ABSTRACT

This paper presents a data mining approach for describing Satellite Image Time Series (SITS) spatially and temporally. It relies on pixel-based evolution and sub-evolution extraction. These evolutions, namely the frequent grouped sequential patterns, are required to cover a minimum surface and to affect pixels that are sufficiently connected. These spatial constraints are actively used to face large data volumes and to select evolutions making sense for end-users. In this paper, a specific application to fully polarimetric SAR image time series is presented. Experiments performed on a RADARSAT-2 SITS covering the Chamonix Mont-Blanc test-site are used to illustrate the proposed approach.

Index Terms— Satellite Image Times Series (SITS), Data mining, Polarimetric SAR images, Frequent grouped sequential patterns.

1. INTRODUCTION

Satellite Image Time Series (SITS) are of high interest as they capture both spatial and temporal information. Numerous works aim at exhibiting this information so that end-users can interpret these data without having to browse the whole dataset. Some techniques provide end-users with the evolutions of the regions that are identified in each one of the images (e.g., [1]). Other techniques link descriptors to each image of the SITS and uncover evolutions and sub-evolutions of these descriptors that match temporal and frequency constraints (e.g., [2]). Pixel-based techniques have also been proposed, focusing either on a specific type of evolution occurring at some timestamp, i.e. pixel change detection techniques (e.g., [3]) or on the characterization of the whole sequence of pixel values (e.g., synthetic channels-based techniques or clustering techniques as detailed in [4]). Change detection techniques also work at the object/region level but they still need assumptions about the type of the evolutions. Though similar to our approach, in the sense that generally both the temporal and the 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 features (e.g., [1], [2]) and/or without making any assumption about the type of evolution (e.g., [3]). Furthermore, when searching for sub-evolutions, we wish to extract them without giving any priority to any date of acquisition, which prevents us from using clustering techniques. In [5], frequent sequential patterns are used to express spatiotemporal relations. It allows to retrieve all evolutions and sub-evolutions, but it requires end-users to set both temporal and spatial constraints. In this paper, a sequential pattern-based approach is used without making any assumption about the temporal information. This approach is detailed in [6]. Thanks to the all-weather
capability of SAR sensors, large time series of radar images have been acquired for 20 years all around the world. The main applications of SAR SITS are vegetation and crop monitoring, change detection [BBM05], especially after disasters, and crustal deformation measurements (faults, volcanoes, urban subsidence...) by interferometric techniques. Since the launch of ALOS, RADARSAT-2 and Terrasat-X satellites in 2006 and 2007 respectively, fully polarimetric space-borne SAR data are acquired for different applications. PolSAR image time series represent large volumes which are even more difficult to handle since 4 different axes have to be explored simultaneously: the 2 spatial directions of the imaged area, the polarimetric direction to analyze the backscattering mechanisms and the time direction to extract the temporal information. In this paper, the proposed data mining SITS analysis technique is applied to a series of PolSAR images. The proposed processing chain consists in using first the Cloude and Pottier decomposition [7] to extract discriminating features such as the entropy H and _ angles and to perform a 2D quantification according to the usual partitioning of the H alpha plane into 9 zones. The resulting symbolic description of the different dates is mined by the proposed approach to detect frequent grouped temporal evolutions.

