once a region is characterized by a (multidimensional) feature value, it is then possible to affect this value to all the pixels composing the region. Let $\mathcal{Z}$ be the number of region-features chosen to describe every sensed area (x, y) of every image.

3.3. Construction of vector images

At this step, each pixel of a multivalued image I^n can be characterized by two types of information:

- directly sensed values (*i.e.*, $\mathcal{B}$ values, denoted I_b^n with $b \in [\![1, \mathcal{B}]\!]$;

- region-associated values (*i.e.*, $\mathcal{Z}$ values, denoted F_a with $a \in [\![1, \mathcal{Z}]\!]$).

All these values are normalized over the image time series by using the extrema values of the attributes in the dataset. It is then possible to combine these features to build "enriched" pixels in order to better characterize them. To process, a vector of features is created and associated to each one of the pixels contained in the image I^n. Finally, by applying this step to each image of the series, we build N vector images defined as:

$$V^n : \begin{array}{ccc} E & \to & [0, 1]^{\mathcal{B}+\mathcal{Z}} \\ (x, y) & \mapsto & \prod_{b=1}^{\mathcal{B}} I_b^n(x, y) \times \prod_{a=1}^{\mathcal{Z}} F_a(I_R^n(x, y)) \end{array} \tag{4}$$

3.4. Construction of time series

Let $\mathcal{S}$ be the dataset built from the image time series. $\mathcal{S}$ is the set of sequences defined as:

$$\mathcal{S} = \left\{ \langle V^1(x, y), \cdots, V^N(x, y) \rangle \mid x \in [\![1, W]\!], y \in [\![1, H]\!] \right\} \tag{5}$$

In these sequences, each element is $(\mathcal{B}+\mathcal{Z})$-dimensional. Since high-dimensional spaces do not often provide the best solutions, we will study different subspaces of this $(\mathcal{B} + \mathcal{Z})$-dimensional space in the experiment part (*e.g.*, time series where each pixel is characterized by a 5-tuple composed of three directly sensed values and two region-associated values).

3.5. Classification of the time series

The extraction of relevant temporal behaviors from satellite image time series can be realized using a classification algorithm. Once these time series have been built, it becomes possible to classify them into different clusters/classes of interest.

To this end, the proposed methodology makes it possible to use either supervised or unsupervised classification algorithms.

A classification of a set of sequences $\mathcal{S}$ is a partition $\mathfrak{C} = \{C_i\}_{i=1}^C$ of E; broadly speaking, as each temporal sequence is associated to a sensed area (x, y), the whole scene can be "decomposed" into C distinct parts C_i, which are supposed to represent similar temporal evolution behaviors. We will denote C_i as a *cluster/class*. The classification can be modeled by a label image $I_C : E \to [\![1, C]\!]$, which associates to each sensed area (x, y) a class value $I_C(x, y)$ among the C possible ones

$$I_C : \begin{array}{ccc} E & \to & [\![1, C]\!] \\ (x, y) & \mapsto & I_C(x, y) \end{array} \tag{6}$$

4. Material and experimental settings

To assess the relevance of the proposed generic spatio-temporal analysis methodology, we have applied it to the classification of agronomical areas. Starting from the different issues raised by this applicative context, we show in this section how the proposed methodology can be used as a potential solution to address them.

4.1. Applicative context: Crop monitoring

The analysis of agronomical areas is important for the monitoring of physical variables, in order to give information to the experts about pollution, vegetation health, crop rotation, *etc.* This monitoring is usually achieved through remote sensing. Indeed, by using classification processes, satellite image time series actually provide an efficient way to monitor the evolution of the Earth's surface. Moreover, when the classes of interest are temporal (*e.g.*, wheat crop, maize crop), the time dimension of the data has to be taken into account by the classification algorithms. For instance, the reflectance levels of the maize crop and of the wheat crop are very similar while their temporal behaviors are quite different (*i.e.*, wheat grows earlier in the year than maize).

Thus, the usual strategy for land-cover mapping consists of classifying the temporal radiometric profiles of the sensed areas (x, y). With the arrival of SITS with high spatial resolution (HSR), it becomes necessary to use the spatial information held in these series, in order to either study the evolution of spatial features, or to help characterizing the different land-cover classes. Our experiments focus on the second point. The underlying idea is that several spatially-built features can be used in the classification process. For example, some crops are usually cultivated in smaller parcels than others, while having the same radiometric temporal behavior (*e.g.*, sunflower crop *vs.* wheat crop). Another (non restrictive) example could be the use of the smoothness of the regions, which could help, for instance, to distinguish between tree-crop and forest.

In the remainder of this section, we experimentally demonstrate how the proposed generic spatio-temporal analysis methodology can be instantiated to enable the use of such spatially-built features.
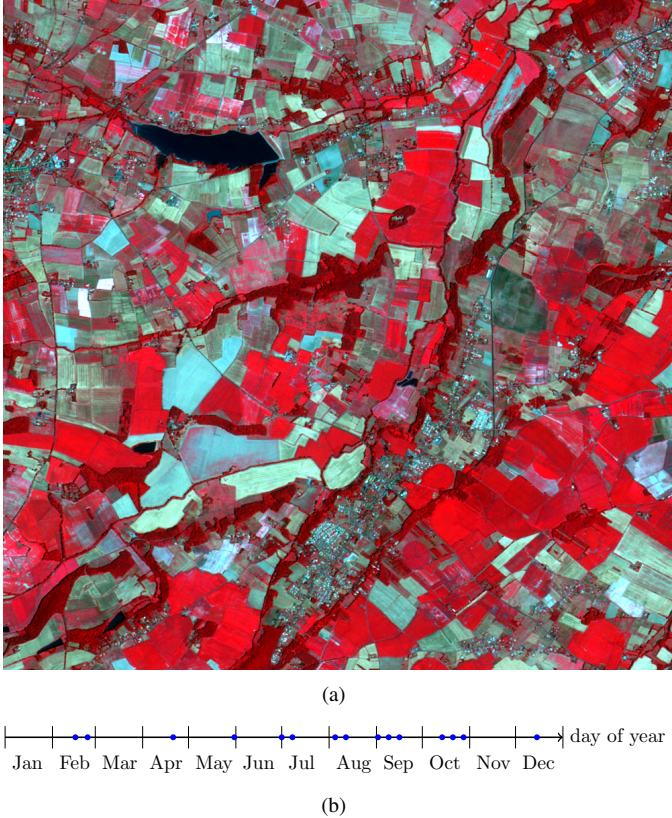
(a)

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

(b)

Figure 2: Presentation of the dataset. (a) One image from the series (August, 4th 2007). (b) Sensing distribution of images sensed over 2007. Each spot represents a sensed image.

4.2. Dataset description

We detail hereafter the main information concerning the images used for this work. The area of study is located near the town of Toulouse in the South West of France. 15 cloud-free FORMOSAT-2 images sensed over the 2007 cultural year are analyzed. These images cover an area of 64 km^2 ($1,000 \times 1,000$ pixels). One image of the series is given in Figure 2(a) while the temporal distribution of the sensed images is given in Figure 2(b).

