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

François Petitjean, Camille Kurtz, Pierre Gançarski

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

François Petitjean[a,*], Camille Kurtz[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 be available. In the classical case, these data are studied by analysing 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 generalisation to a 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 characterise each pixel of the data with a spatial dimension (*i.e.* with contextual information). Providing spatially characterised 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) is a major resource for Earth monitoring. Since the last decades, these image series are either sensed with a high temporal resolution (daily coverage at a kilo-metric spatial resolution) or with a high spatial resolution (weekly coverage at a metric spatial resolution). However, since 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's 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 land-cover maps of a given investigated site.

In order to efficiently use the huge amounts 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 behaviour) with minimal time (by automating the tasks which do not require human expertise), and minimal effort (by reducing the parameters).

In the standard methods, these data are studied by analysing the radiometric evolution of the pixels through the time series. Due to the high spatial resolution of these future images, the geometrical information of the scene could also be considered in the classification process by using region-based approaches. In these methods, a segmentation process is required to extract segments based on radiometric homogeneous hypotheses. Once these segments are extracted, it is possible to characterise them using spatial/geometrical properties, to enhance the classification process.

*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), gancarski@unistra.fr (Pierre Gançarski)

However, the problem of using a segmentation step in a temporal classification 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 issue. This article aims at addressing this issue by characterising a pixel with spatial properties in order to improve the analysis of SITS.

This article is organised as follows. Section 2 gives an overview of existing methods for SITS analysis. Section 3 introduces our method for spatio-temporal analysis of SITS. Section 4 describes the experiments 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 allows the analysis, through observations of land phenomena, of a broad range of applications such as the study of land-cover or even 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 should refer to [1, 2].

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 [3], image ratioing [4], image composition [5] or change vector analysis (CVA) [6]. A second family of **mixed methods**, mainly statistical methods, applies to two or more images. They include methods such as post-classification comparison [7], linear data transformation (PCA and MAF) [8], image regression or interpolation [9] and frequency analysis (*e.g.*, Fourier, wavelets) [10]. Then, we find methods designed more for image time series and based on **radiometric trajectory analysis** [11–14].

Whatever the type of method used in order to analyse satellite image time series, there is a gap between the amount of data representing these time series, and the ability of algorithms to analyse them. Firstly, these algorithms are often dedicated to the study of a change in a scene from bi-temporal representation. Secondly, even if they can map areas of change, they are not able to characterise them. Thirdly, and this point is even more problematic, the geometrical/spatial properties of the data are never taken into account, except for the use of the pixel coordinates. High Spatial Resolution SITS have given rise to the need for spatially and temporally dedicated methods.

To improve the analysing process by using the spatial relationships of the data, object based methods have been more recently proposed [15]. In a first step, the images are segmented/partitioned into sets of connected regions. Then for each region, spatial and/or geometric features (*e.g.* area, elongation, smoothness) are computed in order to characterise the regions [16, 17]. Finally, these regions are classified using these features [18].

Object based methods have shown promising results in the context of mono-image analysis. However, it seems currently difficult to extend these methods to deal with the spatial properties of SITS. Indeed, although several methods have been proposed in order to map segments from one image to another [19], or to directly build spatio-temporal segments [20], their scalability to wide sensed areas and their robustness to local disturbance (temporally and spatially) remains problematic.

This article aims at solving the above-mentioned issues. We therefore propose to classify SITS as the radiometric evolution of atomic sensed areas (*i.e.*, pixels) with time. Then, in order to take into account the spatial properties of the data, we propose to characterise 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 behaviour maps both by preserving the geometrical details in the images and by properly considering the spatial context information.

This paradigm, using spatially characterised pixels, was previously introduced and studied for multi-level segmentation of a single image in [21] and has shown promising results. We propose, in this article, to extend it to the analysis of multi-temporal satellite images.

3. Spatio-temporal analysis methodology

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

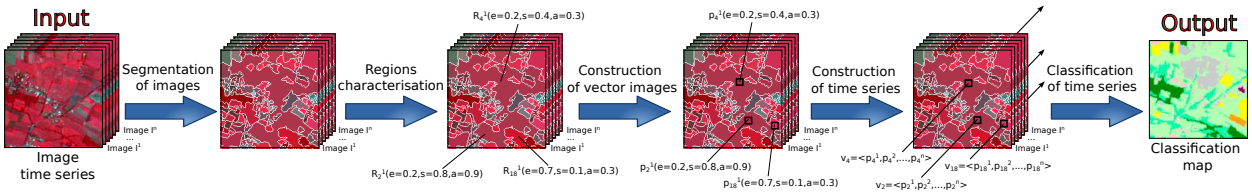


Figure 1: Workflow of the proposed approach.

