

SITS-P2miner: Pattern-Based Mining of Satellite Image Time Series

Tuan Nguyen^{1(✉)}, Nicolas Méger¹, Christophe Rigotti², Catherine Pothier³,
and Rémi Andreoli⁴

¹ Université Savoie Mont Blanc, Polytech Annecy-Chambéry, LISTIC,
74944 Annecy-le-vieux, France

{hoang-viet-tuan.nguyen,nicolas.meger}@univ-smb.fr

² Univ Lyon, INSA-Lyon, CNRS, Inria, LIRIS, UMR5205,
69621 Villeurbanne, France

³ Univ Lyon, INSA-Lyon, SMS-ID, 69621 Villeurbanne, France

⁴ Bluecham S.A.S., 98807 Nouméa, Nouvelle-Calédonie, France

Abstract. This paper presents a mining system for extracting patterns from Satellite Image Time Series. This system is a fully-fledged tool comprising four main modules for pre-processing, pattern extraction, pattern ranking and pattern visualization. It is based on the extraction of grouped frequent sequential patterns and on swap randomization.

1 Introduction

A Satellite Image Time Series (SITS) is a series of images covering a same area acquired by satellites over time. SITS analysis is a still growing research field, stimulated by the enhancement of the spatial resolution, the reduction of the time intervals between acquisitions and the development of new acquisition modes. Considering the large volume and the raw nature of such SITS, it is not possible to process them manually, and unsupervised mining techniques demonstrate their potential to describe and discover spatiotemporal phenomena in SITS. These techniques rely either on global models such as clustering (e.g. [2]) or on local patterns such as sequential patterns (e.g. [3]).

This paper presents *SITS-P2miner* (Pattern maP miner), a system that implements the pattern mining method introduced in [4] and the swap randomization ranking presented in [5], together with appropriated pre-processing and visualization tools. The salient features of the resulting system, with respect to other state-of-the-art methods, are: (1) its ability to process both optical and radar satellite images; and (2) its robustness against frequent quality degradation sources inherent to satellite images (atmospheric perturbations, missing values, sensor defects, irregular time spacing).

In a SITS, the covered area is represented as a grid of pixels, and, for each pixel, the SITS contains the sequence of values (integers or floating point numbers) acquired over time for that location. In *SITS-P2miner*, the input SITS is quantized in a pre-processing step to replace pixel values by symbols denoting discrete levels (1, 2, 3, ...). This symbolic SITS is then mined to extract GFS-patterns [4], where a GFS-pattern is a sequential pattern [1] satisfying the two

following constraints. First, it must occur in a sufficient number of sequences (being frequent, in the usual sense). Secondly, the occurrences of the pattern have to be somehow coherent over space, i.e., if the pattern occurs in the sequence of values of a pixel, then it must also tend to occur in the spatial neighborhood of this pixel (but eventually with a shift in time).

2 System Description

The architecture of the system is presented in Fig. 1. As an input, it takes a SITS expressed as a single synthetic band of interest such as the ground motion magnitude in the line of sight of a radar satellite, or a vegetation index when dealing with optical images. First, the SITS is quantized by the *pre-processing* module, using one of the different available discretization strategies, to produce a *Symbolic SITS*. This symbolic SITS is in turn processed by the *pattern extraction* module to extract *maximal GFS-patterns*. These patterns are then assessed by swap randomization of the symbolic SITS and ranked using a normalized mutual information measure reflecting the impact of the randomization upon the pattern occurrences. Finally, different *maps* depicting the location in space and time of the occurrences of the top-ranked patterns are computed by the *visualization* module. The reader is referred to [4, 5] for the complete definition of the patterns, the description of the extraction/ranking steps and the guidelines for parameter settings.

The whole process is driven by a single human readable parameter file. The output is stored in folders whose hierarchy is structured according to the processing steps, the parameter values and the execution time stamps. This output includes maps, patterns, intermediate ranking information as well as monitoring logfiles that are organized for quick result browsing and easy iterative mining.

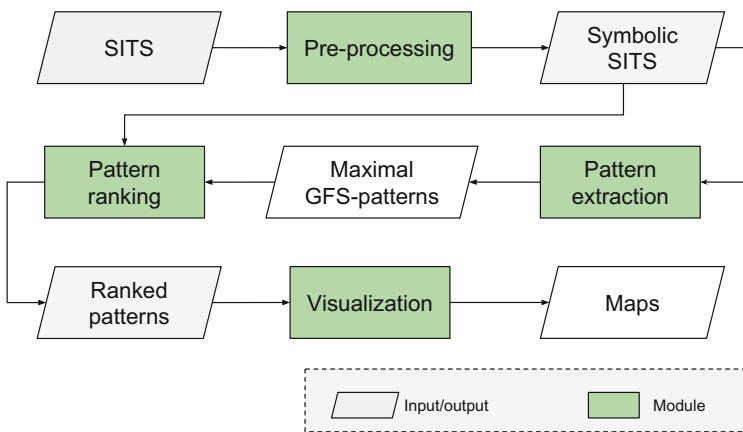


Fig. 1. System architecture.

The most resource consuming steps are the GFS-pattern extraction and the swap randomization. Therefore, the corresponding modules are implemented in C. The other ones are implemented in Python. All modules are chained in Python, which allows to add new modules simply. The system can be run on *Windows*, *Linux* and *Mac OS X* operating systems using a standard computing platform (e.g., single core on 2.7 GHz Intel Core i7, 8 GB memory).