From these images, we use the multi-spectral product at a spatial resolution of 8 m with the four bands Near-Infrared, Red, Green and Blue. Before being used in this work, the FORMOSAT-2 products have been ortho-rectified (guaranteeing that a pixel (x, y) covers the same geographic area throughout the image series). All images also undergo processes in order to make the radiometric pixel values comparable from one image to another. These processes consist of converting the digital counts provided by the sensor into a physical magnitude and in restoring their own contribution to the surface by correcting for atmospheric effects. This procedure is detailed in (Hagolle et al., 2010).

4.3. Experimental settings

This section aims at describing how the proposed generic approach has been instantiated to deal with the presented crop

monitoring issue. However, we recall that the presented approach is not limited to this instantiation. The five steps described in Section 3 have been performed as follows.

Segmentation of images. The segmentation of HSR satellite images is not a trivial task since the different objects of interest (and thematic ground areas) which are sensed by these images, cannot be necessarily segmented at the same scale (*i.e.*, scale issue). For instance, the main environments, such as urban areas, rural zones, or forests, can be identified at coarsest scales, while more detailed structures, such as buildings and roads, will emerge at the finest ones (Blaschke, 2010). It is then difficult to correctly segment all these thematic ground areas by using only one segmentation result.

For the last decade, it has been shown that hierarchical segmentation algorithms provide accurate results adapted to process HSR images (Pesaresi & Benediktsson, 2001; Gaetano et al., 2009). In particular, their combinations can provide an efficient way to deal with the scale issue (Akcay & Aksoy, 2008; Kurtz et al., 2011a,b). However, the parameters of such algorithms have to be tuned according to the characteristics of the image modality (used as input) and the features of the objects to be segmented. To avoid this parametrization problem (which falls outside the scope of this article), we have chosen to use the Mean-Shift segmentation algorithm (Comaniciu & Meer, 2002) to segment each image of the series. Indeed, this algorithm is intuitive to configure and has shown satisfactory results in the context of the segmentation of remote sensing images (Huang & Zhang, 2008). Although we know that considering a single segmentation map for each image is, in most of the cases, a sub-optimal approach (since the spatial arrangement of the objects in the image is intrinsically hierarchical), we assume that when dealing with agricultural territories, the fields could be efficiently extracted at similar scales and thus, by using only one segmentation map per image. We plan to address this aspect in a future development of the work as stated in the conclusions.

The Mean-Shift segmentation algorithm performs as follows. For a given pixel, this algorithm builds a set of neighboring pixels within a given spatial radius and color range. The spatial and color center of this set is then computed and the algorithm iterates with this new spatial and color center. There are three main parameters: the spatial radius (denoted by h_s) used for defining the neighborhood, the range radius (denoted by h_r) used for defining the interval in the color space and the minimum size M for the regions to be kept after segmentation. We have used the OTB implementation of the Mean-Shift algorithm. ORFEO Toolbox (OTB) is an open source library of image processing algorithms developed by the French Space Agency (CNES). `http://www.orfeo-toolbox.org`

In order to assess the robustness of the proposed approach with regard to the segmentation step, the influence of the segmentation parameters has been studied. Since the level of geometrical information extracted by the segmentation algorithm depends on its parametrization, we have run the algorithm using different configurations of the parameters. In practice, the minimum size M of the regions has been fixed to $M = 25$, corresponding to the minimum expected size of the studied objects of

5

interest. The ranges of the possible values for the other parameters h_s and h_r have been scanned exhaustively with a quite low step ($h_s \in \{1, 3, 5, \cdots, 28, 30\}$ and $h_r \in \{1, 5, 10, \cdots, 55, 60\}$).

The underlying idea of this experiment is to study the influence of the segmentation parameters on the classification results. To this end, only the radiometric mean of the regions has been used to characterize the pixels for the classification (during the construction of the time series).

Region characterization. Several characteristics can be useful for the classification of agronomical scene. For instance, the size of the regions could be used to discriminate small/large fields, while the smoothness could be used to separate forest regions from fields. In this way, the following region-associated features have been computed:

- the mean of the infra-red band of the region (F_{NIR});
- the mean of the red band of the region (F_R);
- the mean of the green band of the region (F_G);
- the mean of the blue band of the region (F_B);
- the area of the region (F_{Area});
- the elongation of the region ($F_{Elong.}$);
- the smoothness of the region ($F_{Smooth.}$);
- the compactness of the region ($F_{Comp.}$).

The elongation is computed as the highest ratio between the width and the length of several bounding boxes (computed for different directions, *i.e.*, each $\pi/8$). The smoothness is computed as the ratio between the perimeter of the morphologically opened region and the original region. To this end, we use a square-shaped opening structuring element invariant to the scale (*i.e.*, with a size depending on the area of the original region). The size of the structuring element was set to $\sqrt{F_{Area}}$. The compactness is computed as the square root of the area of the region multiplied by the length of the perimeter of the region.

Construction of vector images. As explained previously, each pixel composing the SITS can be characterized by two types of information: directly sensed values (denoted as I_{NIR}, I_R, I_G, I_B), and region-associated values (denoted as $F_{NIR}, \ldots, F_{Smooth.}$). All the values are normalized in $[0, 1]$, attribute by attribute over the series. This allows each attribute to be of comparable weight for the classification step.

Construction of time series. In order to find the best separation of thematic classes and to assess (globally and independently) the interest of the different contextual attributes, we tested several combinations of twelve attributes over the time series. All the possible combinations of the spatial attributes ($F_{Area}, F_{Elong.}, F_{Smooth.}, F_{Comp.}$) were tested with either the pixel radiometric values (I_{NIR}, I_R, I_G, I_B), or the mean region ones (F_{NIR}, F_R, F_G, F_B). The resulting 32 combinations are presented in Table 1. In particular, the combination $\star$ (only the pixel radiometric values without any region-associated feature) represents the "classical" combination for pixel-based classification of SITS.

Classification of time series. Classification problems are usually addressed using supervised or unsupervised algorithms. Supervised classification algorithms require training examples to learn the classification model. In our case, as we want to demonstrate the relevance of the proposed data representation, the choice and the suitability of the examples would create a bias, which would make difficult to identify the benefits provided by the spatial features. In this way, choosing an unsupervised classification step 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, *etc.*). We have then applied the classical K-MEANS clustering algorithm (MacQueen, 1967) to classify the time series previously constructed. The distance used to compare the time series of S is the Euclidean distance. Note that other distances (and more relevant temporal ones (Petitjean et al., 2011b,a, 2012)) could also be used.

The K-MEANS algorithm has been used with as many classes (see Table 2) as in the reference map (*i.e.*, 25 seeds), and with 15 iterations; the process has generally converged afterwards (Bottou & Bengio, 1995). Note that any clustering algorithm dealing with numerical data could also be used.

4.4. Validation

To assess the quality and the accuracy of the results, the classification maps obtained have been compared to:

1. a field survey (*i.e.*, ground-truth) of the 2007 cultural year (produced by the European Environment Agency; see http://ec.europa.eu/agriculture/index_en.htm for more details about the Common Agricultural Policy.) covering a partial part of the studied area (Figure 3(a));
2. a land cover reference map (produced by a supervised classification method described in (Idbraim et al., 2009)) covering the totality of the studied area (Figure 3(b)).

