EXTRACTION OF FREQUENT GROUPED SEQUENTIAL PATTERNS FROM SATELLITE IMAGE TIME SERIES

A. Julea[1,2,3], N. Méger[1], C. Rigotti[4], M-P. Doin[5], C. Lasserre [6], E. Trouvé[1], P. Bolon[1] and V. Lăzărescu[3]

[1]Université de Savoie, Polytech'Savoie, LISTIC Laboratory
BP 80439, F-74944 Annecy-le-Vieux Cedex, France
{andreea.julea|nicolas.meger|emmanuel.trouve|philippe.bolon}@univ-savoie.fr
[2]Institutul de Stiinte Spatiale, Bucharest, Romania
[3]Politehnica University of Bucharest, Romania
Faculty for Electronics, Telecommunications and Information Technology
Applied Electronics and Information Engineering Departement
vl@elia.pub.ro
[4]Université de Lyon, CNRS
INSA-Lyon, LIRIS, UMR5205, F-69621, France
christophe.rigotti@insa-lyon.fr
[5] Ecole Normale Supérieure, Laboratoire de Géologie, UMR 8538
24 rue Lhomond, F-75231 Paris CEDEX 05, France
doin@geologie.ens.fr
[6] Laboratoire de Géophysique Interne et Tectonophysique, CNRS
Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9
cecile.lasserre@ujf-grenoble.fr

1. INTRODUCTION AND RELATED WORK

Remote sensing techniques provide end-users with ever growing volumes of data. Indeed, the resolution of acquisitions is continually improved while the number of available channels also increases. In addition, acquisition rates have been boosted during the last few years. It is thus possible to gather loads of images concerning a given geographical zone. This kind of dataset is termed as a Satellite Image Time Series (SITS). The analysis of SITS raises new challenges as data volumes to be processed are huge and as both the temporal and the spatial dimensions have to be taken into account. Various techniques aiming at characterizing evolutions in SITS have been proposed. Some of those techniques explore the data at the region level, more precisely, they extract regions from all the images so as to provide end-users with the evolutions of these regions (e.g., [1]). Other techniques link descriptors to each image of the SITS. A time sequence of descriptors is thus build and sub-evolutions that match temporal and frequency constraints are output as the result (e.g., [2]). Pixel-based techniques have also been proposed, focusing either on specific evolution occurring at some time stamp, i.e., pixel change detection techniques (e.g., [3], [4]) or on the characterization of the whole sequence of pixel values and not of the sub-evolutions (e.g., synthetic channels-based techniques as proposed in [5] or clustering techniques as detailed in [6]). It is to notice that change detection techniques also work at the object/region level (but still needing assumptions about the type evolutions). Though similar to our approach, in the sense that generally both temporal and spatial dimensions are taken into account, none of these techniques can extract sets of grouped pixels sharing a same evolution or sub-evolution without first extracting objects/regions (e.g., [1, 2]) and/or without making any assumption about the type of evolution. For example, change detection techniques look for specific change classes while other pixel-based techniques only consider full evolutions and not sub-evolutions (e.g., [5, 6]). Furthermore, when

The authors wish to thank the French Research Agency (ANR) for supporting this work through the EFIDIR project (ANR-2007-MCDC0-04, http://www.efidir.fr). They also thank the ADAM project and the CNES agency for making data available. Finally, the authors express their gratitude to Roxana Vintila (Research Institute for Soil Science and Agrochemistry - Bucharest, Romania) and to Gheorghe Petcu (National Agricultural Research and Development Institute Fundulea, Romania) for supplying the ground truth of the regions that we studied through acquisitions of the ADAM project.

searching for sub-evolutions, we aim at extracting them without giving any priority to any date of acquisition, which prevents us from using clustering techniques.

