Urban land cover analysis from satellite image time series based on temporal stability

Mohamed Chelali*, Camille Kurtz*, Anne Puissant†, Nicole Vincent*
*Université Paris Descartes, LIPADE, France
†Université de Strasbourg, LIVE, France
{firstname.lastname}@{parisdescartes,unistra}.fr

***Abstract*—Satellite Image Time Series (SITS) provide valuable information for the automatic mapping of our territories. In this article we focus on the analysis of urban land covers from SITS, trying to evaluate the density of artificialized areas. We hypothesize that such areas do not evolve significantly through time (over the interval of a year) compared to other non-artificialized areas (*e.g.*, agricultural crops, vegetation). The proposed approach is based on a spatio-temporal characteristic measuring the temporal stability of a zone, extracted using the Run Length Encoding method. Preliminary results obtained on a series of 41 SENTINEL-2 images highlight the ability of our approach to discriminate different urban land-cover classes (*e.g.*, artificial areas, high density vs. low density housing areas).**

***Index Terms*—urban density, satellite image time series, spatio-temporal features, Run Length Encoding, SENTINEL-2**

I. Introduction

The terrestrial surface is regularly observed by satellites, leading to a large amount of images. Novel satellite constellations periodically acquire medium-resolution images around the globe, on the same geographical area, and with increasing frequency. For example, the SENTINEL-2 sensors produce optical Satellite Image Time Series (SITS) with a revisit time of 5 days and a spatial resolution of 10 – 20 meters.

One important application of SITS is the mapping of land cover (*e.g.*, urban areas, agricultural zones) and the identification of land use changes (*e.g.*, urbanization, deforestation). The growing availability of such temporal data makes it possible to produce and update accurate land-cover maps of a territory.

This article focuses on the analysis of the land-cover of urban areas from SITS trying to evaluate the density of artificialized areas. We hypothesize that artificialized areas (*e.g.*, housing, industrial areas) in urban environments do not evolve significantly through time compared to other non-artificialized areas (*e.g.*, agricultural crops, vegetation). An approach based on a spatio-temporal characteristic measuring the temporal stability of a zone is proposed. We do not aim to study temporal land-cover maps or land use changes (*e.g.*, urbanization) but to analyze the actual land-cover.

This article is organized as follows. Section II recalls some existing methods for SITS analysis. Section III introduces our spatio-temporal characteristic for the analysis of urban density from SITS. Section IV describes the experimental validation. Conclusions and perspectives will be found in Section V.

The French ANR supported this work under Grant ANR-17-CE23-0015.

II. Existing approaches for SITS analysis

SITS analysis allows the analysis, through observations of land phenomena with a broad range of applications such as the study of land-cover or even the mapping of damage following a disaster. These changes may be of different types, origins and duration. For a detailed survey, see [5].

Pioneer methods for analyzing SITS operated on single images or stacks of images. On each image, the different measurements per pixel were considered as independent features and involved in classical machine learning-based classifiers. In such approaches, the date of the measurements was ignored in the feature space. Bi-temporal analysis can locate and study abrupt changes occurring between two observations. These methods include image differencing [4], ratioing [10] or change vector analysis [11].

Then, we find methods designed more for image time series and based on multi-date classification approaches such as radiometric trajectory analysis [18]. Such approaches exploit the notion that land-cover can vary through time (*e.g.*, because of seasons, vegetation evolution [17]), and they take into account the order of measurements by using dedicated time series analysis methods [3]. Every pixel is viewed as a temporally ordered (and aligned) series of measurements, and the changes of the measurements through time are analyzed to find (temporal) patterns.

Concerning the type of features, "frequency-domain" approaches include spectral analysis, wavelet analysis [2] while "time-domain" approaches involve auto-correlation and cross-correlation analysis. Concerning the classification method, the classical way is to measure similarity between any incoming sample and the training set; and assign a label to the most similar class using *e.g.*, the Euclidean distance based on a nearest neighbor algorithm or the Dynamic Time Wrapping method [13]. Deep learning approaches have also been recently considered [6], [8]. Finally some methods first transform the SITS into a new space to extract more discriminative "hand-crafted" features [14], [16] in order to be used by a classifier.