Note that these two reference maps reflect the temporal behavior of the considered crops over the 2007 cultural year and do not reflect a static land cover state (*i.e.*, representing a single snapshot of the scene at a particular date). Such property is necessary since we want to assess the accuracy of temporal classification results. We also want to underline that, through the year, the land cover types do not change (*i.e.*, no crop rotation). This fact justifies why the considered classes are designated by static terms (*e.g.*, corn, wheat, meadow) instead of being described by dynamic ones (*e.g.*, the class " bare soil $\rightarrow$ growth of corn $\rightarrow$ harvest").

The classification maps obtained have been compared to these maps using several evaluation indexes. To assess the global accuracy of the obtained classification results, we have computed respectively:

- the average F-measure $\overline{\mathcal{F}}$;
- the Kappa index $\mathcal{K}$;
- the overall classification accuracy $\mathcal{A}$.
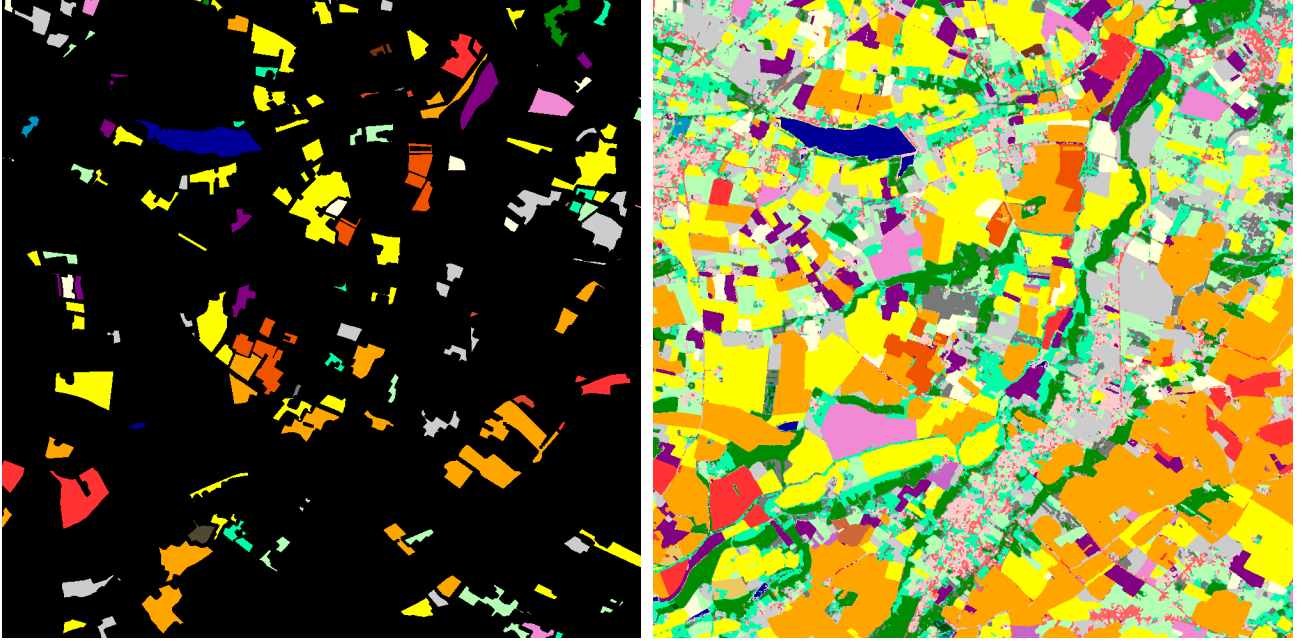
(a)   (b)

Figure 3: Land cover reference maps of the 2007 cultural year. (a) Ground truth (covering a partial part of the studied area) related to a field survey produced by the European Common Agricultural Agency. (b) Land cover reference map (covering the totality of the studied area) produced by the method described in (Idbraim et al., 2009).

The average F-measure $\overline{\mathcal{F}}$ corresponds to the mean, for each class, of the F-measures obtained. To this end, for each thematic class, the best corresponding clusters (in terms of partitions) were extracted. Then, we have computed: the percentage of false positives (denoted by $f^{(p)}$), the percentage of false negatives (denoted by $f^{(n)}$) and the percentage of true positives (denoted by $t^{(p)}$). These measures are used to estimate the precision $\mathcal{P}$ and the recall $\mathcal{R}$ of the results obtained by using the proposed method:

$$\mathcal{P} = \frac{t^{(p)}}{t^{(p)} + f^{(p)}} \quad \text{and} \quad \mathcal{R} = \frac{t^{(p)}}{t^{(p)} + f^{(n)}} \quad (7)$$

For each experiment, we have then computed the geometrical mean $\overline{\mathcal{P}}$ of the precisions obtained and the geometrical mean $\overline{\mathcal{R}}$ of the recalls obtained. Finally, we have computed the mean F-measure $\overline{\mathcal{F}}$ which is the harmonic mean of the mean precision and the mean recall:

$$\overline{\mathcal{F}} = 2 \cdot \frac{\overline{\mathcal{P}} \cdot \overline{\mathcal{R}}}{\overline{\mathcal{P}} + \overline{\mathcal{R}}} \quad (8)$$

The computation of these class-specific indexes requires the matching of classes of interest with clusters extracted by the unsupervised classification approach. To this end, we have used an automatic strategy, which consists of selecting the clusters that maximize the overlapping with the corresponding class.

To assess the global relevance of the results, we have also computed the Kappa index (Congalton, 1991) $\mathcal{K}$, which is a measure of global classification accuracy:

$$\mathcal{K} = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (9)$$

where $\Pr(a)$ is the relative agreement among the observers, and $\Pr(e)$ is the hypothetical probability of chance agreement. The Kappa index takes values in [0, 1] and decreases as the classification is in disagreement with the ground-truth map. Note that the Kappa index is an agreement measure between two partitions and thus does not require to "align" the clusters with the reference classes.

To assess separately the accuracy of each thematic class, we also provide (for each one of these classes) the precision $\mathcal{P}$, the recall $\mathcal{R}$ and their averages.

5. Results

This section presents the results obtained with the proposed contextual approach in the context of the multi-temporal analysis of agronomical areas. The first sub-section describes the study of the influence of the segmentation step on the obtained classification results. The second sub-section proposes an exhaustive analysis of the interest of the different contextual attributes for multi-temporal analysis. Finally, the third sub-section presents an experimental study about the time complexity.