A. Segmentation of the images;
B. Characterisation 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.

Input / Output

Let us briefly define the input (resp. the output) of the proposed method.

Input. The method takes as input a series $\mathcal{S}_{image} = \langle I^1, \ldots, I^{\mathcal{N}} \rangle$ of $\mathcal{N}$ ortho-rectified[1] images of width $\mathcal{W}$ and height $\mathcal{H}$. Let $\mathcal{B}$ be the number of bands of the images composing the series. Each multivalued (*i.e.* with multiple bands) image I^n ($n \in [\![1, \mathcal{N}]\!]$) can be seen as a function:

$$\begin{array}{rccc} I^n : & [\![1, \mathcal{W}]\!] \times [\![1, \mathcal{H}]\!] & \rightarrow & \mathbb{Z}^{\mathcal{B}} \\ & (x, y) & \mapsto & \prod_{b=1}^{\mathcal{B}} I_b^n(x, y) \end{array} \tag{1}$$

with $\prod$ be the Cartesian product and $[\![a, b]\!]$ denote the interval on $\mathbb{Z}$, bounded by $a, b \in \mathbb{Z}$.

Output. The method gives as output a classification/clustering of the major behaviours of evolution. Such classification can be modelled by a label image $I_C : [\![1, \mathcal{W}]\!] \times [\![1, \mathcal{H}]\!] \rightarrow [\![1, C]\!]$, which associates to each sensed area (x, y) a class value $C(x, y)$ among the C possible ones.

Addendum: Each class from the classification is also modelled by a centroid sequence, which provides another source of information of the main evolution behaviours.

3.1. Segmentation of images

A segmentation of a multivalued image I^n is a partition $\mathfrak{R}^n = \{R_i^n\}_{i=1}^{\mathcal{R}^n}$ of $[\![1, \mathcal{W}]\!] \times [\![1, \mathcal{H}]\!]$; broadly speaking, the scene visualised in I^n is "decomposed" into $\mathcal{R}^n$ distinct parts R_i^n, which are supposed to present specific radiometrical 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

$$\begin{array}{rccc} I_R^n : & [\![1, \mathcal{W}]\!] \times [\![1, \mathcal{H}]\!] & \rightarrow & [\![1, \mathcal{R}^n]\!] \\ & (x, y) & \mapsto & I_R^n(x, y) \end{array} \tag{2}$$

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

[1] Note that, the radiometric levels of the images do not have to be comparable from one image to another.

3.2. Regions characterisation

Numerous features (spectral, geometrical, topological, *etc.*) can be computed for the regions of a segmentation in order to characterise 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 $F(R_i^n) \in \mathbb{R}$.

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

Once a region is characterised by a feature value, it is then possible to affect this value to all the pixels composing the region. Let $\mathcal{A}$ be the number of 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 characterised by two types of information:

- directly sensed values (*i.e.*, $\mathcal{B}$ values, denoted P_b^n with $b \in [\![1, \mathcal{B}]\!]$ where $P_b^n(x, y) = I_b^n(x, y)$);
- region-associated values (*i.e.*, $\mathcal{A}$ values, denoted F_a with $a \in [\![1, \mathcal{A}]\!]$).

All of these values are normalised over the image time series. It is then possible to combine these features to build "enriched" pixels in order to better characterise 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 $\mathcal{N}$ vector images defined as:

$$\begin{array}{rccc} V^n: [\![1,\mathcal{W}]\!] \times [\![1,\mathcal{H}]\!] & \rightarrow & [0,1]^{\mathcal{B}+\mathcal{A}} \\ (x, y) & \mapsto & \prod_{b=1}^{\mathcal{B}} P_b^n(x, y) \times \prod_{a=1}^{\mathcal{A}} F_a(I_R^n(x, y)) \end{array} \tag{4}$$

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:

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

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

3.5. Classification of the time series

The extraction of relevant temporal behaviours from satellite image time series can be realised 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 $[\![1, \mathcal{W}]\!] \times [\![1, \mathcal{H}]\!]$; 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 behaviours. We will denote C_i as a *cluster/class*. The classification can be modelled by a label image $I_C : [\![1, \mathcal{W}]\!] \times [\![1, \mathcal{H}]\!] \rightarrow [\![1, C]\!]$, which associates to each sensed area (x, y) a class value $I_C(x, y)$ among the C possible ones

$$\begin{array}{rccc} I_C: & [\![1,\mathcal{W}]\!] \times [\![1,\mathcal{H}]\!] & \rightarrow & [\![1, C]\!] \\ & (x, y) & \mapsto & I_C(x, y) \end{array} \tag{6}$$

4. Experiments

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 behaviours are quite different (*i.e.*, the wheat crop is ahead of the maize crop).

