MINING IMAGE TEMPORAL CHANGES

Inés María Gómez Muñoz and Mihai Datcu
Remote Sensing Technology Institute, Oberpfaffenhofen Germany

ABSTRACT

Important information on the scene changes is contained in temporal image sequences. The generation of the map containing the targets in multitemporal high resolution hyperspectral images is not an easy task. The problematic is the detection of "risk" targets like cars or tracks in situation of multitemporal observations in different illumination conditions and strong background clutter.

In the article are introduced unsupervised techniques for target detection. They are based on the extraction of the basic image primitives as spectral signatures or texture parameters, on the analysis of the spectral bands applying different operations between them, as differences, ratios or temporal spectral angular distance, and also on the analysis of the principal components. Other techniques more related with the composition of color of each band as vegetation indexes could help in target detection. Following this approach, illumination invariant indexes are developed in order to detect the strong false alarms. Finally, the different algorithms are combined to optimize the final results.

However, images can not only contain the quantitative and objective information obtained by unsupervised algorithms, but also subjective based on knowledge. Knowledge-driven Information Mining System (KIM) is built in order to formalize the knowledge acquisition and the knowledge driven interpretation. It provides solutions how to access to large image data sets through information mining, and content based image retrieval.

1. INTRODUCTION

The goal of the research is to identify the sensor limits in the field of target detection by change analysis of multitemporal high resolution hyperspectral images. The requirements on data pre-processing are realistic, i.e. the primitive feature shall be generic such that by different combinations is able to describe a large variety of changes and the data could be expected to be acquired from a satellite sensor in an operational system.

The creation of the archive catalogues or scene understanding paradigms are based on the primitive feature extraction. It has not been considered as relevant external knowledge in order to implement a robust and general method to identify targets and reject non interesting changes independent of the data set.

The article is organized into three sections. The first one presents the *Knowledge-driven Information Mining System (KIM)*. The knowledge consists in the ensemble of existing information, know causalities and other type of associations between information and concepts. In the system, the user-defined semantic image content interpretation is linked with Bayesian networks to the completely unsupervised models. The meaning of image objects or structures is obtained by an interactive learning process fusing the relevant information extracted from the sensor image data set. A right combination of models and a good interaction by the user with the system goes to a clear detection of the targets.

The next section describes a library of unsupervised algorithms for characterization of scene changes. There are three kinds of changes: those representing the targets, those non interesting changes, also called false alarms, that correspond mostly to shadows, and finally, the changes due to strong clutter on the scenes.

2. KNOWLEDGE-DRIVEN INFORMATION MINING SYSTEM (KIM)

In general, by image we understand picture thus relating it to the (human) visual perception and understanding. A picture is characterized by its primitive features such as colour, texture, and shape at different scales. However, images can not only contain this quantitative and objective information, but also subjective based on knowledge. The knowledge consists in the ensemble of existing information, know causalities and other type of associations between information and concepts.

Knowledge-driven Information Mining System (KIM) [6][7] is built in order to formalize the knowledge acquisition and the knowledge driven interpretation. The *KIM* concept and system provide solutions how to access to large image data sets through information mining, and content based image retrieval. In the system, the user-defined semantic image content interpretation is linked with Bayesian networks to a completely unsupervised content-index. Based on this stochastic link, the user can query the archive for relevant images and obtains a probabilistic classification of the entire image archive as an intuitive information representation.

The hierarchy of information representation of the KIM System can be summarize in the following way:

data - features - metafeatures - cluster (class) models - semantics

The concept of information representation is based on a 5 level Bayesian learning model (see figure 1):

- *Level 0*: Representation of the image data *D*. In our research it consists on optical hyperspectral data formed with 12 bands.
- *Level 1*: Primitive image features θ extracted from image data D (level 0) through application of stochastic models (*sp, tx*).
- *Level 2*: Meta-features extracted from image data D (level 0) using different signal models (*sp,tx*), i.e. the type of model used for the extraction of the primitive features, and the scale at which the analysis was done.
- *Level 3*: Completely unsupervised clustering of the pre-extracted image parameters θ, we obtain a vocabulary of signal classes ω_i that describe characteristic groups of points separately for each model *(sp, tx)*. For each image, this results in as many classification maps as models are used, and they are stored in a relational database system. The used classification algorithm is called *k-means*.
- *Level 4*: User-specific interests, that is, semantic cover-type labels A, are linked to combinations of these vocabularies ω_i by probabilities $p(\omega_i \mid A_i)$ (Bayesian networks).

Levels 1 to 3 are obtained in a completely unsupervised and application-free way during data ingestion in the system. The information at level 4 can be interactively defined by users with a learning paradigm that links (objective) signal classes and user (subjective) labels A.