In this work we adopt this last strategy. Our hypothesis relies on the fact that artificialized areas (housing areas, industrial zones) in urban environments do not evolve significantly through time compared to other non-artificialized areas (agricultural crops, urban vegetation, etc.). For instance the radiometric value of a pixel representing a building roof

978-1-7281-0009-8/19/$31.00 ©2019 IEEE

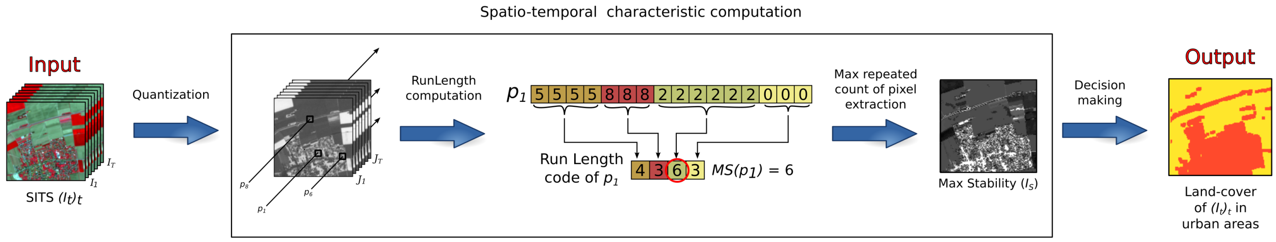


Fig. 1. Flowchart of the method for urban land-cover analysis from SITS based on temporal stability.

should remain stable through time over the interval of a year, if no abrupt change appears. We propose an approach based on a spatio-temporal characteristic measuring the temporal stability of a zone using a classical compression technique. This classic lossless compression technique is known as Run Length Encoding (RLE) [7]. It was already used for time series analysis in [1], [15].

III. Proposed approach

We present a labeling method not relying on an explicit learning phase. This labeling is based on a "hand-crafted" spatio-temporal characteristic we define, it measures the temporal stability of a zone. This enables to apply the method in various sensed regions without requiring a learning phase. Clustering method can be applied with a minimum of parameters to be tuned and transferred. Besides, the labeling can be performed at pixel level or at block level, depending on the application. Pixels or blocks will be characterized and then a clustering approach enables to take a decision in urban zones. The flowchart of the method is presented in Figure 1.

A. Spatio-temporal stability feature

The material being a SITS $(I_t)_{t\in[\![1,T]\!]}$, images are labeled with the acquisition date t, and a spatial domain D on which the images are registered, the purpose is to define spatial domains with rather homogeneous properties in terms of temporal stability with respect to the pixel value, either scalar or vectorial (*e.g.*, panchromatic / multi-spectral values, NDVI).

The characteristic we are to define is the duration of the longest laps of time where the pixel labels stay "constant", it will be noted MS for "Max Stability". It can be defined for a pixel p and we extend it to a spatial domain Z. The core of the method is twofold, it lays in the definition of "constant" and of the spatial domains on which the computation is performed.

First let us consider a pixel p. The temporal evolution of the pixel value along the $(I_t)_t$ series can be compressed using the sequence RLE [7]. We define $MS(p)$ as the length of the longest run expressed in days. The run can be placed at any time, only its length is considered.

When a zone is considered, $Z = \{p_i\}_{i=1}^{n}$ the extension of MS function is achieved using an operator $\perp$ applied to all $MS(p)$ in the zone as

$$MS(Z) = \perp_{i=1}^{n} MS(p_i) \tag{1}$$

B. Equality definition

The equality of values is not always significant, since the pixel values being either continuous or discrete in intervals such as $[0, 255]$ or hypercube when vectorial values are considered.. Furthermore pixel values at different dates are compared whereas the acquisitions are different. Then, a new definition has to be given to the term *equal* and we have achieved this, based on a quantization of the pixel values. The quantization cannot be applied to each image, it would lead to non comparable values when the acquisition dates are different. It has to be done at a global level of all the pixels of $(I_t)_t$. The quantization could be regular, fixed with respect of the usual distribution of the values or it can be adapted to the image series, to the nature of the characteristic used. The latter has been our choice. This can avoid a learning phase. Independently of the structure of the characteristics considered, a clustering algorithm enables to define the significant values. In our case a k-means algorithm, $k_{quantiz}$ being a parameter of the method, is used and different values have been experimented according to the precision needed in the problem to be solved. Then, the pixel values are replaced by the cluster label belonging to $\{1, 2, \ldots, k_{quantiz}\}$ defining new scalar images $(J_t)_t$. MS is thus computed on $(J_t)_t$.

C. Global or local temporal approach

The previous characteristic can be computed at a global or a local temporal level.