2. FGS-PATTERNS EXTRACTION

This paper presents an alternative and complementary approach, relying on sub-evolutions identification at the pixel level, for finding, in an unsupervised way, groups of pixels that could be of interest for end-users. Moreover, the experiments reported in this paper show that large datasets (e.g., 20 images of $1,000,000$ pixels each) can be handled by such an approach. In order to output pixel sets that make sense both spatially and temporally, we select sets of size greater than a given threshold and whose pixels are spatially connected enough and have the same temporal evolution or sub-evolution. No assumption about temporal evolutions is made beforehand. Since we have to face a huge amount of data and a high number of potential temporal evolutions, a novel approach based on data mining techniques is used. More precisely, we adapt frequent sequential patterns extraction to a spatio-temporal context by considering a SITS as a set of temporal and symbolic sequences, each sequence describing the (sub-)evolution of a given pixel. In [7], we presented a preliminary approach to the one of this paper, but it did not take into account the spatial grouping tendencies of the pixels. It was thus possible to extract a (sub-)evolution that holds for a lot of pixels that are not connected to each other. As a consequence, the (sub-)evolutions that are provided to the end-user could be difficult to interpret. In this paper, pixels sharing a same (sub-)evolution, beside covering a minimum surface, also have to exceed a minimum connectivity threshold. The connectivity measure that is proposed relies on the 8-connectivity contour convention, and enables for search space pruning (based on the so-called *anti-monotonicity* property), which reduces by orders of magnitude execution times and memory consumption. Evolutions and sub-evolutions that are retained are termed as *Frequent Grouped Sequential patterns* (fgs-patterns) and pixels for which a fgs-pattern holds are said to be *covered* by this pattern. A fgs-pattern is of the form: $A_1 \to A_2 \to \ldots \to A_n$, where $A_1, A_2, \ldots, A_n$ are symbols representing discrete pixel states. For instance, consider a pattern of size 3 denoted $A_1 \to A_2 \to A_3$. Then, a pixel p is covered by $A_1 \to A_2 \to A_3$ if p is in state A_1 in an image of the SITS, and in state A_2 in one of the following images (not necessarily the next one), and later in state A_3 in another image. This fgs-pattern $A_1 \to A_2 \to A_3$ is output, when the extraction engine is run on a SITS, if a sufficient number of pixels satisfy to this temporal (sub-)evolution and to a spatial grouping tendency.

3. EXPERIMENTS

This method is complementary to the existing techniques. In practice, it turns out to be effective in finding interesting groups of pixels, sharing meaningful common temporal evolutions, and that would not be exhibited by other approaches. We report experiments on two real datasets. Firstly, on the ADAM (Data Assimilation for Agro-Modeling) [8] dataset, that is a SPOT (Satellite Pour l'Observation de la Terre) SITS containing images covering a rural zone in South Romania, near Bucharest. And secondly, we present experiments on interferograms covering the lake Mead on the Colorado river in the United States. These interferograms have been computed using Synthetic Aperture Radar (SAR) images provided by the ERS satellites.

For the experiments on the ADAM dataset, we selected 20 images between October 2000 and July 2001 containing 1000*1000 pixels each. These images have been captured via three bands by SPOT satellites: B1 in green (0.5 - 0.59 μm), B2 in red (0.61 - 0.68 μm) and B3 in near infrared (NIR 0.78 - 0.89 μm). Their resolution is 20mx20m. For each pixel, and for each date, we consider a synthetic band B4. B4 is established by calculating the *Normalized Difference Vegetation Index* (NDVI) [9] using bands B2 and B3. B4 is thus defined as $B4 = \frac{B3-B2}{B3+B2}$. The NDVI index is widely used for detecting live green plant canopies in multispectral remote sensing data and it partly copes with the basic thematic of the ADAM SITS which is an agricultural one. Having at disposal the ground truth for the fields that belong to the Romanian National Agricultural Research and Development Institute (5.9% of the scene), we were able to evaluate our results. Parameters were set so that extractions times do not exceed 1000 seconds and so that end-users are not overwhelmed by too many fgs-patterns. In the end, no more than 100 fgs-pattern were exhibited. Among them, we found short ones such as $1 \to 1 \to 3 \to 3$. This pattern points out the first part of phenological cycles, meaning that some cultures are seeded, rise and reach maturation (e.g., spring time seeded cultures: corn, pea, chick-pea and Sudan grass). If this pattern is localized, i.e., if all pixel covered by it are enlightened while the others are set to black, then we get the image of Figure 1, for the area where the ground truth is available. Quite homogeneous geometric regions with crisp boundaries appear. White regions correspond to different types of agricultural fields with spring cultures, while black regions correspond to forests and to other types of agricultural fields.

Fig. 1: Localization of the short fgs-pattern $1 \rightarrow 1 \rightarrow 3 \rightarrow 3$.

Among longer patterns, we found $1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 3 \rightarrow 3$. The spatial localization of this pattern and its temporal discrimination are presented in Figure 2. Each colored pixel is a covered one, and each color presents a given distribution of occurrence dates of the pattern. Pixels that are colored in light blue refer to occurrences of the pattern that start at early acquisition dates. They correspond to regions with Sudan grass and pea cultures. Red and dark blue pixels correspond to occurrence starting later, and exhibit corn cultures.