3. FEATURE EXTRACTION FOR TARGET DETECTION

This section is concentrated on the second step of the KIM chain: *feature extraction*. In this phase, different unsupervised techniques are applied in order to obtain illumination independence, to extract the basic image primitives as spectral signatures or texture parameters, to analize the principal components of the images or to obtain the wrong targets. In the following these methods are briefly presented.

3.1. Texture Parameters

This section pretends to explain how to extract information from the observed data as contextual spatial features for false alarms rejection. To extract structural information from remote sensing images, stochastic models are applied, for instance, the *Gibbs-Markov Random Field (GMRF)*. This algorithm assumes that the statistics of the grey level of the pixels belonging to a neighborhood with a restricted dimension.

From the estimated parameters, it is derived several features to describe the image content: the norm of the estimated parameters $|\hat{\theta}|$ as the strength of the texture, the estimate of the variance $\hat{\sigma}^2_M$, the evidence of the model M and the local mean, e.g. the twelve spectral channels in the visible spectrum of sensor result in a twelve-dimensional vector $\theta = \theta_1, \theta_2, ..., \theta_{12}$.

3.2. Spectral Signatures

To describe the image content by using the spectral properties, directly assign the individual spectral channels to elements of the vector ξ, e.g. the twelve spectral channels in the visible spectrum of the sensor result in a twelve-dimensional vector $\xi = \xi_1, \xi_2, ..., \xi_{12}$.

3.3. Principal Component Analysis

Principal Components Analysis technique produces uncorrelated output bands, segregates noise components, and reduces the dimensionality of data sets. Because multispectral data bands are often highly correlated, the Principal Component (PC) Transformation is used to produce uncorrelated output bands. This is done by finding a new set of orthogonal axes that have their origin at the data mean and that are rotated so the data variance is maximized. PC bands are linear combinations of the original spectral bands and are uncorrelated. The first PC band contains the largest percentage of data variance and the second PC band contains the second largest data variance. The third one appears noisy because it contains very little variance, much

of which is due to noise in the original spectral data. Principal Component bands produce more contrast composite images than spectral composite images because the data is uncorrelated.

3.4. Temporal Spectral Angular Distance (TSAD)

The aim of this algorithm is the detection of targets getting illumination independence. The method focuses the attention on pixel by pixel in all spectral bands. A vector for each pixel is formed, taking all its values in all bands. This operation is done for time 1 and time 2, and later, the angle that both vectors forms is computed. Proceeding in this way, a map where the unchanged pixels present small gray-level values and the changed pixels are marked stronger is obtained.

In a more formal representation, each pixel is considered as a vector:

$$B_1 = [b_1, ..., b_n], t_1 \tag{1}$$

$$B_2 = [b_1, ..., b_n], t_2 \tag{2}$$

and it is calculated the cosines of the angle that the vectors form as follows:

$$\cos\alpha = \frac{B_1 B_2}{||B_1||||B_2||} \tag{3}$$

The angle between the hyperspectral signature at the two times reduce dependence on the intensity, thus pixels with similar signature but in different illumination conditions will have small angles. Contrary, new "target" will be characterized by large angle values.

False alarms are mainly produced by registration errors or low signal noise values in shadow areas.

3.5. Normalized Difference Vegetation Index (NDVI)

Especially the Dorsten and Borculo sites are mainly cover by vegetation. For this reason, it is going to be applied another algorithm based on a vegetation index. It is a value that is calculated from sets of remote sensing data that is used to quantify the vegetative cover on the Earth's surface. The NDVI is calculated as the ratio between the red and near infrared (NIR) spectral bands:

$$NDVI = \frac{NIR - red}{NIR + red} \tag{4}$$

These two bands are chosen because they are most affected by the absorption of chlorophyll in leafy green vegetation in the case of the red band and by the density of green vegetation on the surface, this is the case of the near infrared band.

In order to maximize the range of NDVI values and provide numbers that are appropriate to display in an 8 bit image, the NDVI value must be scaled. We will do it applying the following formula:

$$ScaledNDVI = 100(NDVI + 1) \tag{5}$$

Now, the NDVI computed value is scaled to the range of 0 to 200. As a result, NDVI values less than 100 now represent clouds, snow, water, and other non-vegetative surfaces, and values greater than 100 represent vegetative surfaces. After scaling the ndvi images in time 1 and in time 2, the ratio between both data in order to enhance the spectral differences and to reduce the effects of topography is calculated.

The NDVI results in a good invariant transformation of the data, however it is using only part of the hyperspectral signature.

3.6. Gaussian Normalize Algorithm

In this section, it is discussed a method to detect non relevant changes in order to reduce the false alarms.

