

Image Time-Series Mining

Patrick Heas¹, Philippe Marthon²

Institut de Recherche en
Informatique de Toulouse
Toulouse, France

¹Email: patrick.heas@enseiht.fr

²Email: philippe.marthon@enseiht.fr

Mihai Datcu

DLR, German Aerospace Center
Oberpfaffenhofen, Germany
Email: mihai.datcu@dlr.de

Alain Giros

Centre National d'Etudes Spatiales
Toulouse, France
Email: alain.giros@cnes.fr

Abstract—A visual information mining concept is proposed for spatio-temporal patterns discovery in remotely sensed image time-series (ITS). An information theory framework is adopted to first model information content. It results in the inference of a relevant directed graph characterizing ITS. Then the user conjecture is modeled via visual information representations : similarity measures between sub-graphs, which represents spatio-temporal events are derived and included in an interactive learning and probabilistic retrieval procedure of user-specific event-types. The present concept for ITS mining is demonstrated on multitemporal SPOT data.

I. INTRODUCTION

During the last decades, the imaging satellite sensors have acquired huge quantities of data enabling the elaboration of spatio-temporal databases. Image time-series (ITS) contained in them are complex objects containing interesting information. However, our capability to access and store large volume of data has highly exceeded our capability to extract the relevant information from the data. Therefore, ITS information systems are needed to bridge the gap between database theory, which is not adapted to manipulate temporal and pictural multidimensional data, and computer vision, which does not offer query possibilities. To cover such a broad problematic, a visual information mining concept is proposed to help a user to explore and understand spatio-temporal behaviors present in ITS.

The article begins by presenting unsupervised methods for coding the information content in the observations. Thus, the dynamics of the feature space (FS) is modeled which results in the inference of a relevant directed graph representation, characterizing ITS. It continues by describing sub-graphs, contained in the graph, with model parameters. Combinations of the parameters correspond to semantic descriptions of spatio-temporal events. In this perspective, a distance model between sub-graph parameters is developed to enable user specific similarity measurements. Then, a Bayesian interactive learning on graphs is presented to enable reasoning on ITS. The learning is performed by the incremental definition of a spatio-temporal pattern type via user-provided positive and negative sub-graph examples. In parallel, the interactive search of spatio-temporal events is guided by accuracy probabilistic measurements derived from the current state of learning via the user positive and negative semantic. Finally, experiments

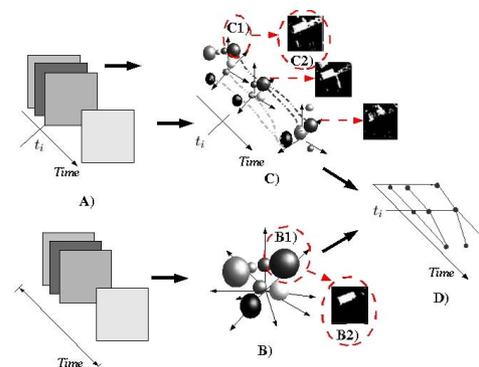


Fig. 1. Different representations of the image time-series $x(r, t, \Theta)$. A) Spatio-temporal space $x(r, t, \Theta)$; B) MT feature space $x_r(\Theta_{t_1}, \dots, \Theta_{t_d})$, B1) MT cluster C_{t_1, \dots, t_d}^j , B2) MT spatial class $C_{t_1, \dots, t_d}^j(r)$; C) TL feature spaces $\{P_{t_i}(x_r(\Theta_{t_1}, \dots, \Theta_{t_d}))\}$, where P_{t_i} denotes the operator projecting the MT feature space on the TL feature component Θ_{t_i} , C1) TL cluster $C_{t_i}^j$, C2) TL spatial class $C_{t_i}^j(r)$; D) Graph characterizing the dynamic clusters $x_r(t, \Theta)$.

performed on a SPOT ITS, demonstrate the presented ITS mining concept.

II. GRAPH REPRESENTATION OF IMAGE TIME-SERIES

Image time-series are stochastic processes which have different representations in a multidimensional space comprising two spatial axes r , a time axis t , and several feature dimensions Θ (radiometric values, texture parameters,...). The multidimensional signal denoted by $x(r, t, \Theta)$ can have various interpretations depending on its representations. The spatio-temporal representation $x(r, t, \Theta)$ is simply the time-series of images. Denoting by d the number of time samples, the multitemporal (MT) feature space representation $x_r(\Theta_{t_1}, \dots, \Theta_{t_d})$ is a multidimensional space composed by the union of all the time-localized (TL) feature components $\{\Theta_{t_i}; i \in [1, d]\}$.