Thus, the usual strategy for land-cover mapping consists of classifying the temporal radiometric profiles of atomic 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 characterising 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 behaviour (*e.g.*, sunflower crop *vs.* wheat crop). Another (non restrictive) example could be the use of the smoothness of the regions, which could help to distinguish between tree-crop and forest.

4.2. Material: The Satellite Image Time Series used

We detail here the main information concerning the images used for this work. The area of study of this work is located near the town of Toulouse in the South West of France. We have at our disposal 15 cloud-free Formosat-2 images sensed over the 2007 cultural year. These images cover an area of 64 km^2. An image from the series is given in Figure 2 while the temporal repartition of the sensed images is given in Figure 3.

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.

From the instrument radiometric model, digital numbers are first converted into reflectances (normalised physical quantities of solar irradiance). The absolute calibration coefficients used in this step come from the monitoring of Formosat-2 sensors conducted by the French Space Agency (CNES). The inversion of the surface reflectance is then made by comparing the measured reflectance in simulations at the top of the atmosphere, carried out for atmospheric and geometric conditions of measurement. The elevation is taken into account by carrying out simulations for various altitudes, including a weighting of the atmospheric pressure and the amounts of aerosols and water vapour. The state of the atmosphere at the time of the sensing is in turn characterised 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.

4.3. Settings

The five steps described in Section 3 have been performed as follows.

Segmentation of images. In the literature, the most efficient segmentation algorithms are obtained by combining several basic/standard ones. Furthermore, the parameters of these basic segmentation algorithms are tuned according to the characteristics of the image modality (used as input) and the features of the objects to be segmented. To avoid this parametrisation problem, we have chosen to use the Mean-Shift segmentation algorithm [22] to segment each image. Indeed, this algorithm is simple to configure and has shown promising results in the context of the segmentation of remote sensing images [23].

For a given pixel, this algorithm builds a set of neighbouring pixels within a given spatial radius and colour range. The spatial and colour centre of this set is then computed and the algorithm iterates with this new spatial and colour

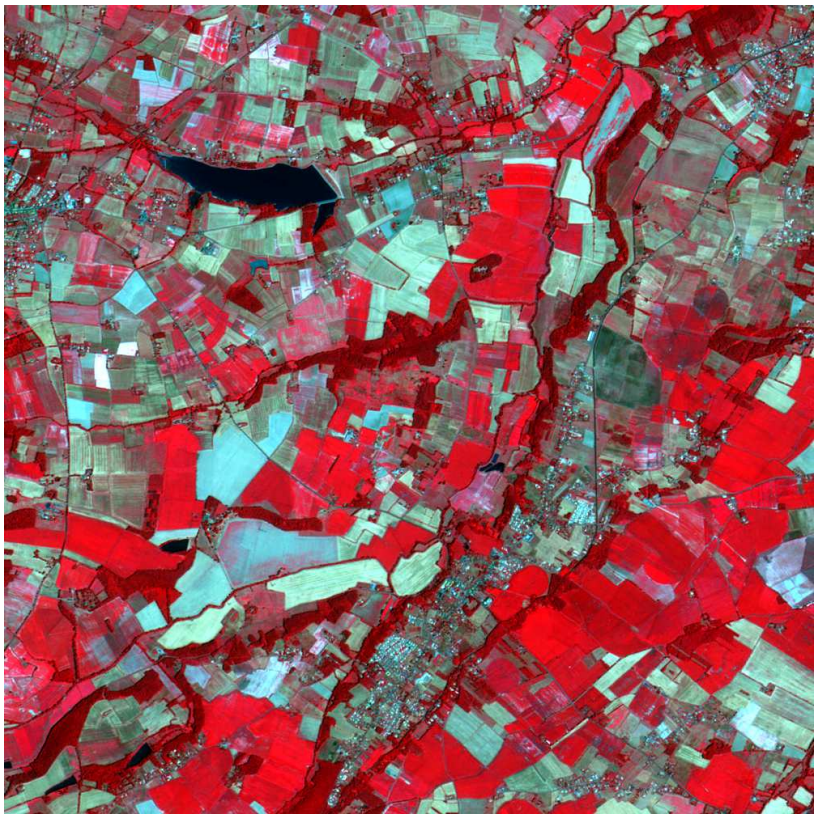


Figure 2: One image from the series (August, 4th 2007).