As the images in the collection are not regularly distributed along the year, an interpolation makes possible to complete the series to build one image each day figuring a virtual acquisition. In our case a linear interpolation was simply performed. Then, the temporal interval of study can be recovered with N overlapping or non-overlapping sub-intervals of the same duration in order to make comparisons possible with equal temporal distances between consecutive images. In our case we considered a local duration of one month. The max stability values at each pixel on the local temporal blocks were obtained as previously defined at the global level, leading then to a *spatio-temporal* characteristic. This local study enables, either at the pixel level or at the zone level to compute N max stability values associated with these entities. The entity representation of the image domain is done in $\mathbb{R}^N$.

Whether we consider the local or global approach to compute MS, an image I_s is defined on the domain D. It is either a scalar (temporal global approach) or a vectorial image (temporal local approach).

D. Decision strategy

From the information of the max stability value or the temporal vector of local max stability values, a decision on the urban land use of the soil has to be taken with more or less precision. The strategy differs according to the targeted goal. We considered two different objectives, either the finding of homogeneous regions with respect to the urban density or the labeling of (already delineated) urban blocks.

1) Pixel level: The brighter a pixel of the I_s image is, the more stable the associated space is. As only temporal information has been used and no spatial information, a smoothing of the I_s image is needed. A gray level morphological dilation with a square structural element of size d is performed. Then the decision is deduced from a clustering, a k-means clustering in our case. As the parameter $k_{decision}$ is monotonic with the longest stable duration, the center of the classes can be ordered. Thus, the pixels can be labeled according to the value of their cluster inertia center. We hypothesize that highest values correspond to the most dense zones.

2) Block level: In order to label a block Z according to its max stability, first a $\perp$ operator has to be defined. A vectorial characteristic is defined as the (composition) histogram of the $MS(p)$ in the zone. Such "composition histogram" strategy already showed its interest in the case of urban areas in [12]. The bins of the histogram are defined at a global level on the whole image I_s by way of a k-means algorithm using the k_{bin} parameter. The final label associated with the block is obtained by applying a final k-means algorithm on the vectorial characteristic, $k_{decision}$ being the parameter.

IV. Experimental study

Our approach has been employed for the analysis of urban land-covers from a series of satellite images. As validation, two scales of analysis were considered: the large urban districts and the urban blocks, with finer thematic classes.

A. Material

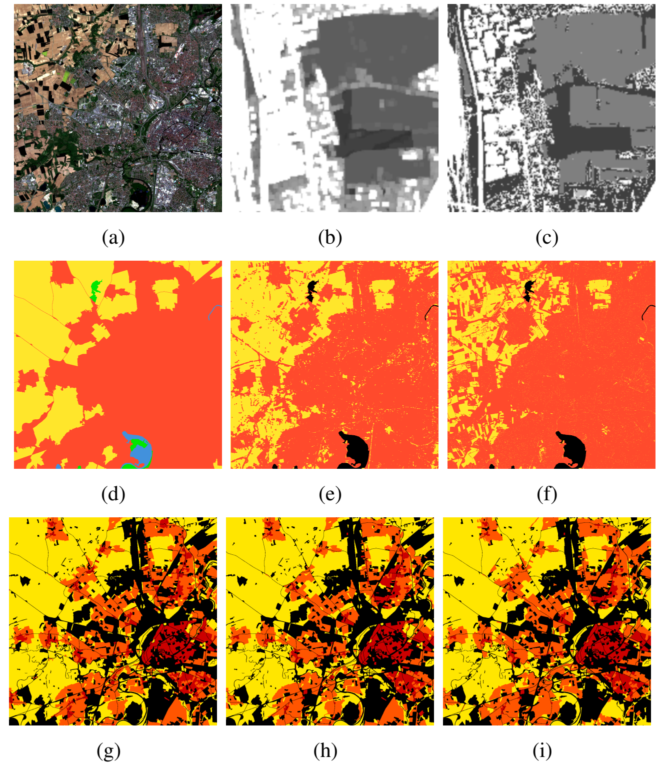
The study area is the city of Strasbourg, France. There are 41 SENTINEL-2 images (1000×1000 pixels) sensed in 2017 (Figure 2(a) shows an image of the SITS). The optical products have been orthorectified.

From these images, we considered different characteristics, the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Water Index (NDWI) and the Brightness Index (BI) built from the multi-spectral product, at a spatial resolution of 10m. Actually, these physical variables values are quite comparable to each other through the series, and are quite consistent indexes for sensed urban scene description.