2. FGS-PATTERNS

A pixel evolution or sub-evolution is described using a sequential pattern, denoted $A_1 \rightarrow A_2 \rightarrow \ldots \rightarrow A_n$, where $A_1, A_2, \ldots, A_n$ are symbols representing discrete pixel states at $n$ different dates which are not necessarily consecutive. These patterns were initially proposed in [8] to mine sequences of commercial transactions. We intend to use these pixel evolutions and subevolutions, to find, in an unsupervised way, groups of pixels that could be of interest for end-users. In order to output pixel sets making sense spatially and temporally, sets having at least $\sigma$ pixels (i.e. a minimum surface) sharing the same temporal evolution $\gamma$ are selected. Pixels sharing $\gamma$ are said to be covered by $\gamma$ and are denoted $\text{cov}(\gamma)$. Furthermore, these same pixels are also required to exceed a minimum connectivity threshold $\kappa$. The used connectivity measure is called the average connectivity. It gives, for the pixels sharing $\gamma$, the average number of neighbor pixels also sharing $\gamma$. The 8 nearest neighbors (8-NN) are taken into consideration. Let us consider a local connectivity function $\text{LC} ((x, y), \gamma)$ that returns, for a pixel $(x, y)$, the number of neighbors covered by $\gamma$. The average connectivity of $\gamma$ is defined as: $\text{AC}(\gamma) = \frac{\sum_{(x,y) \in \text{cov}(\gamma)} \text{LC} ((x, y), \gamma)}{\text{lcov}(\gamma)}$. Formally, an evolution (or sub-evolution) $\gamma$ is thus retained if $\text{lcov}(\gamma) \geq \sigma$ and if $\text{AC}(\gamma) \geq \kappa$. In this case, it is called a Frequent Grouped Sequential Pattern (FGSPattern). These constraints can be actively pushed into data mining algorithms (e.g., [8], [9]) to prune the search space and to make extractions tractable. More details can be found in [10]. Our prototype, based on the PrefixGrowth algorithm [9], has been written in C language using our own data structures.

3. EXPERIMENTS

In [6], this approach has been applied to different SITS, related to optical data covering an agricultural area and to interferometric SAR data: SITS from ERS and ENVISAT differential interferograms covering the lake Mead area (USA) for crustal deformation monitoring and for discarding random phenomena such as atmospheric perturbations [10]. More recently, a PolSAR SITS from RADARSAT-2 satellite (4 image acquisitions) has been analyzed and this paper is intended to extend the processing for a larger SITS composed by 7 fully polarimetric images acquired from winter 2009, on January 29 to summer 2009, June 22 covering a high mountain area: the Chamonix Mont-Blanc test-site of the EFIDIR project. PolSAR images may provide different measurements based on the radar specificity: snow and ice penetration which yields to subsurface observation and polarized coherent waves which allow the backscattering media to be analyzed. However, polarimetric SAR acquisitions provide multi-channel complex images which require a heavy processing chain, especially in the context of Alpine glaciers with high relief, fast motion and surface changes and a rather unknown SAR signature of the ice/snow/rock mixture. The processing chain can be divided into two main stages: a pre-processing stage to obtain higher level information than the initial SAR data, that can be performed by conventional methods now available in distributed software and an information mining stage which can be performed by different approaches depending on the application requirements. Several supervised or unsupervised classification techniques based on the analysis of the PolSAR feature images have been proposed. The proposed approach is based on similar initial classification rules, complemented by a data mining exploration technique to analyze multi-temporal PolSAR data and to introduce a contextual constraint. The first pre-processing step consists in transforming the complex channels into 3 channels (in monostatic configuration $S_{IV} \approx S_{VH}$) expressed in the Pauli basis [7] which is usually preferred to obtain a coherent scattering vector $[k]$ closer to the physical phenomena of wave scattering. Figure 1 illustrates the 3 amplitudes in the Pauli basis of a RADARSAT-2 image acquired over Chamonix Mont-Blanc. This sub-image includes 3 glaciers (Argentière, Taléfre and Mer-de-Glace), the Chamonix valley at about 1000 m high (left down) and mountains up to 4000 m.
Except for a coherent target, the speckle phenomenon which affects the distributed targets makes it difficult to work directly with the scattering vectors. The second preprocessing step consists in estimating the 3x3 Hermitian positive semidefinite coherency matrix. In this experiment, the conventional Lee filter is applied to reduce the speckle noise effect on the coherency matrix. The third preprocessing step consists in applying the Cloude and Pottier decomposition [7]. Several PolSAR features are usually derived from this decomposition to discriminate the different backscattering mechanisms. The H and $\alpha$ PolSAR features indicate the random behavior of the global scattering and the mean scattering mechanism from surface to double bounce scattering. They are strongly related to the geophysical properties of the ground target area providing reliable information for further classification. In [7], nine clustering zones are proposed to describe the H and $\alpha$ plane. The boundaries of such crisp predefined clustering illustrated in Fig. 2-(a) are often used as references to interpret the attribute spaces in terms of backscattering mechanisms or to initialize clustering techniques as in the so-called “Whishart classifier” [11]. To follow a conventional PolSAR preprocessing chain, the proposed data mining approach uses the (H, $\alpha$) features as input information. The 2D distribution of these features over the Mont-Blanc area is illustrated in Fig. 2-(b). The “temporal” information carried by these features can be observed by combining 3 different dates in a RGB color composition in Figs. 2-(c) and 2-(d). The results show that significant changes occurred between the 3 dates, and that the evolutions are related to the different parts of the images: the glaciers, the mountains and the Chamonix valley appear with different colors, which also vary with the height or slope orientation.