A. Modeling and clustering multitemporal and time-localized feature spaces

The goal is to understand the evolving scene. Moreover, the MT feature space contains the global information on the image time-series. Therefore, the dynamic of image structures and spatial objects is analyzed in the multidimensional space. Figure 1 illustrates the analysis [2]. The signal is modeled

and clustered in different FS representations. In particular, TL representations are considered in order to study the evolution in time of image structures. They are obtained by projecting the multidimensional FS on a TL feature component Θ_{t_i} , using a defined projector operator. Thus, for each time t_i , a TL feature space representation is obtained. When modeling the MT feature space, dimension reduction techniques such as principal component analysis or projection pursuit are used to condense the information contained in the MT feature space into a sub-space of lower dimensionality. They provide an efficient solution for parameter and model inference in multidimensional spaces with limited sample sizes. Stochastic models are required to learn about this multidimensional signal. A Gaussian mixture model is able to approximate efficiently any probability distribution function (PDF); thus it is well suited to model PDF of signal $x_r(\Theta_{t_1}, \dots, \Theta_{t_d})$ for which no prior knowledge is available. The Gaussian mixture modeling measures similarities used for clustering the FS. Each Gaussian will be considered as a grouping of similar data points and thus, will define a cluster. In order to perform clustering without any constraints neither on the number of Gaussians present in the mixture nor on their parameters, a minimum description length criterion has been used to select the best model among all the possible Gaussian mixture configurations [2].

B. Inference of the dynamic feature space modeling

Each MT cluster has a particular behavior while observing its evolution into successive TL representations. For example, some MT clusters may share the same TL cluster at a given time and split or/and merge with other MT clusters in the future. The problem is to quantify, at a given time, the similarity of these projected MT clusters with the goal to infer temporal relations. By associating the TL and the MT information, the problem of quantifying the distance between projected MT clusters can be reformulated in defining an information measurement, able to compare two different models and learn about their similarity. Using Kullback-Leibler divergence, which is relevant for these model comparisons, the dynamic FS modeling is inferred resulting in weighted MT cluster trajectories formalized through TL evolution of clusters [2]. Therefore, this analysis results in a graph \mathcal{G} representation of $x_r(t, \Theta)$ characterizing the dynamic clusters.

C. Graph characterizing the dynamic clusters

The directed graph \mathcal{G} is a specific multidimensional temporal feature representation which characterizes the trajectories of the dynamic clusters. The information is condensed in its vertices and edges. A vertex is representing a TL cluster and is related to a given MT Cluster. It is characterized by a pixel weight π , Gaussian parameters $\xi = (M, A)$ and a divergence measurement. Moreover, spatial information is contained in the corresponding TL classes. An edge, representing the evolution of the MT cluster between two image samples, is characterized by a time sampling delay T , a pixel flow γ , Gaussian parameter evolution $\delta(\xi)$ and mutual information

MI . Furthermore, the graph is composed by the union of all MT evolutions. Thus MT clusters and classes are assigned to vertices and edges.

Unsupervised methods for coding the information content in the observation results in a graph characterization of an ITS. Therefore, the graph can be included in an information-mining system enabling to add the user semantic which helps for interpretations by supervised methods which are using relevant information.

III. SIMILARITIES BETWEEN SUB-GRAPHS

Sub-graphs, contained in the graph, are statistical models for spatio-temporal processes, which at their turn describe physical changes in the observed scene. The sub-graphs are described by model parameters and combinations of the parameters correspond to semantic descriptions of spatio-temporal events. Thus, a parametric distance model between sub-graph is developed to enable user specific similarity measurements.

A. Parametric distance model for sub-graph similarity

Inexact graph matching has been proposed in [1]. It enables the evaluation of a similarity measurement between sub-graphs. The idea is to transform one of the graphs into the other one by assigning a cost to each distortion. Thus, a cost is assigned to each vertex or edge addition/removal. Graphs characterizing ITS are supporting several attributes on their vertices and edges. Thus, a sub-graph distance is defined by adapting the inexact graph matching algorithm into a parametric model, weighting the different attribute contributions.