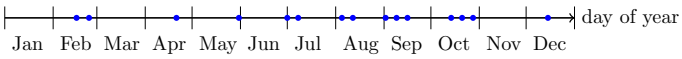


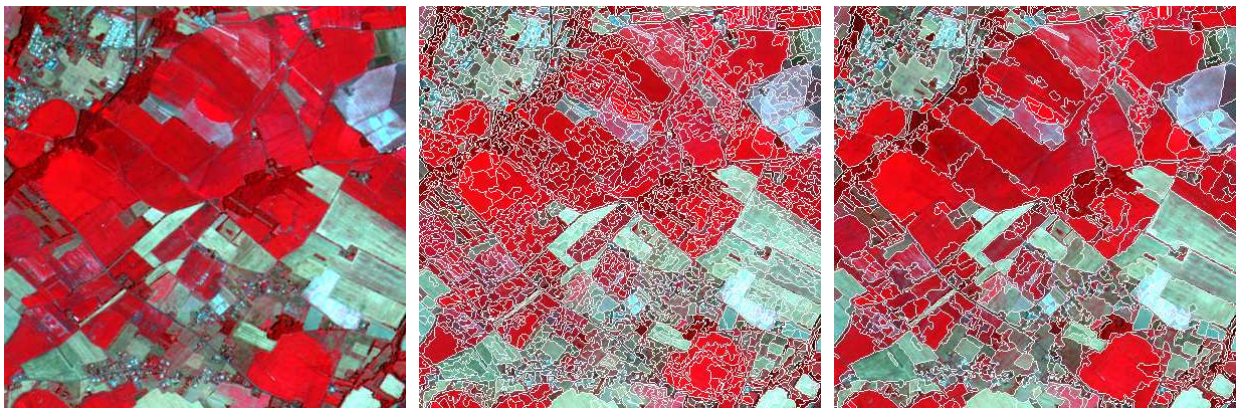
Figure 3: Sensing distribution of images sensed over 2007. Each spot represents a sensed image.

centre. There are three main parameters: the spatial radius (denoted by h_s) used for defining the neighbourhood, the range radius (denoted by h_r) used for defining the interval in the colour space and the minimum size M for the regions to be kept after segmentation.

Since the level of geometrical information extracted by the segmentation algorithm depends on its parametrisation, we have run the algorithm using different parameter configurations. Experiments have shown that the best parameter configuration for the extraction of agricultural areas from these data is ($h_s = 5, h_r = 25, M = 25$). Figure 4 illustrates different segmentation results obtained on an extract of the image presented in Figure 2. One can note that the parameter configuration ($h_s = 5, h_r = 5, M = 25$) provides over-segmented results while the parameter configuration ($h_s = 5, h_r = 25, M = 25$) provides satisfactory results to deal with for the extraction of agricultural areas. We have used the OTB implementation[2] of the Mean-Shift algorithm.

Region characterisation. 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 dissociate forest regions from fields. In this way and without loss of generality, the following region-associated features have been computed:

[2]ORFEO Toolbox (OTB) is an open source library of image processing algorithms developed by the French Space Agency (CNES). `http://www.orfeo-toolbox.org`



(a) (b) (c)

Figure 4: Results of the Mean-Shift segmentation algorithm carried out on an extract of the satellite image presented in Figure 2. (a) Extract of one image of the series. (b) Segmentation result ($h_s = 5, h_r = 5, M = 25$). (c) Segmentation result ($h_s = 5, h_r = 25, M = 25$). Note that the boundaries of the regions are depicted in white.

- 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 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 an 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 $\sqrt{F_{Area}}$.

Construction of vector images. As it was explained previously, each pixel can be characterised by two types of information: directly sensed values (denoted as P_{NIR}, P_R, P_G, P_B), and region-associated values (denoted as $F_{NIR}, \ldots, F_{Smooth.}$). All the values are normalised 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, we tested several combinations of the nine attributes over the time series. Twelve combinations were selected:

