Mining Image Time-Series

Patrick Heas[1,2]
[1]IRIT, Toulouse, France
Patrick.Heas@enseeiht.fr

Mihai Datcu[2], Alain Giros[3]
[2]German Aerospace Center, DLR Oberpfaffenhofen, Wessling, Germany
[3]Centre National d'Etudes Spatiales CNES, Toulouse, France

I. Introduction

The process of searching and analyzing data in order to discover potentially useful information is a crucial matter when dealing with large databases. Considering the huge amount of data collected by satellite observation systems, opportunities to generate multispectral image time-series are increasing. Exploratory methods are needed to understand the dynamics of Earth observation scenes and of their objects.
In this paper, information mining methods are proposed. They first rely on a hierarchical model based information representation. Entropic measurement similarity in a modeled dynamic feature space characterizing the dynamics of objects and of image structures are then researched.

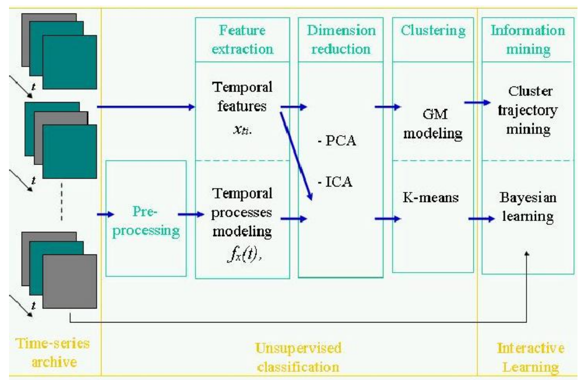

Fig. 1. *Synergy of methods for exploration.*

II. Definition of a concept for spatio-temporal information mining

A hierarchical model based information representation must constitute the fundments of a concept designed for dynamic scene understanding. Data modeling is a wide field of investigations. Thanks to an information theory approach [4], selection between models and parameter estimation are efficient. Unsupervised model inferences are performed using Minimum Description Length (MDL) principle, model order and parameters are inferred using a maximum likelihood criterion. Supervised Bayesian learning infers on the features and models.
The exploration procedure is splitted into different processing levels. The general synergy is described in figure 1.
Once the data has been radiometrically calibrated and geometrically registered, the first processing level is feature extraction. According to assumptions on the stochastic processes, model parameters are learnt from the data. They form temporal and spatio-temporal features characterizing the spectral and textural pixel located stochastic processes.
Image time-series have different representations in a multidimensional space comprising two spatial axes, a time axis and the several feature dimensions. The spatio-temporal representation is simply the time-series of images. The multitemporal feature space representation is a multidimensional space composed by the union of all the time localized feature components and for which the spatial index is hidden.
The two following levels of processing perform the analysis of the multitemporal feature space. It can be decomposed into the following two steps:
1. According to criterion of decorrelation or independence, dimension reduction infers interesting projections of the multidimensional space.
2. The features can then be grouped by similarities. Clusterings are performed assuming that the feature space projections have Gaussian mixture (GM) probability distribution functions (PDF). The parameters of the mixture are learnt from the data using the MDL principle. Paralally, in the spatio-temporal representation, classifications are generated.

Finally, Information mining can be performed in the multitemporal feature space (visual information mining) or/and in the spatio-temporal representation (image information mining).
Image information mining is driven by the interaction of a user through a graphical user interface (GUI). Bayesian learning enables a user to visualize, for each pixel location, the probability of a defined semantic conditioned by the data decomposed into several classifications [1].
Visual information mining enables a user to mine other data representations and enhances the data understanding. Exploring a graph describing cluster trajectories allows the user to access hidden information characterizing spatio-temporal patterns such as their trajectory parameters and their interactions with other trajectories. By exploiting the hierarchical modeling previously described, the dynamic

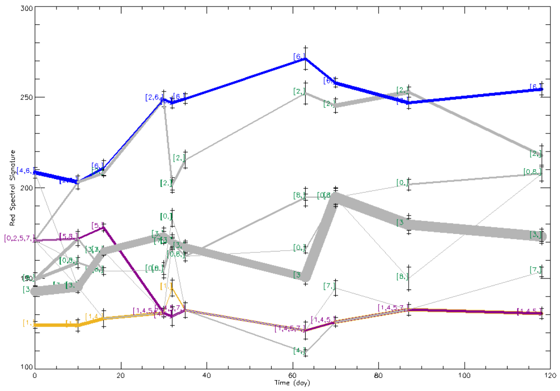

Fig. 2. *Example of graph characterizing the dynamic clusters.*

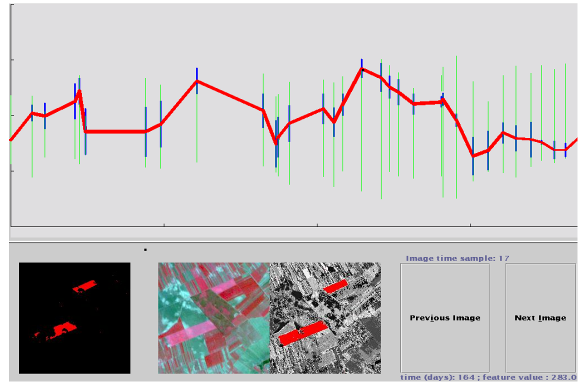

Fig. 3. *GUI for the exploration of the graph : The colza crops associated cluster most probable trajectory is plotted above in red on a period of 265 days; it is a 1-dimensional projection of the graph on the red spectral band; the variance evolution appears in blue, and in green, the other multitemporal object variance evolutions are plotted. Below the correspondent multitemporal class (in red), the original image sample at a given time and the associated time-localized class(in red) are displayed from left to right. The increasing gray level values in the time-localized classification are significant of a decreasing distance with the projected multitemporal class.*