5.1. Influence of the segmentation step

The graph represented in Figure 4 summarizes the accuracy scores (mean F-measure $\overline{\mathcal{F}}$ values) of the classification results obtained as a function of the parameters of the segmentation algorithm (the spatial radius h_s and the range radius h_r). For each series of resulting segmentations, the classification is obtained by using the radiometric mean of the regions to characterize the

7

| (a) | (b) | (c) |
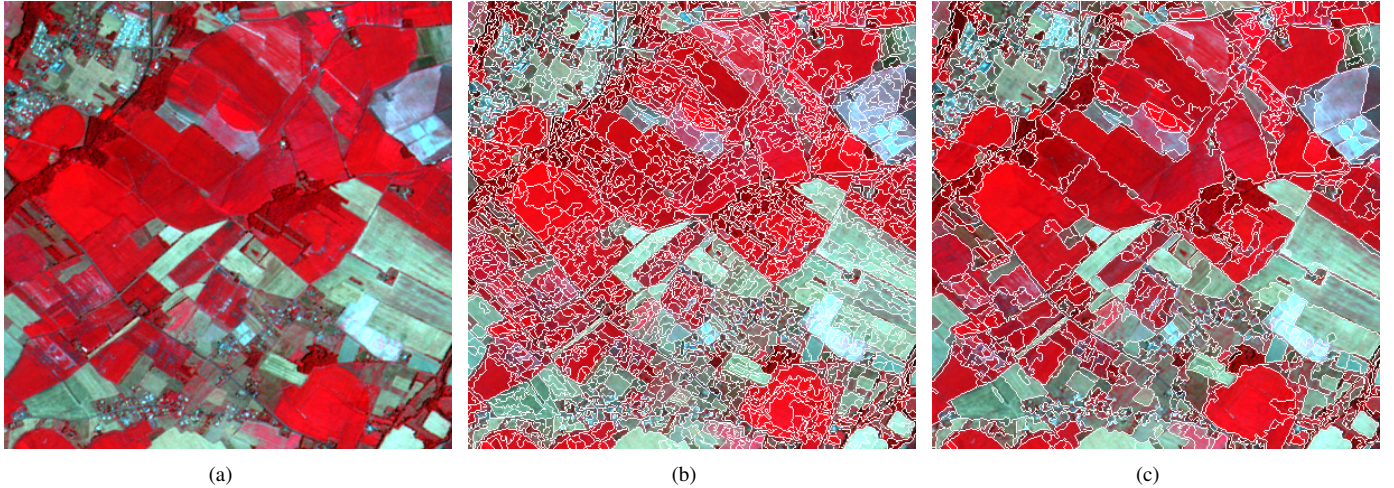
Figure 5: Results of the Mean-Shift segmentation algorithm carried out on an extract of the satellite image presented in Figure 2(a). (a) Extract of one image of the series. (b) Segmentation result ($h_s = 3, h_r = 3, M = 25$). (c) Segmentation result ($h_s = 10, h_r = 15, M = 25$). Note that the boundaries of the regions are depicted in white.
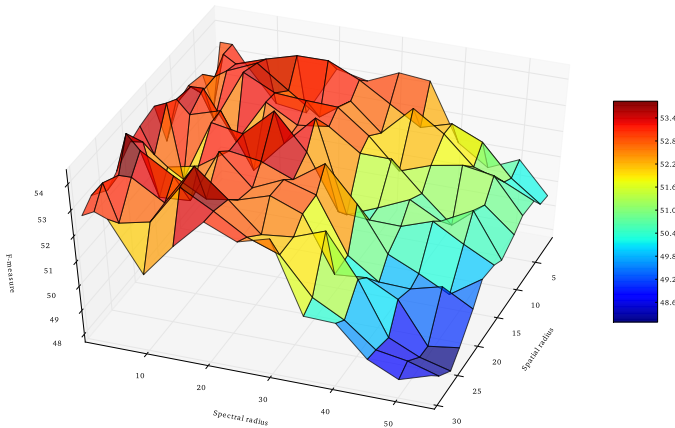
Figure 4: Influence of the parameters of segmentation (the spatial radius h_s and the range radius h_r) on the classification results. The accuracy of the classification results obtained is assessed using the mean F-measure $\overline{\mathcal{F}}$ computed with the land-cover reference map presented in Figure 3(b).

pixels through time. The best scores obtained (and thus, the best configurations of parameters) correspond to the orange-red area while the worst ones correspond to the green-blue area. The large size of the orange-red area allows us to assess the robustness of the approach to the segmentation step. As reference, the base-experiment $\star$ obtained a F-measure score of 51.3% (corresponding to the green area), which is lower than most of the scores obtained by using the radiometric mean of the regions. For the remainder of this experimental study, the configuration ($h_s = 10, h_r = 15, M = 25$) has then been kept.

In order to visually confirm the choice of this configuration, Figure 5 illustrates different segmentation results obtained on an extract of an image. One can note that the parameters configuration ($h_s = 3, h_r = 3, M = 25$) provides over-segmented results while the parameters configuration ($h_s = 10, h_r = 15, M = 25$) provides satisfactory results for the extraction of agricultural areas.

5.2. Results analysis

Table 1 summarizes the F-measure values, the Kappa values and the overall classification accuracy values obtained for the experiments with several subsets of attributes. Experiment $\star$ gives the reference score obtained by a pixel-based classification of the SITS. From this table, one can first note that these *baseline* scores are quite high, demonstrating the relevance of the temporal dimension for land-cover classification.

These experiments show that the radiometrical mean values of the regions (F_{NIR}, F_R, F_G, F_B), as well as the smoothness and the compactness of the regions (F_{Smooth}, F_{Comp}), are the more relevant features for the classification of this studied area. On the contrary, the area feature F_{Area} seems not relevant for this dataset; a possible reason could be that each crop class contains different sizes of fields. A similar observation can be made for the elongation feature F_{Elong}. This being so, these observations are not questioning the interest of the approach, since these spatial characteristics (and others) could be used in other application cases. In our case, the best result has been obtained with the use of the radiometric mean of the regions combined with the smoothness feature (Experiment 18).

In order to visually assess the results provided by the proposed method, Figure 6(b) shows the best clustering result obtained on sequences from 2007. We also provide in Figure 6(a) the result provided by a naive fusion classification approach. This approach performs by fusing (with a majority vote) the different clustering maps obtained independently for each image of the series. A visual comparison between these two results directly emphasizes the potential of using a pixel-enriched temporal classification approach instead of a naive temporal fusion one. Actually, considering the temporal dimension of the data makes it possible to obtain more accurate classification results.

To separately assess the accuracy of each thematic class, we also provide for the best clustering result obtained (Experiment 18), the precision $\mathcal{P}$, the recall $\mathcal{R}$, the F-measure $\mathcal{F}$ as well as their averages (see Table 2). For comparison purpose,

8

Table 1: Results of the experiments.