$\star$. (P_{NIR}, P_R, P_G, P_B), used as reference;
1. (F_{NIR}, F_R, F_G, F_B);
2. $(P_{NIR}, P_R, P_G, P_B, F_{NIR}, F_R, F_G, F_B)$;
3. $(P_{NIR}, P_R, P_G, P_B, F_{Smooth.})$;
4. $(P_{NIR}, P_R, P_G, P_B, F_{Area})$;
5. $(P_{NIR}, P_R, P_G, P_B, F_{Elong.})$;
6. $(P_{NIR}, P_R, P_G, P_B, F_{Smooth.}, F_{Area}, F_{Elong.})$;
7. $(F_{NIR}, F_R, F_G, F_B, F_{Smooth.})$;
8. $(F_{NIR}, F_R, F_G, F_B, F_{Area})$;
9. $(F_{NIR}, F_R, F_G, F_B, F_{Elong.})$;

10. $(F_{NIR}, F_R, F_G, F_B, F_{Smooth.}, F_{Area}, F_{Elong.})$;
11. All attributes.

Classification of time series. In the machine learning literature, 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. For these reasons, we have chosen to use an unsupervised classification algorithm. We have then applied the classical K-MEANS clustering algorithm [24] to classify the time series previously constructed. The distance used to compare the time series of $\mathcal{S}$ is the Euclidean distance[3].

In agreement with the expert, the K-MEANS algorithm has been used with as many classes (see Appendix A.) as in the reference map (*i.e.*, 25 seeds), and with 15 iterations. Note that any clustering algorithm dealing with numerical data could also be used.

4.4. Validation

The obtained results have been compared to a land cover map produced by the method described in [26] using a comprehensive ground reference data set. Several evaluation indexes have been computed. To assess the accuracy of the extracted thematic classes, we have computed the mean, for each class, of the F-measures obtained $\overline{\mathcal{F}}$. 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:

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

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:

$$\overline{\mathcal{F}} = 2 \cdot \frac{\overline{\mathcal{P}} \cdot \overline{\mathcal{R}}}{\overline{\mathcal{P}} + \overline{\mathcal{R}}} \tag{8}$$

To assess the relevance of the results, we have also computed the Kappa index [27]. The Kappa index $\mathcal{K}$, which is a measure of global classification accuracy, is defined as:

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

where $\Pr(a)$ is the relative agreement among the observers, and $\Pr(e)$ is the hypothetical probability of chance agreement. The Kappa index takes value in $[0, 1]$ and decreases as the classification is in disagreement with the ground-truth map. We have computed this index as follows. The approach consists of considering all point couples $(\mathbf{x_1}, \mathbf{x_2}) = ((x_1, y_1), (x_2, y_2))$ and see the configuration of these two points in each partition (the clustering result and the ground-truth). There are four possible configurations; for each one, a counter is associated and incremented each time a configuration appears:

1. $\mathbf{x_1}$ and $\mathbf{x_2}$ belong to the same partition both in the clustering and in the reference map (counter s_s);
2. $\mathbf{x_1}$ and $\mathbf{x_2}$ belong to the same partition in the clustering but not in the reference map (counter s_d);
3. $\mathbf{x_1}$ and $\mathbf{x_2}$ belong to the same partition in the reference map but not in the clustering (counter d_s);
4. $\mathbf{x_1}$ and $\mathbf{x_2}$ belong to the same partition neither in the reference map nor in the clustering (counter d_d).

[3]Note that other distances (and more relevant temporal ones [25]) could also be used. However, we remind that this article is focused on the data representation and not on the relevance of the classification algorithm.

Table 1: Results of the experiments.

	Experiment	$\overline{\mathcal{F}}$	$\mathcal{K}$
⋆	(P_{NIR}, P_R, P_G, P_B)	51.9	87.7
1	(F_{NIR}, F_R, F_G, F_B)	**52.3**	**87.8**
2	$(P_{NIR}, P_R, P_G, P_B, F_{NIR}, F_R, F_G, F_B)$	51.8	**87.8**
3	$(P_{NIR}, P_R, P_G, P_B, F_{Smooth.})$	**52.7**	**88.0**
4	$(P_{NIR}, P_R, P_G, P_B, F_{Area})$	51.7	86.6
5	$(P_{NIR}, P_R, P_G, P_B, F_{Elong.})$	46.0	87.3
6	$(P_{NIR}, P_R, P_G, P_B, F_{Smooth.}, F_{Area}, F_{Elong.})$	48.5	87.1
7	$(F_{NIR}, F_R, F_G, F_B, F_{Smooth.})$	**53.2**	**88.0**
8	$(F_{NIR}, F_R, F_G, F_B, F_{Area})$	51.8	86.6
9	$(F_{NIR}, F_R, F_G, F_B, F_{Elong.})$	45.7	87.3
10	$(F_{NIR}, F_R, F_G, F_B, F_{Smooth.}, F_{Area}, F_{Elong.})$	47.6	86.9
11	All attributes	51.2	87.2

The best scores are shown in boldface.