3 Demonstration

During the demonstration, we will present the analysis of two real SITS. The first one (provided by Marie-Pierre Doin, ISTerre lab., CNRS), is an ENVISAT-based SITS covering Mount Etna (16 radar images 598×553 from 2003 to 2010). In this series, the effects of the stratified atmosphere have been corrected, but not the ones due to the turbulent atmosphere. The pixel values give ground motion magnitudes in the satellite line of sight. The other series is a Landsat 7 SITS (16 optical images 513×513 from 2004 to 2011) covering the area of Yat e in New Caledonia and containing values expressing the presence/absence of vegetation (NDVI index). The limited size of these series allows for live computation of the maps during the demonstration on a standard laptop. Four images of the second series are shown in Fig. 2, illustrating typical problems of satellite data such as missing values, artifacts, sensor defects, presence of clouds, etc.

Two examples of maps of occurrences of GFS-patterns, selected among the best ranked maps found in these series, are shown in Fig. 3. The colored pixels denote the locations of the occurrences in space while the colors correspond to the ending dates of the occurrences (middle of the SITS in blue and end of the SITS in pink). The map of Fig. 3a corresponds to GFS-pattern 1-2-2-2-2-2-3-3 over the Mount Etna motion series. It sketches a trend from low magnitude motion (symbol 1) to high magnitude motion (symbol 3). The upper part of the map exhibits a moving part of the volcano flank and the lower part unveils a fault system. Figure 3b shows the map of GFS-pattern 2-2-3-2-2-2-3 over the New Caledonia vegetation series. It denotes a cycling variation from normal vegetation index (symbol 2) to high vegetation index (symbol 3) and corresponds to the presence of maquis (evergreen vegetation). The best ranked maps over the New Caledonia vegetation series highlight various phenomena related not only to the vegetation

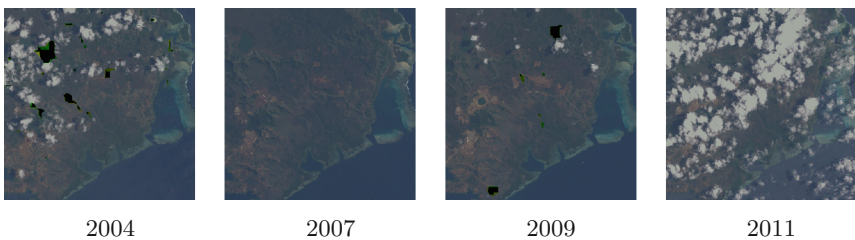


Fig. 2. Landsat 7 images (RGB color space). (Color figure online)

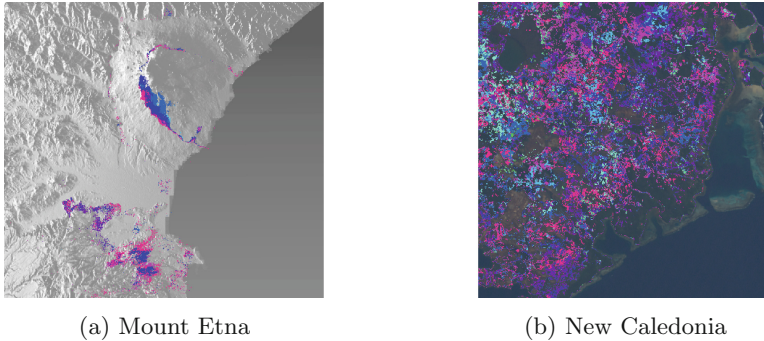


Fig. 3. Examples of top-ranked maps of GFS-patterns. (Color figure online)

but also to anthropic activities, and have been integrated as data layers in the *Qëhnelö environmental management platform* (<http://www.yate.nc/>).

The system used in this demonstration is available at: https://www.polytech.univ-savoie.fr/fileadmin/polytech_autres_sites/sites/listic/projets/sitsmining/SITSP2MINER.zip

Acknowledgments. Funding for this project was provided by a grant from la Région Rhône-Alpes (Tuan Nguyen’s grant). Other support: C. Rigotti and C. Pothier are members of LabEx IMU (ANR-10-LABX-0088).

References

1. Agrawal, R., Srikant, R.: Mining sequential patterns. In: Proceedings of Conference ICDE, pp. 3–14 (1995)
2. Gueguen, L., Datcu, M.: Image time-series data mining based on the information-bottleneck principle. *IEEE Trans. Geosci. Remote Sens.* **45**(4), 827–838 (2007)
3. Honda, R., Konishi, O.: Temporal rule discovery for time-series satellite images and integration with RDB. In: Siebes, A., De Raedt, L. (eds.) PKDD 2001. LNCS (LNAI), vol. 2168, pp. 204–215. Springer, Heidelberg (2001)
4. Julea, A., Méger, N., Bolon, P., Rigotti, C., Doin, M.P., Lasserre, C., Trouvé, E., Lăzărescu, V.N.: Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns. *IEEE Trans. Geosci. Remote Sens.* **49**(4), 1417–1430 (2011)
5. Méger, N., Rigotti, C., Pothier, C.: Swap randomization of bases of sequences for mining satellite image times series. In: Appice, A., Rodrigues, P.P., Santos Costa, V., Gama, J., Jorge, A., Soares, C. (eds.) ECML PKDD 2015. LNCS, vol. 9285, pp. 190–205. Springer, Heidelberg (2015)