	Experiment	Ground truth			Reference map		
		$\overline{\mathcal{F}}$	$\mathcal{K}$	$\mathcal{A}$	$\overline{\mathcal{F}}$	$\mathcal{K}$	$\mathcal{A}$
★	I_{NIR}, I_R, I_G, I_B	69.6	61.9	74.2	51.3	44.2	56.0
1	$I_{NIR}, I_R, I_G, I_B, F_{Comp.}$	**70.5**	60.9	73.7	50.8	43.7	55.3
2	$I_{NIR}, I_R, I_G, I_B, F_{Smooth.}$	69.1	59.7	72.5	**51.5**	**45.2**	**56.1**
3	$I_{NIR}, I_R, I_G, I_B, F_{Smooth.}, F_{Comp.}$	**72.0**	59.1	73.6	51.1	**44.8**	54.4
4	$I_{NIR}, I_R, I_G, I_B, F_{Elong.}$	63.2	52.0	68.6	45.7	37.4	51.1
5	$I_{NIR}, I_R, I_G, I_B, F_{Elong.}, F_{Comp.}$	63.2	49.5	67.0	46.0	36.3	50.9
6	$I_{NIR}, I_R, I_G, I_B, F_{Elong.}, F_{Smooth.}$	62.9	49.5	67.5	46.3	36.2	50.7
7	$I_{NIR}, I_R, I_G, I_B, F_{Elong.}, F_{Smooth.}, F_{Comp.}$	64.6	49.3	68.3	46.4	36.1	50.6
8	$I_{NIR}, I_R, I_G, I_B, F_{Area}$	67.9	56.2	71.2	50.5	44.2	53.7
9	$I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Comp.}$	68.0	55.3	70.7	50.4	**44.6**	53.7
10	$I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Smooth.}$	67.2	56.3	71.1	50.8	43.3	53.4
11	$I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Smooth.}, F_{Comp.}$	67.3	56.1	71.3	50.3	43.2	53.3
12	$I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Elong.}$	65.1	52.1	69.1	47.4	38.3	52.0
13	$I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Elong.}, F_{Comp.}$	64.3	51.2	68.9	46.5	39.0	52.0
14	$I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Elong.}, F_{Smooth.}$	64.8	51.7	69.0	46.9	38.1	51.5
15	$I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Elong.}, F_{Smooth.}, F_{Comp.}$	64.4	50.0	69.0	46.6	38.8	51.7
16	F_{NIR}, F_R, F_G, F_B	**72.7**	**63.0**	**75.7**	**52.1**	**44.5**	**56.0**
17	$F_{NIR}, F_R, F_G, F_B, F_{Comp.}$	**70.1**	58.2	73.4	50.8	**45.6**	55.7
18	$F_{NIR}, F_R, F_G, F_B, F_{Smooth.}$	**72.8**	**66.1**	**77.2**	**52.4**	**45.2**	55.9
19	$F_{NIR}, F_R, F_G, F_B, F_{Smooth.}, F_{Comp.}$	**71.1**	60.9	72.8	50.2	**45.2**	55.7
20	$F_{NIR}, F_R, F_G, F_B, F_{Elong.}$	64.6	52.6	69.1	45.3	37.0	50.3
21	$F_{NIR}, F_R, F_G, F_B, F_{Elong.}, F_{Comp.}$	65.0	52.5	68.9	45.3	37.0	50.4
22	$F_{NIR}, F_R, F_G, F_B, F_{Elong.}, F_{Smooth.}$	65.1	50.8	68.7	46.0	36.6	50.0
23	$F_{NIR}, F_R, F_G, F_B, F_{Elong.}, F_{Smooth.}, F_{Comp.}$	65.0	48.4	67.8	45.4	36.1	49.6
24	$F_{NIR}, F_R, F_G, F_B, F_{Area}$	67.9	53.9	69.5	50.6	44.0	54.7
25	$F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Comp.}$	66.3	52.4	68.8	49.9	43.0	53.8
26	$F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Smooth.}$	66.7	52.3	69.4	50.3	43.8	54.0
27	$F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Smooth.}, F_{Comp.}$	65.8	52.5	69.3	49.7	43.6	54.0
28	$F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Elong.}$	65.1	52.8	68.7	47.8	39.1	52.4
29	$F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Elong.}, F_{Comp.}$	64.8	53.0	68.7	47.8	38.9	52.6
30	$F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Elong.}, F_{Smooth.}$	64.6	51.8	67.9	47.2	38.3	51.8
31	$F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Elong.}, F_{Smooth.}, F_{Comp.}$	63.7	49.6	67.7	47.0	38.1	51.8

The scores that outperform the ones obtained with Experiment ★ (*i.e.*, the reference scores obtained by a pixel-based classification of the SITS) are shown in boldface.

Table 2: Detailed results for Experiment 18.

Class	Colour	Ground truth				Reference map			
		# ($\times 10^3$)	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	# ($\times 10^3$)	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
corn	■	25	89,5	94,3	91,8	192	91,5	83,8	87,5
wheat	■	35	78,3	91,4	84,3	179	67,2	86,5	75,6
temp. meadow	■	6	28,3	59,3	38,3	104	34,7	64,9	45,2
fallow land	■	13	61,5	78,1	68,8	104	28,9	48,3	36,2
meadow	■	3	12,8	63,4	21,3	81	22,7	72,6	34,6
broad-leaved tree	■	1	79,4	91,2	84,9	77	58,8	95,0	72,6
wild land	■	< 0.5	2,4	59,8	4,6	46	26,4	37,4	31,0
sunflower	■	5	54,4	63,6	58,6	45	48,7	58,4	53,1
dense housing	■	< 0.5	0,9	12,5	1,7	36	31,3	58,7	40,8
housing	■		**n/a**			33	15,2	56,1	23,9
barley	■	2	9,2	57,5	15,9	27	10,4	50,5	17,3
soybean	■	9	77,6	69,1	73,1	23	46,5	65,6	54,4
rape	■	3	15,4	68,8	25,2	21	18,9	91,1	31,4
corn for silage	■	7	83,6	99,1	90,7	9	30,6	95,0	46,2
lake	■	6	100,0	98,9	99,5	9	86,5	96,0	91,0
non-irrigated corn	■		**n/a**			6	3,1	18,0	5,2
pea	■		**n/a**			2	1,9	43,4	3,6
sorghum II	■		**n/a**			2	5,0	79,4	9,4
eucalyptus	■	< 0.5	18,2	99,5	30,8	1	1,7	9,8	2,9
conifer	■		**n/a**			1	1,3	9,8	2,3
sorghum	■	< 0.5	4,1	70,2	7,7	1	1,4	33,7	2,6
specific surface	■		**n/a**			< 0.5	1,4	39,2	2,6
water	■		**n/a**			< 0.5	2,1	81,5	4,1
mineral surface	■		**n/a**			< 0.5	0,9	88,8	1,9
gravel pit	■		**n/a**			< 0.5	1,1	99,1	2,1
poplar tree	■	< 0.5	8,3	100,0	15,3	< 0.5	0,4	25,0	0,8
Average		**n/a**	64,5	83,6	72,8	**n/a**	42,1	69,2	52,4