Denoting by $\nu_1 = \{\nu_i^1\}$ and $\nu_2 = \{\nu_j^2\}$ the vertex sets of sub-graphs \mathcal{G}_1 and \mathcal{G}_2 , and denoting an extra set of vertices by $\lambda = \{\lambda_i\}$, a mapping function $\mathcal{F} = \{f\}$ composed by a given combination of elementary mapping functions $f: \nu^1 \rightarrow \nu^{2\lambda} = \nu^2 \cup \lambda$ is defined. Note that f is an injective function. A cost $C_\Theta(f(\nu_i^1) = \nu_j^{2\lambda})$ is assigned to each elementary transformations. The cost function depends on the vector parameter $\Theta = \{\theta_k\}$ and is composed by a weighted sum of similarities between vertices ν_i^1 and $\nu_j^{2\lambda}$ and related edges. Preserving notation of paragraph II-C, when considering only incoming edges related to a node ν_i , flows and Gaussian parameter evolutions related to edges are denoted by γ_{ν_i} and $\delta_{\nu_i}(\xi)$. The cost is defined by the expression

$$C_\Theta(f(\nu_i^1) = \nu_j^{2\lambda}) = \theta_1 \Delta(\pi_{\nu_i^1}, \pi_{\nu_j^{2\lambda}}) + \theta_2 \Delta(\xi_{\nu_i^1}, \xi_{\nu_j^{2\lambda}}) \quad (1) \\ + \theta_3 \Delta(\gamma_{\nu_i^1}, \gamma_{\nu_j^{2\lambda}}) + \theta_4 \Delta(\delta_{\nu_i^1}(\xi), \delta_{\nu_j^{2\lambda}}(\xi))$$

where $\Delta(\cdot)$ represents a distance functions which is either a difference for a scalar or a similarity measure between PDFs such as Kullback-Leibler divergence. Because time sampling delay and mutual information are characterizing all edges of a given MT class in a given interval, it must be added in the cost function only once per time interval. Furthermore, when considering 2 MT cluster at a time, their distance denoted by $\delta_{MT}(\xi)$ is evaluated using Gaussian parameter similarity and reported once per time sampling in the cost function.

Thus, when satisfying the given conditions, the following terms $\theta_5\Delta(MI) + \theta_6\Delta(T) + \theta_7\Delta(\delta_{MT}(\xi))$ are added to equation (1). The sub-graphs similarity is then defined, for a given vector parameter Θ , by finding the less expensive elementary mapping function combination over all possible mapping functions:

$$S_{\Theta}(\mathcal{G}_1, \mathcal{G}_2) = \min_{\mathcal{F}} \left(\sum_i C_{\Theta}(f(\nu_i^1) = \nu_j^{2\lambda}) \right). \quad (2)$$

In order to estimate the minima, an optimization procedure is performed searching a minimum cost path in a tree structure containing all possible mapping functions configurations. Because of the combinatorial explosion of possible configurations, the tree is regularly pruned during the search, with the potential drawback that the optimization can lead to a local minima.

B. Semantic description of spatio-temporal events

Similarities between sub-graphs have been defined using a parametric distance model. As each term of the cost function characterizes a physical behavior in the observed scene, one can attach a semantic to each component of equation(1). Thus, in order to assign a physical meaning, the different components are reviewed sequentially. Factor θ_1 is weighting the pixel quantity concerned by the event. Factor θ_2 is weighting the temporal feature value characterizing the event. Factor θ_3 is weighting the pixel exchanges during the event. Factor θ_4 is weighting the feature value change. Factor θ_5 is weighting the inter-class change during the event. Factor θ_6 is weighting the delay in which occurs the event. Factor θ_7 is weighting the MT class interactions. However, tuning correctly these parameters in order to define a semantically consistent distance model, might not be obvious for a user. Therefore a supervised learning procedure is needed to estimate the parameter vector characterizing a specific semantic.

IV. BAYESIAN INTERACTIVE LEARNING ON GRAPHS

A concept of interactive learning and probabilistic retrieval of user-specific event-types is developed for reasoning in ITS using visual representations. The interactive learning scheme has been developed in the framework of a concept presented in [3].

A. Mining a sub-graph collection

Let the graph \mathcal{G} be decomposed into a collection of p sub-graphs $\{\mathcal{G}_k\}$. The similarity measure of equation (2) between a given sub-graph \mathcal{G}_0 and a sub-graph \mathcal{G}_k can be rewritten as $S_{\Theta}(\mathcal{G}_0, \mathcal{G}_k) = \sum_i S_{\theta_i}(\mathcal{G}_0, \mathcal{G}_k)$, where $S_{\theta_i}(\mathcal{G}_0, \mathcal{G}_k)$ is the contribution related to the parameter θ_i , to the cost function. For a particular sub-graph \mathcal{G}_k , the parameter θ_i is assumed proportional to the partial cost function $S_{\theta_i}(\mathcal{G}_0, \mathcal{G}_k)$. Therefore, the learning step is performed by the incremental definition of a spatio-temporal pattern type via user-provided positive and negative sub-graph examples. Let the positive and negative semantic of a user be denoted by \mathcal{A}_{ν} and $\neg\mathcal{A}_{\nu}$. From these examples, we infer probabilities of a Bayesian

network, based on a Dirichlet model, that links user interest to specific parameter PDFs $p(\Theta | \mathcal{A}_{\nu})$ and $p(\Theta | \neg\mathcal{A}_{\nu})$. Then, the minimum mean square estimator is used to estimate the parameter vector.