Thus, the Kappa index is computed with:

$$\Pr(a) = \frac{s_s + d_d}{s_s + s_d + d_s + d_d} \tag{10}$$

and

$$\Pr(e) = \frac{(s_s + s_d) \cdot (s_s + d_s) + (s_d + d_d) \cdot (d_s + d_d)}{(s_s + s_d + d_s + d_d)^2} \tag{11}$$

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.

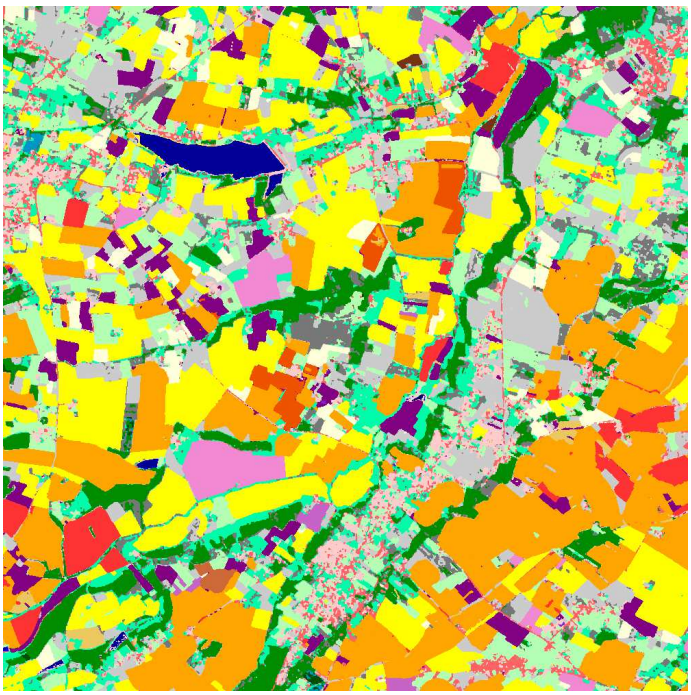
5. Results

This section presents the results of the proposed contextual approach for multi-temporal analysis. These experiments aim at showing a non-restrictive example of the use of our approach in the classification of an agronomical HSR SITS.

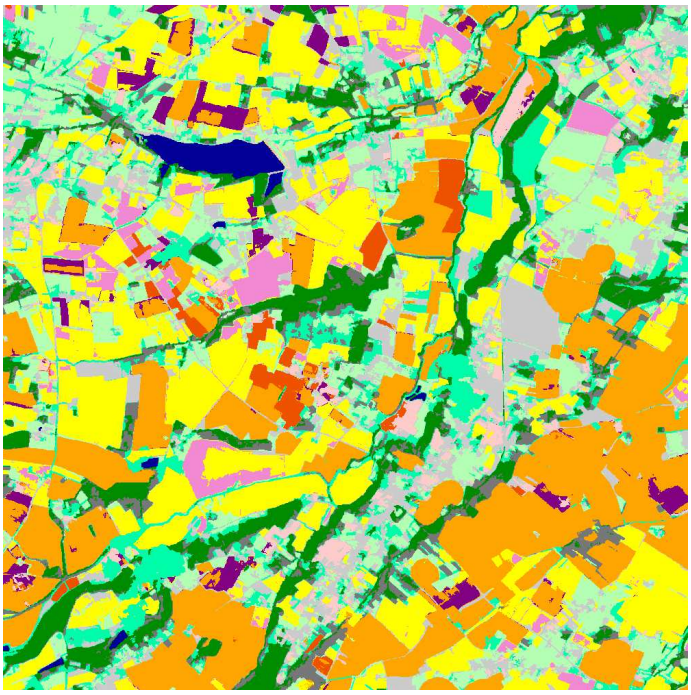
Table 1 summarises the F-measure and the Kappa values obtained for the experiments with several attributes projections (see Section 4). Experiment ⋆ gives the reference score obtained by a pixel-based classification of the SITS. From this table, one can first note that these *baseline* scores are quite high, demonstrating the relevance of the temporal dimension for land-cover classification. In addition, note that due to the significant number of classes in the reference map, the Kappa scores are quite high and a small increase of the score can correspond to a major difference in the classification.