The symbol **n/a** means that the considered value is either not available or not relevant. The symbol # corresponds to the cardinal (number of pixels) of the thematic class (we recall that each image is composed of $1,000 \times 1,000$ pixels).
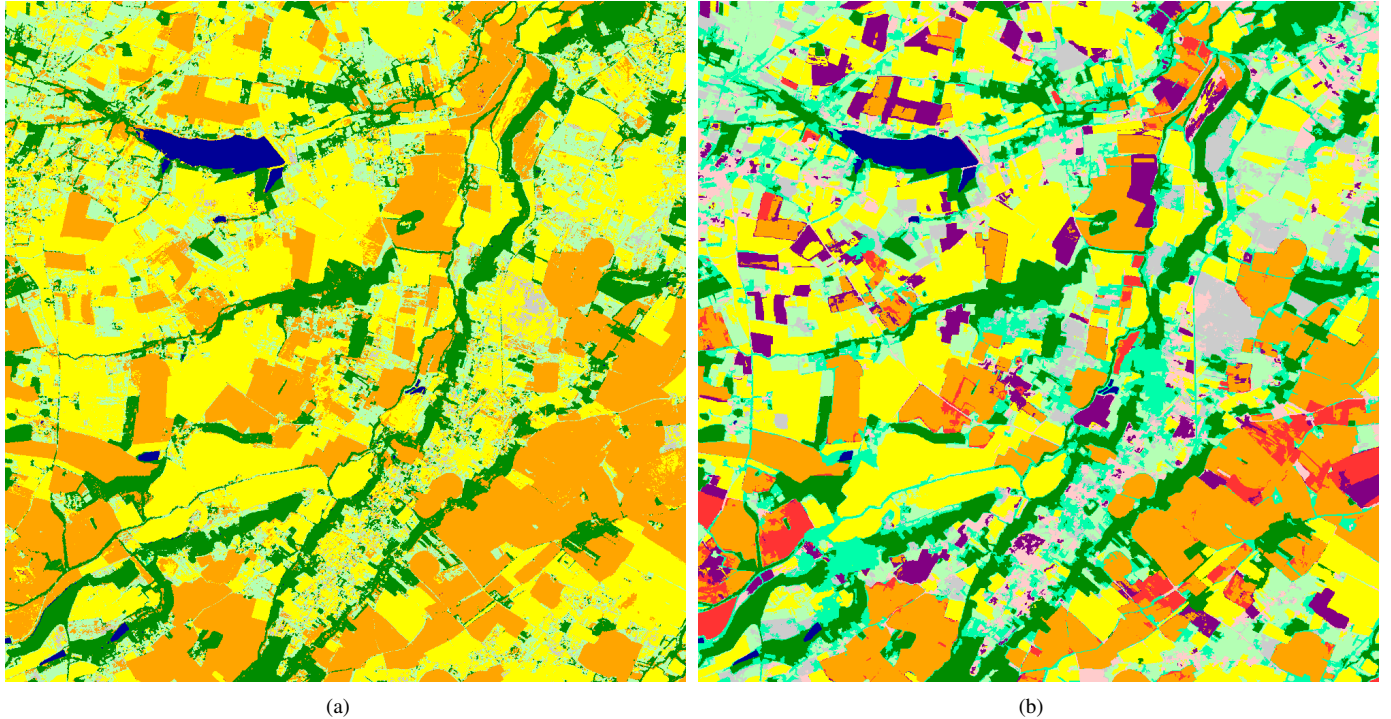
| (a) | (b) |

Figure 6: Clustering maps obtained on the satellite image time series. (a) Result obtained with a naive fusion approach. (b) Result obtained with the proposed method (Experiment 18). Note that these maps have been recolored according to the corresponding land cover reference map (Figure 3(b)).

the confusion matrix obtained by comparing this result to the considered land cover reference map is provided in Table 3. From these two tables, one can note that most of the major considered temporal classes have been correctly extracted by the proposed approach. Table 2 highlights that the proposed approach provides results with high values of precision, recall, and F-measure for most of the extracted classes. For instance these values reach approximatively 85% for the corn and wheat classes which are the most represented ones. Furthermore, the confusion matrix obtained shows that these two temporal classes are mainly regrouped in two clusters by the K-MEANS algorithm. The same observation can be given for the broad-leaved tree class. Such comparisons enable to assess the accuracy of the classification results provided by the proposed pixel-enriched approach. Note that, as the proposed approach provides a clustering of the sensed area, no one-to-one mapping between thematic classes and clusters is guaranteed. In this way, it is not possible to provide statistical accuracies from this matrix. For instance, cluster 8 is predominantly representing the wheat class, but also represents the barley and rape classes. In fact, this cluster represents the broader class of winter crops (*i.e.*, of higher semantic level), precisely composed of these three classes.

Moreover, Figure 7 focuses on a restricted area in order to visualize the differences between the pixel-based approach and the proposed pixel-enriched approach. One can see that, in the details, the land-cover map obtained with the proposed pixel-enriched approach is spatially more consistent and regular than the result obtained with the pixel-based approach. Furthermore, one can note that the orange and yellow classes, correspond-

ing respectively to corn and wheat crop fields, as well as the dark green class corresponding to hardwoods, are well separated. More generally, these results demonstrate visually the relevance of the proposed pixel-enriched approach compared to the pixel-based analysis.

Finally, in order to statistically study the correlation of the considered features, a correlation matrix between these features has been computed (Table 4). To this end, all pixels of all images were characterized by the spatial features (computed on the segmentation ($h_s = 10, h_r = 15, M = 25$)). Not surprisingly, the radiometric features F_R, F_G and F_B are highly correlated (due to the similar reflectances of usual sensed objects in these radiometric bands). This matrix also shows that the spatial features are generally not correlated, except for the couple ($F_{Elong.}, F_{Comp.}$).

5.3. Computation time study

As it is quite difficult to provide a relevant theoretical complexity study of the proposed methodology, we present hereafter an experimental evaluation of the complexity.

Table 5 provides the run-time and the memory usages for the processing of the images contained in the studied dataset sensed over the 2007 cultural year. Experiments have been run on an Intel® Core™2 Quad running at 2.4 GHz with 8 GB of RAM. The algorithms have been implemented using the Java programming language and different threading strategies. From Table 5, one can note that the proposed approach makes it possible to classify a whole HSR SITS in less than 15 minutes. Furthermore the memory consumption remains tractable since it does not exceed 2.3 GB when processing a dataset composed

Table 3: Confusion matrix obtained by comparing the result of Experiment 18 to the considered land-cover reference map (Figure 3(b)).

Class	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	card
corn	7	84	4	4	-	1	-	-	-	-	19 %
wheat	-	-	2	1	2	1	6	2	87	-	18 %
temporary meadow	-	-	15	-	3	3	53	12	14	-	10 %
fallow land	-	-	18	2	10	4	34	17	14	-	10 %
meadow	-	-	35	-	2	23	31	6	2	-	8 %
broad-leaved tree	-	-	4	-	-	96	-	-	-	-	8 %
wild land	-	-	29	-	-	42	13	14	1	-	5 %
sunflower	7	7	11	58	10	1	1	1	3	-	4 %
dense housing	-	-	15	15	53	2	6	4	3	2	4 %
housing	-	-	29	2	32	11	17	6	3	-	3 %
barley	-	-	2	9	4	-	8	-	75	-	3 %
soybean	66	23	6	3	2	-	-	-	-	-	2 %
rape	-	-	1	-	1	-	3	-	95	-	2 %
corn for sillage	-	47	2	49	-	-	-	-	-	-	1 %
lake	-	-	3	-	1	-	-	-	-	96	1 %
non-irrigated corn	13	20	28	3	5	13	1	17	-	-	1 %

The thematic classes covering less than one percent of the sensed surface are not represented in the matrix.

Table 4: Correlation matrix of the features corresponding to the segmentation parameters ($h_s = 10, h_r = 15, M = 25$)).