feature space can be modeled, and hence, cluster trajectories can be characterized [2]. Indeed, the multitemporal feature space containing the global spatio-temporal information can be projected in time-localized feature spaces. The complementary of the representations is exploited and an inference of the dynamic feature space modeling is performed using Kullback-Kleiber divergence. An Example of graph obtained is displayed in figure 2. Moreover, the modeling can be mapped back in the image space and leads to dynamic classifications of images time-series.

III. Exploring a graph characterizing the dynamic clusters and dynamic classifications

The hierarchical modeling of the image time-series produces a graph characterizing the dynamic clusters. To each dynamic cluster trajectory is indexed a dynamic classification. The association of these two objects characterizes fully the image time-series. However, the information contained in the graph and in the dynamic classifications is huge and a GUI is required to explore paralally both representations. A Snapshot of the GUI, presented in figure 3, displays the colza crops associated cluster most probable trajectory together with its correspondent multitemporal class and time-localized class at a given time.

IV. Entropy measures for searching sub-graph similarities

The GUI enhances the comprehension of the graph. However, the density of information contained in the graph is high. Furthermore, plotting cluster most probable trajectories is a simple visualization which does not describe fully the GM evolutions. Consequently, similarities measures need to be defined in order to characterize the graph for a better understanding.

A. Searching similar degree of change using mutual information

Mutual information $I(X_{t_1}, X_{t_2})$ between two random variables X_{t_1} and X_{t_2} quantifies the amount of information contributed by X_{t_1}, about X_{t_2}. It is well suited to measure the reduction of uncertainty between two GM relative to a trajectory in consecutive times. As the marginal PDF $P(X_{t_1})$ and $P(X_{t_2})$ are already known, only a Gaussian mixture estimation approximating the joint PDF $P(X_{t_1}, X_{t_2})$ remains for the calculation of mutual information. The MDL based GM estimation algorithm is used for this approximation [2].
Mutual information evolutions are significant of the degree of change of the multitemporal cluster during time. Set of consecutive nodes can be grouped by similarity according to this measurement.

B. Searching entropic and morphologic similarities in an oriented graph

Denoting by v_i the nodes of the sub-graph with $V = \cup_i v_i$ and by $pa(v_i)$ the parent nodes of v_i, the oriented sub-graph PDF $G(X)$ is defined by

$$G(X) = \prod_{v_i \in V} G_{v_i|pa(v_i)}(X_{v_i}|X_{pa(v_i)}) \tag{1}$$

where $G_{v_i|pa(v_i)} = G_{v_i pa(v_i)}/G_{pa(v_i)}$ is the PDF of the node v_i conditioned by the parents nodes PDF. Because the sampling rate in time is irregular, the entropy of the spans of the lattice has to be considered. Shannon entropy of discrete quantized stochastic process considers this additional uncertainty [3]. It is defined for $G(X)$ by

$$H\big(G(X)\big) = -G(X)log\big(G(X)\big) + k\sum_j log(\tau_j) \tag{2}$$

where k is a constant, τ_i defines a length interval with
$\tau_1 = (t_2 - t_1); \tau_m = (t_m - t_{m-1});$
$\tau_j = (t_{j+1} - t_{j-1})/2 \quad \forall j \in [2, m-1].$
Entropies of oriented sub-graphs are thus define in a time window $[1, m]$.

In order to discover similarities in the oriented graph, the mining procedure is decomposed in several levels. First, the collection of sub-graphs is defined by determining a sub-graph of interest and thus, a temporal window size. Then, the sub-graphs which do not possess the same morphology are removed from the collection. The sub-graph entropies are estimated and sorted by decreasing entropic similarity measurements.

V. Conclusion

This work is an attemp to solve the complex problem of spatio-temporal reasoning in image time-series. A modeling of the dynamic feature space is achieved. Entropic measurement are applied in order to define similarity measures in sub-graphs and thus, to enable visual information mining by the dynamic feature space exploration of image time-series.

References

[1] M.Datcu, K.Seidel, M.Walessa (1998). *Spatial information retrieval from remote-sensing Images -part I : information theoritical perspective*. IEEE transactions on geoscience and remote sensing, vol. 36. N^o.5. september 1998.*

[2] P.Heas, M.Datcu, A.Giros (2003).*Trajectory of dynamic clusters in image time-series* Second International Workshop on the Analysis of Multi-temporal Remote Sensing Images, Ispra ITALY.

[3] G.Jumarie (1990). *Relative Information, theories and applications.* Editor : Hermann Haken. Publishers / Tokyo.

[4] A.Spartaru(1987). *Fondements de la théorie de la transmission de l'information.* Lausanne: Presse Polytechniques Romandes.