These experiments show that the mean value of the regions, as well as the smoothness of the regions, are relevant features for the classification of the specific studied area. However, the area feature seems actually not relevant for this dataset, since each crop class contains different size of fields. A similar observation can be made for the elongation feature. Nevertheless, these observations are not questioning the genericity of the approach, since these spatial characteristics (and others) could be used in other applications. In our case, the best result has been obtained with the use of the radiometric mean of the regions combined with their smoothness.

In order to visually assess the results provided by the proposed method, Figure 5 shows the best clustering result obtained on sequences from 2007, as well as the corresponding reference land cover map. Finally, Figure 6 focuses on a restricted area in order to visualise 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



(a)



(b)

Figure 5: Results of the proposed method carried out on the satellite image time series. (a) Land cover reference map. (b) Clustering map obtained (Experiment 7). Note that this clustering map has been recoloured according to the corresponding land cover reference map.

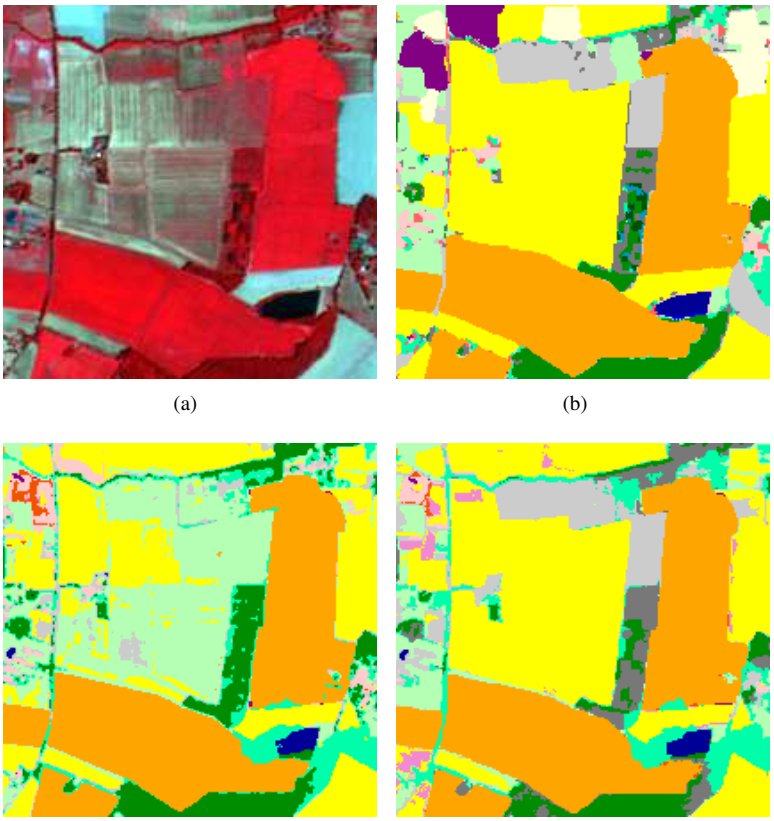


Figure 6: Zoom on an extract of the results provided by the proposed method carried out on the satellite image time series. (a) Zoom on the considered ground surface in one image of the series. (b) Zoom on the Land cover reference map. (c) Zoom on the clustering map obtained with Experiment ⋆. (d) Zoom on the clustering map obtained with Experiment 7.

spatially more consistent and regular than the result obtained with the pixel-based approach. Furthermore, one can note that the orange and yellow classes, corresponding 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 statistically and visually the relevance of the proposed pixel-enriched approach compared to the pixel-based analysis.

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 characterising pixels with contextual features computed on segments, allows us to enhance the classification process. This methodology has been carried out on a SITS; the first results obtained have shown the relevance of this approach.

We believe this work opens up a number of research directions. Firstly, the choice of spatial features and of their weighting 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 another segmentation strategies. For instance, it has been proposed in [28] a new segmentation approach enabling to decompose the scene at different semantic levels. Such an approach could be extended to SITS to analyse the scene in a multi-temporal/multi-level fashion. 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 images couples Multi-Spectral / Panchromatic. The spatial accuracy of Panchromatic images will help to preserve the fine details and structures.

This hybrid paradigm combines the possibilities offered by the (per-pixel) multi-temporal analysis and the relevance of the (mono-image) region-based frameworks for spatio-temporal analysis. The coming SENTINEL-2 satellite will provide at the same time images with different spatial and radiometric resolutions at a high temporal frequency. In this context, the methodology proposed in this article provides a first trend to deal with such data.