Feature	F_{NIR}	F_R	F_G	F_B	F_{Area}	$F_{Elong.}$	$F_{Smooth.}$	$F_{Comp.}$
F_{NIR}	1	-0.21	-0.05	-0.15	-0.04	-0.2	-0.07	-0.21
F_R		1	0.96	0.93	-0.06	0.24	-0.05	0.11
F_G			1	0.96	-0.04	0.23	-0.04	0.12
F_B				1	-0.04	0.23	-0.04	0.12
F_{Area}					1	-0.52	0.34	-0.33
$F_{Elong.}$						1	-0.12	0.83
$F_{Smooth.}$							1	0.02
$F_{Comp.}$								1

Table 5: Run-time and memory usage for the processing of the considered dataset.

Step	Runtime	Memory (RAM)
A. Segmentation of the images	9 min 21 s	1.1 GB
B. Characterisation of the regions	2 min 18 s	2.3 GB
C. Construction of the vector images	**n/a**	2.3 GB
D. Construction of the time series	**n/a**	2.3 GB
E. Classification of the time series	1 min 3 s	1.6 GB
Total	≈ 13 min	≈ 2.3 GB

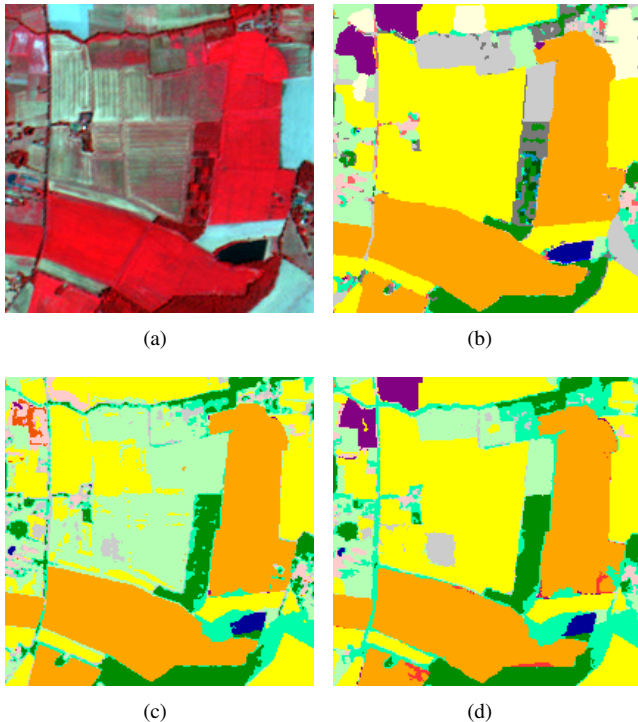The symbol **n/a** means that the considered run-time is not significant.

Figure 7: Extract of the results provided by the proposed method carried out on the satellite image time series. (a) Zoom on the considered ground surface in one image of the series. (b) Zoom on the land cover reference map (Figure 3(b)). (c) Zoom on the clustering map obtained with Experiment ★. (d) Zoom on the clustering map obtained with Experiment 18.

of 15 images of $1,000 \times 1,000$ pixels. For comparison purpose, the classification of the same HSR SITS, without considering the spatial context of the pixels (Experiment ★), requires less than 2 minutes.

6. Conclusion

This article has introduced a novel approach for the analysis of satellite image time series. The originality of this approach lies in its consideration of spatial relationships between pixels in each remotely sensed image. We have seen that characterizing pixels with contextual features computed on segments, allows us to enhance the classification process. This methodology has been carried out on a SITS composed of 15 HSR images. The different classification results obtained have shown the relevance of this approach in the context of the analysis of agronomical areas.

This hybrid paradigm combines the possibilities offered by the (per-pixel) multi-temporal analysis and the relevance of the (single-image) object-based frameworks for spatio-temporal analysis. The coming pair of SENTINEL-2 satellites will provide at the same time images with different spatial and radiometric resolutions (four bands at 10 m, six bands at 20 m and three bands at 60 m) at a high temporal frequency. In this context, the methodology proposed in this article provides a first trend to deal with such data.

We believe this work opens up a number of research directions. Firstly, the choice of the considered spatial features in the classification process has to be deeply studied. For instance, textural and topological features could be used. Secondly, we also plan to validate the proposed methodology by using other segmentation strategies. For instance, it has been proposed in (Kurtz et al., 2012) a new segmentation approach enabling to decompose the scene at different semantic levels. Such an approach could be extended to SITS to analyze the scene in a multi-temporal/multi-level fashion. We also plan to automate the choice of the parameters of segmentation. Indeed, different approaches (supervised or unsupervised) have been proposed to evaluate the quality of a segmentation (Clinton et al., 2010; Özdemir et al., 2010) and thus to select the "best" segmentation result relatively to a particular partitioning task. Finally, the higher the spatial and temporal resolution, the more relevant our approach will be. In this way, the next step of this study could consists of applying this paradigm to a series of Multi-Spectral/Panchromatic images couples. The spatial accuracy of Panchromatic images will help to preserve the fine details and structures.

References

Akcay, H. G., & Aksoy, S., 2008. Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE Transactions on Geoscience and Remote Sensing*, *46(7)*, 2097–2111.

Andres, L., Salas, W., & Skole, D., 1994. Fourier analysis of multi-temporal AVHRR data applied to a land cover classification. *International Journal of Remote Sensing*, *15(5)*, 1115–1121.

Bahirat, K., Bovolo, F., Bruzzone, L., & Chaudhuri, S., 2012. A novel domain adaptation bayesian classifier for updating land-cover maps with class differences in source and target domains. *IEEE Transactions on Geoscience and Remote Sensing*, *(In press)*. 10.1109/TGRS.2011.2174154.

Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65(1)*, 2–16.

Bottou, L., & Bengio, Y., 1995. Convergence properties of the k-means algorithms. In *Proceedings of the Conference on Advances in Neural Information Processing Systems 7*, pp. 585–592. MIT Press.

Bovolo, F., 2009. A multilevel parcel-based approach to change detection in very high resolution multitemporal images. *IEEE Geoscience and Remote Sensing Letters*, *6(1)*, 33–37.

Bruzzone, L., & Carlin, L., 2006. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, *44(9)*, 2587–2600.

Bruzzone, L., & Prieto, D., 2000. Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*, *38(3)*, 1171–1182.

Carleer, A., & Wolff, E., 2006. Urban land cover multilevel region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing*, *27(6)*, 1035–1051.

Clinton, N., Holt, A., Scarborough, J., Yan, L., & Gong, P., 2010. Accuracy assessment measures for object-based image segmentation goodness. *Photogrammetric Engineering and Remote Sensing*, *76(3)*, 289–299.

Comaniciu, D., & Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24(5)*, 603–619.

Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37(1)*, 35–46.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E., 2004. Digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing*, *25(5)*, 1565–1596.

Fan, J., Wang, R., Zhang, L., Xing, D., & Gan, F., 1996. Image sequence segmentation based on 2D temporal entropic thresholding. *Pattern Recognition Letters*, *17(10)*, 1101–1107.