4. INFORMATION MINING STAGE

The proposed data mining technique requires the input sequences of pixel values to be transformed into sequences of symbols. The quantization strategy must preserve the useful information. In the case of the PolSAR data, the two input channels for each date are the entropy value in [0, 1] and the $\alpha$ angle value in [0, 90°]. Both values could be quantified into a few intervals, resulting into a regular partitioning of the H – $\alpha$ space. Instead of this general quantification, the proposed approach consists in using the H-$\alpha$ space partitioning illustrated in Fig. 2-(a). This symbolic representation provided by the PolSAR domain makes the interpretation of the output FGS-patterns easier for the end-user. The experiments have been run on the 7 RADARSAT-2 images available on the Chamonix Mont-Blanc test site by first applying the preprocessing chain presented in Section 3. Then the resulting H and $\alpha$ images (7 pairs of 2048 x 2048-pixel images) are mined by coding the pixel positions according to the 9 zones and searching for FGS-patterns. 46 different FGS-patterns come out with a minimum surface $\sigma$ set to 4000 and a minimum average connectivity threshold $\kappa$ set to 4. Among them, 21 are maximal and 4 FGS-patterns retain our attention: (a) $6 \rightarrow 6 \rightarrow 6 \rightarrow 6 \rightarrow 6$ which often appears on the glacier lower parts and on the surrounding moraine; (b) $5 \rightarrow 6 \rightarrow 6 \rightarrow 6 \rightarrow 6$ and $5 \rightarrow 5 \rightarrow 6$ zones 5 and 6 which is related to the glacier lower
parts; (c) 4→4 which seems to be complementary to the previous pattern by appearing on the glacier upper parts; (d) 6→9→9→9 which often appears on the fold-over sides of the mountains. The two first FGS-patterns are illustrated in Fig. 3. Pixels where the FGS-pattern occurs appear with a color which depends on its occurrence dates. For example, the 6→6→6→6→6 pattern displayed in Fig. 3-(a) appears most of the time with the same color which corresponds to the presence of zone 6 in images 1, 2, 3, 4 and 5. This visualization allows the end-user to observe for each pattern where and when it appears. If the same dates are involved, it provides a temporal localization of the events. If the pattern is obtained with the contribution of different groups of dates, it reveals a progressive evolution which may occur not at the same time in different places. This also allows noisy values breaking the pattern to be discarded when other dates can provide the missing values in the same order. Fig. 3-(b) gives the spatio-temporal localization of pattern 5→6→6→6. As it can be observed, most of the pixels are simultaneously affected by this pattern. The change from zone 5 to zone 6 can be associated with surface state change due to snow melting between winter and summer seasons over Chamonix Mont-Blanc area.

Regarding the information mining stage, our research will focus on FGS-pattern post analysis methods to provide a single clustering of the whole SITS associated to thematic classes.

6. REFERENCES


This work was supported by the French Research Agency (ANR) through the EFIDIR project (ANR-2007-MCDOC0-04, http://www.efidir.fr) and FOSTER project (ANR-2010-COSI-012-04). The authors wish to thank the Canadian Space Agency (CSA) for the Radarsat-2 SAR data (SOAR project # 3399) over the Mont-Blanc test-site and the cooperation project Brazil – France CAPES COFECUB.