Acknowledgements

The authors would like to thank the French Space Agency (CNES) and Thales Alenia Space for supporting this work under research contract n°1520011594 and the researchers from CESBIO (Danielle Ducrot, Claire Marais-Sicre, Olivier Hagolle and Mireille Huc) for providing the land-cover maps and the geometrically and radiometrically corrected FORMOSAT-2 images.

Appendix A. Legend of the maps

Colour	Class
	corn
	corn for silage
	non-irrigated corn
	wheat
	sunflower
	sorghum
	sorghum II
	soybean
	barley
	pea
	rape
	broad-leaved tree
	conifer
	poplar tree
	eucalyptus
	water
	lake
	gravel pit
	meadow
	temporary meadow
	fallow land
	wild land
	high density housing surface
	specific urban surface
	low density housing surface
	mineral surface

References

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

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

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

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

[5] Y. O. Ouma, S. Josaphat, R. Tateishi, 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) (2008) 715–737.

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

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

[8] P. Howarth, J. Piwowar, A. Millward, Time-Series Analysis of Medium-Resolution, Multisensor Satellite Data for Identifying Landscape Change, Photogrammetric engineering and remote sensing 72 (6) (2006) 653–663.
[9] R. E. Kennedy, W. B. Cohen, T. A. Schroeder, Trajectory-based change detection for automated characterization of forest disturbance dynamics, Remote Sensing of Environment 110 (3) (2007) 370–386.
[10] L. Andres, W. Salas, D. Skole, Fourier analysis of multi-temporal AVHRR data applied to a land cover classification, International Journal of Remote Sensing 15 (5) (1994) 1115–1121.
[11] P. Jönsson, L. Eklundh, Timesat–a program for analyzing time-series of satellite sensor data, Computers & Geosciences 30 (8) (2004) 833–845.
[12] J. Verbesselt, R. Hyndman, G. Newnham, D. Culvenor, Detecting trend and seasonal changes in satellite image time series, Remote Sensing of Environment 114 (1) (2010) 106–115.
[13] F. Petitjean, P. Gançarski, F. Masseglia, G. Forestier, Analysing satellite image time series by means of pattern mining, in: 11th International Conference on Intelligent Data Engineering and Automated Learning, Vol. 6283 of Lecture Notes in Computer Science, Springer, 2010, pp. 45–52.
[14] R. Kennedy, Z. Yang, W. Cohen, Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr – Temporal segmentation algorithms, Remote Sensing of Environment 114 (12) (2010) 2897–2910.
[15] T. Blaschke, Object based image analysis for remote sensing, ISPRS Journal of Photogrammetry and Remote Sensing 65 (1) (2010) 2–16.
[16] M. Herold, X. Liu, K. Clarke, Spatial metrics and image texture for mapping urban land use, Photogrammetric Engineering and Remote Sensing 69 (9) (2003) 991–1001.
[17] A. Carleer, E. Wolff, Urban land cover multilevel region-based classification of VHR data by selecting relevant features, International Journal of Remote Sensing 27 (6) (2006) 1035–1051.
[18] C. Kurtz, N. Passat, P. Gançarski, A. Puissant, Multiresolution region-based clustering for urban analysis, International Journal of Remote Sensing 31 (22) (2010) 5941–5973.
[19] L. Gueguen, C. Le Men, M. Datcu, Analysis of satellite image time series based on information bottleneck, in: Bayesian Inference and Maximum Entropy Methods In Science and Engineering, Vol. 872, 2006, pp. 367–374.
[20] F. Moscheni, S. Bhattacharjee, M. Kunt, Spatio-temporal segmentation based on region merging, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (9) (1998) 897–915.
[21] L. Bruzzone, L. Carlin, A multilevel context-based system for classification of very high spatial resolution images, IEEE Transactions on Geoscience and Remote Sensing 44 (9) (2006) 2587–2600.
[22] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5) (2002) 603–619.
[23] X. Huang, L. Zhang, An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery, IEEE Transactions on Geoscience and Remote Sensing 46 (12) (2008) 4173–4185.
[24] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.
[25] F. Petitjean, J. Inglada, P. Gançarski, Clustering of satellite image time series under time warping, in: IEEE International Workshop on the Analysis of Multi-temporal Remote Sensing Images, 2011, pp. 69–72.
[26] S. Idbraim, D. Ducrot, D. Mammass, 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 4 (2) (2009) 165–176.
[27] R. Congalton, A review of assessing the accuracy of classifications of remotely sensed data, Remote Sensing of Environment 37 (1) (1991) 35–46.
[28] C. Kurtz, N. Passat, P. Gançarski, A. Puissant, Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology, Pattern Recognition 45 (2) (2012) 685–706.