Foody, G., 2001. Monitoring the magnitude of land-cover change around the southern limits of the Sahara. *Photogrammetric Engineering and Remote Sensing*, *67(7)*, 841–848.

Gaetano, R., Scarpa, G., & Poggi, G., 2009. Hierarchical texture-based segmentation of multiresolution remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *47(7)*, 2129–2141.

Gueguen, L., Le Men, C., & Datcu, M., 2006. Analysis of satellite image time series based on information bottleneck. In *Proceeedings of the 27th workshop on Bayesian Inference and Maximum Entropy Methods In Science and Engineering*, pp. 367–374. volume 872.

Hagolle, O., Huc, M., Pascual, D. V., & Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment*, *114(8)*, 1747–1755.

Hall, O., & Hay, G. J., 2003. A multiscale object-specific approach to digital change detection. *International Journal of Applied Earth Observation and Geoinformation*, *4(4)*, 311–327.

Herold, M., Liu, X., & Clarke, K., 2003. Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering and Remote Sensing*, *69(9)*, 991–1001.

Hofmann, P., Lohmann, P., & Müller, S., 2008. Concepts of an object-based change detection process chain for GIS update: IntArchPhRS. In *21st International Society for Photogrammetry and Remote Sensing Congress*, pp. 305–312. volume XXXVII.

Howarth, P., Piwowar, J., & Millward, A., 2006. Time-series analysis of medium-resolution, multisensor satellite data for identifying landscape change. *Photogrammetric Engineering and Remote Sensing*, *72(6)*, 653–663.

Huang, X., & Zhang, L., 2008. An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *46(12)*, 4173–4185.

Idbraim, S., Ducrot, D., Mammass, D., & Aboutajdine, D., 2009. An unsupervised classification using a novel ICM method with constraints for land cover mapping from remote sensing imagery. *International Review on Computers and Software*, *4(2)*, 165–176.

Jensen, J. R., 1981. Urban change detection mapping using Landsat digital data. *Cartography and Geographic Information Science*, *8(21)*, 127–147.

Johnson, R., & Kasischke, E., 1998. Change vector analysis: A technique for the multispectral monitoring of land cover and condition. *International Journal of Remote Sensing*, *19(16)*, 411–426.

Jönsson, P., & Eklundh, L., 2004. TIMESAT – A program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, *30(8)*, 833–845.

Kennedy, R., Yang, Z., & Cohen, W., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr – Temporal segmentation algorithms. *Remote Sensing of Environment*, *114(12)*, 2897–2910.

Kennedy, R. E., Cohen, W. B., & Schroeder, T. A., 2007. Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sensing of Environment*, *110(3)*, 370–386.

Kurtz, C., Passat, N., Gançarski, P., & Puissant, A., 2010. Multiresolution region-based clustering for urban analysis. *International Journal of Remote Sensing*, *31(22)*, 5941–5973.

Kurtz, C., Passat, N., Gançarski, P., & Puissant, A., 2012. Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology. *Pattern Recognition*, *45(2)*, 685–706.

Kurtz, C., Passat, N., Puissant, A., & Gançarski, P., 2011a. Hierarchical segmentation of multiresolution remote sensing images. In P. Soille, M. Pesaresi, & G. K. Ouzounis (Eds.), *Proceedings of the International Symposium on Mathematical Morphology*, pp. 343–354. Springer volume 6671 of *Lecture Notes in Computer Science*.

Kurtz, C., Puissant, A., Passat, N., & Gançarski, P., 2011b. An interactive approach for extraction of urban patterns from multisource images. In *Proceedings of the IEEE Joint Urban Remote Sensing Event*, pp. 321–324.

Lu, D., Mausel, P., Brondizio, E., & Moran, E., 2004. Change detection techniques. *International Journal of Remote Sensing*, *25(37)*, 2365–2401.

Lui, D., & Cai, S., 2011. A spatial-temporal modeling approach to reconstructing land-cover change trajectories from multi-temporal satellite imagery. *Annals of the Association of American Geographers*, *(In Press)*.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In *Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.

Melgani, F., & Serpico, S. B., 2002. A statistical approach to the fusion of spectral and spatio-temporal contextual information for the classification of remote-sensing images. *Pattern Recognition Letters*, *23(9)*, 1053–1061.

Moscheni, F., Bhattacharjee, S., & Kunt, M., 1998. Spatio-temporal segmentation based on region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20(9)*, 897–915.

Niemeyer, I., Marpu, P., & Nussbaum, S., 2008. Change detection using object features. In *Object-Based Image Analysis*, Lecture Notes in Geoinformation and Cartography chapter 2.5. pp. 185–201. Springer Berlin Heidelberg.

Ouma, Y. O., Josaphat, S., & Tateishi, R., 2008. Multiscale remote sensing data segmentation and post-segmentation change detection based on logical modeling: Theoretical exposition and experimental results for forestland cover change analysis. *Computers & Geosciences*, *34(7)*, 715–737.

Özdemir, B., Aksoy, S., Eckert, S., Pesaresi, M., & Ehrlich, D., 2010. Performance measures for object detection evaluation. *Pattern Recognition Letters*, *31(10)*, 1128–1137.

Pesaresi, M., & Benediktsson, J. A., 2001. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *39(2)*, 309–320.

Petitjean, F., Inglada, J., & Gançarski, P., 2011a. Clustering of satellite image time series under time warping. In *Proceedings of the IEEE International Workshop on the Analysis of Multi-temporal Remote Sensing Images*, pp. 69–72.

Petitjean, F., Inglada, J., & Gançarski, P., 2012. Satellite image time series analysis under time warping. *IEEE Transactions on Geoscience and Remote Sensing*, *50(8)*.

Petitjean, F., Ketterlin, A., & Gançarski, P., 2011b. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, *44(3)*, 678–693.

Petitjean, F., Masseglia, F., Gançarski, P., & Forestier, G., 2011c. Discovering significant evolution patterns from satellite image time series. *International Journal of Neural Systems*, *21(6)*, 475–489.

Schopfer, E., Lang, S., & Albrecht, F., 2008. Object-fate analysis: Spatial relationships for the assessment of object transition and correspondence. In *Object-Based Image Analysis*, Lecture Notes in Geoinformation and Cartography chapter 8.4. pp. 785–801. Springer Berlin Heidelberg.

Tiede, D., Lang, S., Füreder, P., Hölbling, D., Hoffmann, C., & Zeil, P., 2011. Automated damage indication for rapid geospatial reporting. *Photogrammetric Engineering and Remote Sensing*, *77(9)*, 933–942.

Tsai, D. M., & Chiu, W. Y., 2008. Motion detection using Fourier image reconstruction. *Pattern Recognition Letters*, *29(16)*, 2145–2155.

Tseng, V. S., Chen, C. H., Huang, P. C., & Hong, T. P., 2009. Cluster-based genetic segmentation of time series with DWT. *Pattern Recognition Letters*, *30(13)*, 1190–1197.

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, *114(1)*, 106–115.

Wu, Q. Z., Cheng, H. Y., & Jeng, B. S., 2005. Motion detection via change-point detection for cumulative histograms of ratio images. *Pattern Recognition Letters*, *26(5)*, 555–563.