The method consists of a color based technique, called *Color Normalize, Brovey* [8], and an *statistical algorithm*. The method is implemented as a mathematical combination of the spectral image and high resolution data. Each band in the hyperspectral signature is multiplied by a ratio of the high resolution data divided by the sum of the spectral bands. The function automatically resamples the three spectral bands to the high-resolution pixel size using a nearest neighbor. It is resulting in an image with the pixel size of the input high-resolution data. The value of each pixel of the resulting data will be between 0.0 and 1.0, representing the probability of occurrence or success.

$$CN = B_n * \frac{data}{B_1 + .. + B_n} \tag{6}$$

The *statistical algorithm* consists of computing to each pixel, the cutoff value V in a standard Gaussian distribution with a mean of 0.0 and a variance of 1.0 such that the probability that a random variable X is greater than V is equal to a user-supplied probability P.

3.7. Synergy of Methods

As a result of the previous methods, a good approach could be the combination of them. The basic idea is to take out the false alarms, obtained through the *gaussian* algorithm, from the *temporal scalar product* and *ndvi* results.

Analyzing the exact pixel where the interesting target or non interesting one is, wrong results will get due to the strong scene clutter of the data. For this reason, a reasonable neighborhood must be considered, in such a way that when a pixel belonging to

a false alarm is found, the method will take out not only that pixel, but also the considered neighborhood from the *temporal spectral angular distance* and *ndvi* results. In order to decide this order, the data must be careful analyzed. In Dorsten case, the pixels belonging to false alarms are far from those belonging to target ones. On the contrary, in Borculo and Marseille data the targets are very closed to false alarms. This act can lead to avoid targets when false alarms are taking out. For this reason, in the experiments the considered neighborhood will be of five pixels around for the case of Dorsten and one pixel for Borculo and Marseille site.

4. CASES STUDIES

The experiments have been done using high resolution hyperspectral images acquired by the airbone sensor Daedalus. However the results are more general can be applied for other airbone or satellite temporal sensor. Daedalus has 12 spectral bands ranging from the visible blue to the thermal infrared. This cover the visible and near-infrared (Ch 1-8) shortwave infrared (Ch 9 and 10) and thermal infrared (Ch 11) and include channels that closely match the seven spectral channels of the Landsat Thematic Mapper. The pixel size (ground spatial resolution) of this sensor will be dependent on the flying altitude of the aircraft, in the present case the resolution is 1 meter.

For the study, data over three sites have been used: *Dorsten* (Germany), *Borculo* (Netherlands) and *Marseille* (France) sites. The first site (fig.2) is situated in the region of Nordrhein-Westfallen in Germany. It includes a natural gas pipeline in area with vegetation and various constructions (houses). From a scene of 600 x 600 pixels, a central sub-image of 466 x 466 pixels containing the relevant targets was selected. The second data set is placed in the vicinity city of Marseille in France. In this case, it is a big data set containing 55 different sites in time 1 and time 2 for the city of Marseille. As example, a relevant sub-image of 466 x 466 pixels is chosen (fig. 3). The third site, the Borculo data set (fig. 4) in Netherlands (1000 x 625 pixels), is a scene containing none gas pipeline in a typical agriculture environment. From the data set, two subsets of 466 x 466 pixels are selected.

Once presented the data sets, an interactive learning process will be applied to them.

The meaning of image objects or structures is obtained by an interactive learning process fusing the relevant information extracted from the sensor image data set. It is based on a learning algorithm to select and combine features and to allow users to give positive and negative examples. This combination will be formed by four models chosen between *spectral in time 1 and time 2, spectral Principal Component Analysis in time 1 and time 2, texture in both times, texture PCA in time 1 and time 2, temporal scalar product, NDVI, gaussian, TSAD-gaussian and NDVI-gaussian models*.

The method refines the user interaction and enhances the quality of the queries.

In figures 5, 6, are shown the results of different interactive learning iterations using, for each case, a distinct combination of models.

REFERENCES

Allan A. Nielsen, Knut Conradsen, and James J. Simpson, "Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data: New Approaches to Change Detection Studies". http://dial.ucsd.edu/projects/mafmad/mafmad.html.

Lorenzo Bruzzone, and Diego Fernandez Prieto, "Automatic Analysis of the Difference Image for Unsupervised Change Detection", IEEE Transactions on Geoscience and Remote Sensing, Vol. 38, No. 3, pp. 1171-1182, May 2000.

Lorenzo Bruzzone, and Diego Fernandez Prieto, "Unsupervised Retraining of a Maximum Likelihood Classifier for the Analysis of Multitemporal Remote Sensing Images", IEEE Transactions on Geoscience and Remote Sensing, Vol. 39, No. 2, pp. 456-460, February 2001.

Alan J. Lipton, Hironobu Fujiyoshi, Raju S. Patil, "Moving target classification and tracking from real-time video", http://www.cs.cmu.edu/ vsam.