In a first step, let us consider only the positive semantic parameter learning. Using the new estimation, the distance model function $S_{\Theta}^{MMSE}(\mathcal{G}_0, \mathcal{G}_k)$ can be evaluated and a graph likelihood probability can be assigned to each sub-graph of the collection according to its evaluated distance. As the search tree structure for the calculation of the distance model have been saved the computation of the distance model is not intensive, which enables the interactivity of the learning process. The likelihood probability of a given sub-graph \mathcal{G}_k conditioned by the user semantic \mathcal{A}_{ν} is defined by

$$p(\mathcal{G}_k | \mathcal{A}_{\nu}) = \Phi \left(S_{\Theta}^{MMSE}(\mathcal{G}_0, \mathcal{G}_k) \right) \quad (3)$$

where $\Phi(\cdot)$ is a linear operator inversely proportional, mapping the similarity function values into probabilities. A similar strategy is adopted for the estimation of the sub-graph likelihood probabilities $p(\mathcal{G}_k | \neg\mathcal{A}_{\nu})$ conditioned by the negative semantic. Note that the probabilities are dependent of the negative and positive reference graphs \mathcal{G}_0 , which are different. Then, using the Bayes rule, we calculate the posterior probabilities $p(\mathcal{A}_{\nu} | \mathcal{G}_k)$ as

$$p(\mathcal{A}_{\nu} | \mathcal{G}_k) = \frac{p(\mathcal{G}_k | \mathcal{A}_{\nu})p(\mathcal{A}_{\nu})}{p(\mathcal{G}_k)} \quad (4)$$

with the graph priors \mathcal{G}_k being $p(\mathcal{G}_k) = \sum_{\nu} p(\mathcal{G}_k | \mathcal{A}_{\nu})p(\mathcal{A}_{\nu})$ where the sum over ν is computed using positive and negative semantics. Thus, assuming a uniform prior for the semantics, using equation 4, the posterior probability of a graph conditioned by the semantic \mathcal{A}_{ν} can be calculates as

$$p(\mathcal{A}_{\nu} | \mathcal{G}_k) = \frac{p(\mathcal{G}_k | \mathcal{A}_{\nu})}{p(\mathcal{G}_k | \mathcal{A}_{\nu}) + p(\mathcal{G}_k | \neg\mathcal{A}_{\nu})}. \quad (5)$$

Therefore, the collection of sub-graphs can be sorted, in ascending order, according to posterior probabilities. Changing of reference graphs must be done regularly when the user adopts a browsing strategy for his search. The interactive training can be iterated until satisfaction of the user.

B. Evaluating the search quality

The search of spatio-temporal events is guided by a measure a strength of stochastic links which can be computed for each parameter. This measure can serve to evaluate the strength of each characteristic of the similarity function. Denoting by $D(\cdot | \cdot)$ the Kullback-Leibler divergence, the strength of stochastic links is defined by $D(p(\theta_i | \mathcal{A}_{\nu}) | p(\theta_i | \neg\mathcal{A}_{\nu}))$.

V. PRACTICAL APPLICATIONS

Experiments were performed on a temporal database composed by 38 superposable 2000x3000 pixels SPOT images which were obtained by daily acquisition and by filtering the cloud-free images in a period of 286 days. The temporal sequence displays a dynamic scene acquired in the East of Bucharest (Romania). A spatial subset of 200x200 pixels is

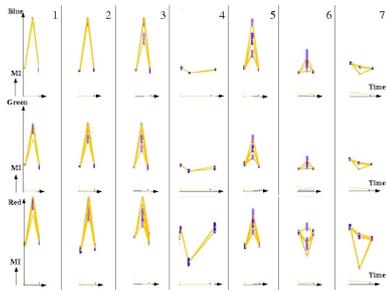


Fig. 2. 7 most probable graph patterns characterizing cloud apparitions

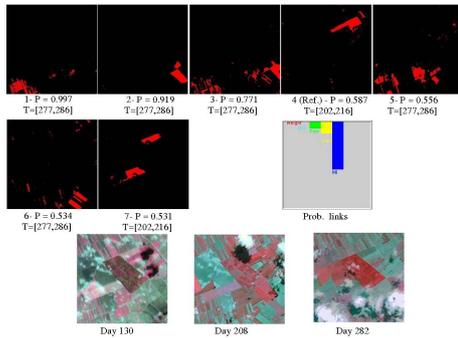


Fig. 3. Cloud search results. Above, probabilities of the 7 most likely spatial class with associated time-windows. Below, retrieved image samples containing clouds