These characteristics are considered individually to compute MS, leading to I_s^{NDVI}, I_s^{NDWI}, I_s^{BI} or combined together $I_s^{\star}$ at the decision stage (each channel of $I_s^{\star}$ is considered as an independent attribute in the final k-means algorithm).

B. Evaluation strategies

Our results have been evaluated with ground-truth (GT) maps, extracted from the CIGAL2011-2012 database (https://www.cigalsace.org/). Two scales were considered:



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 2. Illustration of the data and results: (a) 1^{st} image of the SITS; (b) zoom on a result: global temporal method, I_s^{NDVI}; (c) zoom on a result: local temporal method, I_s^{NDVI}; (d) GT map at the the "urban area" level (Artificial surfaces (red), Agricultural surfaces (yellow)); (e) pixel level decision: global temporal approach, $k_{decision} = 4$, I_s^{NDVI}; (f) pixel level decision: local temporal approach, $k_{decision} = 4$, I_s^{NDVI}; (g) GT map at the "urban blocks" level (3 classes are considered: high density housing areas (red), low density housing areas (orange), Vegetation and agricultural surfaces (yellow)); (h) block level decision: global temporal approach, $k_{decision} = 7$, I_s^{NDVI}. (i) block level decision: local temporal approach, $k_{decision} = 7$, I_s^{NDVI}).

- the "urban area" level (1/250000) to map the territory (see GT map on Figure 2(d)), with a low detail level, enabling to study the large urban districts (artificialized surfaces vs. agricultural surfaces).
- the "urban blocks" level (1/10000), blocks being defined as the minimal cycles formed by communication ways. The GT map (Figure 2(g)) contains 1069 blocks (polygons) labeled with 3 thematic classes (high density housing areas, low density housing areas and agricultural & vegetation surfaces). Miscellaneous zones are in black.

The pixel-based decision level of our approach was validated at the "urban area" scale while the block-based decision level was validated at the "urban blocks" scale.

As quality indexes, we derived from our clustering results and the GT maps a confusion matrix, from which we computed the classical recognition rate (RR), precision (P), recall (R) and F_1-measure (F). We selected automatically the best mapping between clusters and thematic classes by optimizing the RR.

C. Obtained results

As explained in Section III-B, the MS characteristics have been computed on quantized pixel values. We set empirically

TABLE I
QUALITY OF THE RESULTS FROM COMPARISONS WITH GT MAPS.

	Temporal approach	Config.	RR	P	R	F
Pixel level decision	Global temporal	I_s^{NDVI}	89	88.93	89	88.98
		I_s^{NDWI}	86.24	86.86	86.25	86.55
		I_s^{BI}	84.5	84.75	84.50	84.63
		$I_s^{\star}$	89.84	89.76	89.85	89.80
	Local temporal	I_s^{NDVI}	81.7	82.7	81.7	82.2
		I_s^{NDWI}	80.24	80.55	80.25	80.4
		I_s^{BI}	76.47	79.2	76.48	77.81
		$I_s^{\star}$	83.14	84.64	83.15	83.89
Block decision	Global temporal	I_s^{NDVI}	78.48	77.98	78.48	78.23
		I_s^{NDWI}	69.59	60.77	69.60	64.89
		I_s^{BI}	72.02	73.06	72.03	72.54
	Local temporal	I_s^{NDVI}	74.18	74.50	74.18	74.34
		I_s^{NDWI}	68.75	68.06	68.76	68.41
		I_s^{BI}	67.25	66.26	67.26	66.75

$k_{quantiz}$ equal to 4.

We used the pixel level decision strategy from I_s^{NDVI}, I_s^{NDWI}, I_s^{BI} and $I_s^{\star}$ by considering both the global and local temporal approaches. For the local temporal approaches, N was set to 23 with month overlapping temporal intervals. We evaluated the results by experimentally varying $k_{decision}$ from 2 to 10. Best results were obtained when $k_{decision} = 4$ so we kept this value for all experiments. We also evaluated the impact of the size d of the structural element of the dilation. The d value was then fixed at 5.

Table I (first two lines) presents the quantitative results while Figure 2(e, f) illustrates visual results. From the quantitative results, we observe that, overall, the global temporal approach provides higher accuracy scores on this application than the local temporal approach. The combination of the different features in $I_s^{\star}$, slightly improves the accuracy scores, showing the complementary of the features. When we visualize the computed characteristics (Figure 2(b, c)), we observe that the local approach makes it possible to discriminate finer temporal stability behaviors in peri-urban areas than the global approach. However the (coarse) scale of the GT does not allow to take this into account in the quantitative analysis.