G. D. Finlayson, S. D. Hordley and M.S. Drew, "Removing Shadows From Images", School of Information Systems, University of East Anglia, UK, School of Computer Science, Simon Fraser University, Canada, http://www.cs.sfu.ca/ mark/ftp/Eccv02/ shadowlessblurb.pdf

M.Schröder, H. Rehrauer, K. Seidel, M. Datcu, "Interactive Learning and Probabilistic Retrieval in REmote Sensing Image Archives", IEEE Transactions on Geoscience and Remote Sensing, Vol. 38, pp. 2288-2298, 2000.

M. Datcu, K. Seidel, S. DElia, P. G. Marchetti, "Knowledge-driven Information-Mining in Remote Sensing Image Archives", ESA Bulletin 110, pp. 26-35, 2002.

"Spectral Signatures and Multi-spectral Image Interpretation", Note 8.

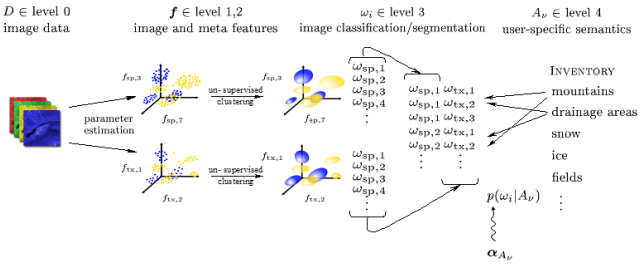

Figure 1. KIM formal representacion.

Figure 2. Dorsten site. The image on the left shows Dorsten site in time 1 and the one on the right shows the Dorsten data at time 2.

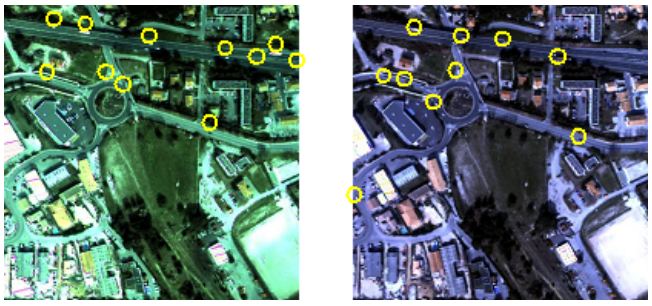

Figure 3. Marseille site. The image on the left shows Marseille site in time 1 and the one on the right shows the Marseille data at time 2.

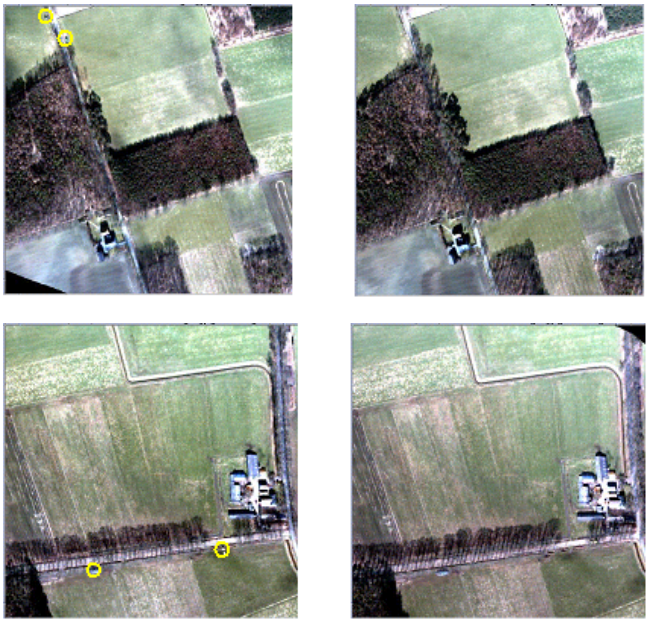

Figure 4. Borculo site. The images on the left show Borculo subsets in time 1 and the ones on the right show the Borculo data at time 2.

Figure 5. Interactive Learning in Dorsten site with models: spectral PCA t_1, spectral PCA t_2, NDVI-Gaussian, texture t_2

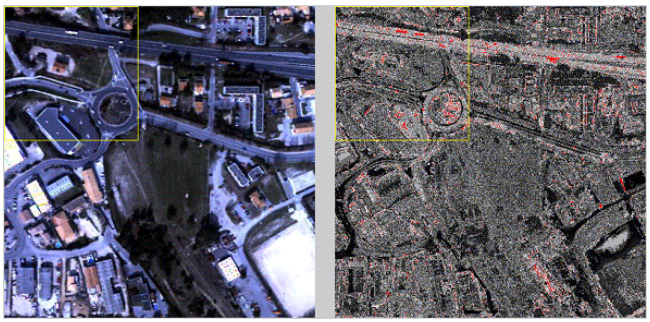

Figure 6. Interactive Learning in Marseille site with models: spectral PCA t_1, spectral PCA t_2, TSAD, TSAD-Gaussian