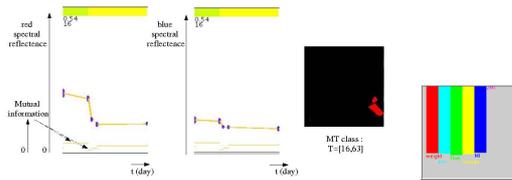


Fig. 4. Harvest reference pattern together with its spatial class & search probabilistic links

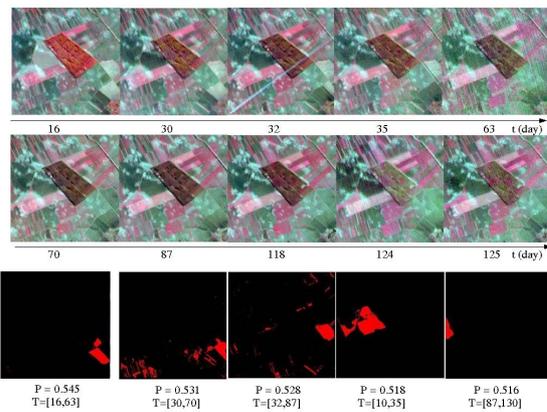


Fig. 5. Harvest search results. Above ITS between day 16 and day 125. Below, Probabilities of the 5 most probable spatial classes with associated time-window

considered. The features are spectral reflectances.

The first step was completed by the modeling and the clustering of the MT and the TL feature spaces resulting in the inference of the graph. Then, an interactive learning step has been performed using a Java visual interface which displays, in a synchronized way, the ITS, the graph, the spatial TL and MT classes. Two different searched semantics are presented in this paper : cloud apparitions and harvest phenomena. A first training was performed using as a reference the 4th cloud pattern in figure 2 which is a cloud shadow partial covering of a field. A time-window of 3 samples was used and time delays were not considered in the similarity function (c.f. section III-B). Figure 2 displays the 7 graph patterns of highest probability resulting from the search . Figure 3 displays the corresponding MT spatial classes together with their probabilities, their time coordinates and the corresponding images where clouds appear. The probabilistic links clearly demonstrate that the cloud apparition semantic is highly characterized by mutual information (low), a little by radiometric signature changes (as the change depends on the occluded object radiometry), pixel exchanges and weights but not by radiometric signatures since we are searching for clouds and shadows. A second training was performed using a harvest semantic. The reference graph pattern is presented in figure 4. Since green and blue spectral signatures are similar, only 2 spectral reflectances are displayed. A time-window of 5 samples was used and time delays were not considered in the similarity function. The probabilistic links show that the harvest phenomenon is characterized almost equally by all the measures presented in section III-B. Figure 5 displays the search results. Above the ITS between day 16 and day 125 is displayed. Almost all the retrieved harvests occurred in this period. Below is presented the resulting 5 most probable spatial classes together with their time-window. A probability is assigned to each spatial class presented in its given time-window.

VI. CONCLUSION

This work is an attempt to solve the complex problem of spatio-temporal reasoning in image times series. A trajectory modeling of dynamic clusters of the image time-series is proposed to extract information related to evolving image structures and characterize the global spatio-temporal signal. The resulting graph is included in an interactive learning and probabilistic retrieval procedure of user-specific event-types. The experiments demonstrate the relevance of the method by the interactive characterization of spatio-temporal events.

REFERENCES

- [1] H. Bunke and G. Allerman, *Inexact graph matching for structural pattern recognition*, Pattern Recognition Lett. 1(4), 245-253, 1983.
- [2] P. Heas, M. Datcu and A. Giros, *Trajectory of dynamic clusters in image time-series*, Proceedings of the 2nd international workshop on the analysis of multitemporal remote sensing images, Ispra(VA), Italy, 2003.
- [3] Schrder M., Rehrauer H., Seidel K., Datcu M., *Interactive learning and probabilistic retrieval in remote sensing image archives*, IEEE Trans. on Geoscience and Remote Sensing, ISSN 0196 2892, Vol. 38, pp. 2288-2298, 2000. .