We used the block level decision strategy from I_s^{NDVI}, I_s^{NDWI} and I_s^{BI} by considering both the global and local temporal approaches. We considered the composition histogram strategy by varying both k_{bin} and $k_{decision}$ from 2 to 10. Best results were obtained with $k_{bin} = 8$ and $k_{decision} = 7$ for I_s^{NDVI}. The results are summarized in Table I (last two lines), and some visual illustrations are shown in (Figure 2(h, i)). Here also the global temporal approach provides higher accuracy scores on this case. This can be due to our equality definition that we plan to improve as mentioned later.

V. CONCLUSION AND PERSPECTIVES

We presented in this article an approach for studying urban land-cover classes from SITS. This method is based on the definition of a spatio-temporal characteristic trying to locate the long period where the pixel values stay stable trough time. An experimental study has been conducted on a series of SENTINEL-2 images. The preliminary results obtained highlight the ability of this approach to discriminate different urban land-cover classes. In another context, in [9], the results obtained on urban land-cover analysis from SITS at the country scale have an accuracy at the same level for high and low density housing classes. As a short-term perspective, we plan to work on the notion of equality, used to decide if a pixel value is stable through time. For instance, when the characteristic is vectorial, it is possible to try different definitions of the equality, leading to a more or less constrained equality definition. Such vectorial approach is more suited when the considered scalar characteristics are independent, that is not yet the case with those involved in our application.

REFERENCES

[1] S. Aminikhanghahi and D. J. Cook. A survey of methods for time series change point detection. *Knowl. Inf. Syst.*, 51(2):339–367, 2017.

[2] L. Andres, W. Salas, and D. Skole. Fourier analysis of multi-temporal AVHRR data applied to a land cover classification. *Int. J. Remote Sens.*, 15(5):1115–1121, 1994.

[3] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Discov.*, 31(3):606–660, 2017.

[4] L. Bruzzone and D. Prieto. Automatic analysis of the difference image for unsupervised change detection. *IEEE Trans. Geosci. Remote. Sens.*, 38(3):1171–1182, May 2000.

[5] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin. Digital change detection methods in ecosystem monitoring: A review. *Int. J. Remote Sens.*, 25(5):1565–1596, 2004.

[6] N. Di Mauro, A. Vergari, T. M. A. Basile, F. G. Ventola, and F. Esposito. End-to-end learning of deep spatio-temporal representations for satellite image time series classification. In *DC@ECML/PKDD, Procs.*, pages 1–8, 2017.

[7] S. W. Golomb. Run-length encodings. *IEEE Trans. Inf. Theory*, 12:399–401, 1966.

[8] D. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel. Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geosci. Remote Sens. Lett.*, 14(10):1685–1689, 2017.

[9] J. Inglada, A. Vincent, M. Arias, B. Tardy, D. Morin, and I. Rodes. Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sens.*, 9(1), 2017.

[10] J. R. Jensen. Urban change detection mapping using Landsat digital data. *Cartogr. Geogr. Inf. Sci.*, 8(21):127–147, 1981.

[11] R. Johnson and E. Kasischke. Change vector analysis: A technique for the multispectral monitoring of land cover and condition. *Int. J. Remote Sens.*, 19(16):411–426, 1998.

[12] C. Kurtz, N. Passat, P. Gançarski, and A. Puissant. Multiresolution region-based clustering for urban analysis. *Int. J. Remote Sens.*, 31(22):5941–5973, 2010.

[13] F. Petitjean, J. Inglada, and P. Gançarski. Satellite image time series analysis under time warping. *IEEE Trans. Geosci. Remote. Sens.*, 50(8):3081–3095, 2012.

[14] F. Petitjean, C. Kurtz, N. Passat, and P. Gançarski. Spatio-temporal reasoning for the classification of satellite image time series. *Patt. Rec. Lett.*, 33(13):1805–1815, 2012.

[15] C. Ratanamahatana, E. Keogh, A. J. Bagnall, and S. Lonardi. A novel bit level time series representation with implication of similarity search and clustering. In *PAKDD, Procs.*, pages 771–777, 2005.

[16] P. Ravikumar and V. S. Devi. Weighted feature-based classification of time series data. In *CIDM, Procs.*, pages 222–228, 2014.

[17] C. Senf, P. J. Leitão, D. Pflugmacher, S. van der Linden, and P. Hostert. Mapping land cover in complex mediterranean landscapes using landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. *Remote Sens. Environ.*, 156:527–536, 2015.

[18] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.*, 114(1):106–115, 2010.