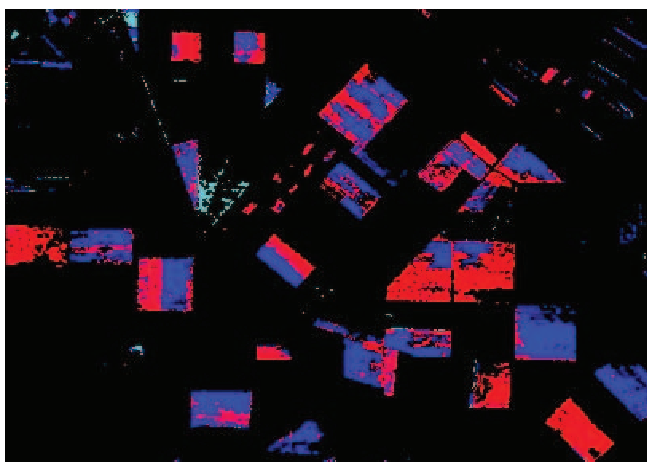

Fig. 2: Localization of pattern $1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 3 \rightarrow 3$ with its temporal discrimination.

For a given distribution of occurrence dates, the purity of the corresponding colored area is the percentage of pixels of this areas that correspond to the main culture of the area (obtained using the ground truth). Then, the purity of the pattern itself is simply the weighted average of the purity of its colored areas (the weight of an area is simply the number of pixels in this area). According to this definition, the purity of the previously presented pattern is 91.13%. Moreover, this pattern is rather general since it covers 66.8% of the pixels that correspond to corn cultures in the ground truth.

The second dataset corresponds to interferograms and cover the lake Mead area, where the soil surface around the lake is affected by a subsidence/uplift motion that is correlated with water level fluctuations. We selected a subset of 20 interferograms obtained from images acquired between 1996 and 2008. Each interferogram gives the interferometric phase difference of its acquisition date relative to the master date 1995-10-08. The atmospheric phase screen of the master image is assigned to the master date. Between 1996 and 1998, the lake water level increases while it drops between 2000 and 2008. The analyzed images (759x716 pixels, 130m$\times$130m resolution) contain phase delays due to both atmospheric and deformation patterns. We show that it is possible to extract long fgs-patterns, having 15 labels, that are correlated to water level fluctuations. Indeed, their localization well corresponds to zones where ground deformation has been identified. Moreover, though atmospheric perturbations were present, none of these patterns reported them, which demonstrates the ability of fgs-patterns to discard such random phenomena.

4. REFERENCES

[1] P. Héas and M. Datcu, "Modeling trajectory of dynamic clusters in image time-series for spatio-temporal reasoning," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 7, pp. 1635– 1647, 2005.

[2] R. Honda and O. Konishi, "Temporal rule discovery for time-series satellite images and integration with RDB," in *PKDD '01: Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery*, London, UK, 2001, pp. 204–215, Springer-Verlag.

[3] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Digital change detection methods in ecosystem monitoring: a review," vol. 25, no. 9, pp. 1565–1596, May 2004.

[4] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Intl. J. of Remote Sensing*, vol. 25, no. 12, pp. 2365–2407, June 2004.

[5] E. Nezry, G. Genovese, G. Solaas, and S. Rémondière, "ERS - Based early estimation of crop areas in Europe during winter 1994-95," in *ERS Application, Proceedings of the Second International Workshop held 6-8 December 1995 in London*, Guyenne T.-D., Ed., 1996, vol. 383 of *ESA Special Publication*, p. 13.

[6] A. Ketterlin and P. Gançarski, "Sequence similarity and multi-date image segmentation," in *4th Intl Workshop on the Analysis of Multitemporal Remote Sensing Images*, Leuven, Belgique, july 2007.

[7] A. Julea, N. Meger, E. Trouve, and P. Bolon, "On extracting evolutions from satellite image time series," in *Proc. of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2008)*, Boston, MA, USA, 2008, vol. 5, pp. 228–231.

[8] Centre National d'Etudes Spatiales, "Database for the Data Assimilation for Agro-Modeling (ADAM) project," online, http://kalideos.cnes.fr/index.php?id=accueil-adam.

[9] T.M. Lillesand and R.W. Kiefer, *Remote Sensing and Image Interpretation*, John Wiley and Sons, New York, fourth edition, 2000.