# SEMANTIC CLASSIFICATION OF VERY HIGH RESOLUTION EARTH OBSERVATION IMAGE CONTENT BASED ON TOPOLOGICAL INFORMATION

Corina VĂDUVA[1], Inge GAVĂŢ[2], Mihai DATCU[3]

*Lucrarea se încadrează în domeniul analizei imaginilor de teledetecţie şi prezintă o metodă ce îşi propune să extragă în mod automat informaţie cu înţeles semantic ridicat ce poate înlocui analiza manuală a utilizatorului uman în vederea înţelegerii şi adnotării semantice a conţinutului unei imagini satelitare. Noutatea metodologiei abordate în aceasta direcţie constă într-o analiză ierarhică a conţinutului imaginii. Trăsaturile spaţiale şi spectrale sunt evaluate în mod statistic pentru a extrage cât mai multă informaţie. Scopul algoritmului propus este de a obţine o clasificare semantică de nivel înalt a conţinutului imaginilor de teledetecţie, astfel încât clasele obţinute să exprime un înţeles concret al scenei pe care o reprezintă. Analiza începe prin extragerea obiectelor de interes de către utilizator prin intermediul sistemului Knowledge Based Information Mining (KIM). Apoi aranjamentele topologice din scenă sunt definite cu ajutorul unor descriptori invarianţi ce caracterizează poziţia relativă a obiectelor din cadrul grupărilor analizate. Aplicând metoda de clasificare k-means acestor descriptori se obţine o clasă nouă ce conţine configuraţii de obiecte cu aranjament spaţial similar. Pentru exemplicare s-a folosit o imagine WorldView2 achiziţionată deasupra oraşului Bucureşti, Romania.*

*In the context of Earth Observation, this paper presents a method whose purpose is to supersede human inductive learning and reasoning in complex scene understanding and characterization by automatic adding high-level ontology to the image. The innovation in this direction of data mining consists in a hierarchical analysis of remote sensing image content. Spectral and spatial information are both assessed in a statistically balanced combination. The aim of the proposed algorithm is to provide high-level semantic classifications for image content. According to a user interest, objects are firstly extracted using a Knowledge Based Information Mining (KIM) System. Further, invariant descriptors for the relative positioning of objects inside a configuration are computed in order to describe topological arrangements inside the scene. New patterns containing spectral and topological information about similar objects configurations are defined through a k-means classification. This approach is exemplified on a WorldView2 scene acquired over Bucharest, Romania.*

[1] Research engineer, Applied Electronics and Information Engineering Department, University POLITEHNICA of Bucharest, Roumania, e-mail: corina.vaduva@gmail.com

[2] Professor, Applied Electronics and Information Engineering Department, University POLITEHNICA of Bucharest, Roumania, e-mail: inge.gavat9@gmail.com

[3] Professor, Applied Electronics and Information Engineering Department, University POLITEHNICA of Bucharest, Roumania, e-mail: mihai.datcu@dlr.de

**Keywords:** high-level semantic classification, hierarchical image modeling, topological analysis

## 1. Introduction

Over the last years, the launched satellite sensors create new challenges in the field of image understanding and data mining. Different perspectives on how to observe and explore our planet are given by the increased amount of information received due to the improvement of spatial, spectral and radiometric resolutions. Innovative approaches are developed in order to support the growing number of applications and requirements coming from the users.

The old pixel-based discovery methods [1] fail to support fully exploitation of all the data recently acquired and the operational applications corresponding to the amount of distinguishing details. New theories are required to process the information inside very high resolution images. The structures visible on the Earth surface are very complex and the objects inside the images have well-defined attributes. Spectral and texture features are not enough anymore for a complete characterizing of the image. In order to understand a certain scene, we first have to assign meaning to its component regions. This is a challenging problem due to the fact that single objects are defined only by their basic features: color, texture and shape. If we try to group objects together, they acquire meaning, according to their relative position and the spatial arrangement inside the scene. For instance, in both images in Fig. 1 there can be identified houses, apartment blocks and very large buildings. While the first scene describes a train station and the afferent halls due to the positioning of the large buildings near a railway in the middle of a residential area, the second image illustrates an industrial area at the town’s periphery.

Fig. 1. a, b: WorldView2 images containing similar areas with different spatial arrangements, corresponding to: a. railway station/train halls; b. industrial area.

Due to the incapacity of previous methods to extract all the information contained inside very high resolution image, the scene analysis is not accurate. A human intervention is thus required to complete the process. Therefore, important research has been done in order to solve this problem of the semantic gap between the original data and the outcome map.

The system proposed in [4] was first to introduce the spatial information concepts in the process of image processing by analyzing fuzzy relationships between regions of the scene. In [3] the histogram of angle was substituted by the histogram of forces [5]. A simplified version of the force histogram underlay the spatial signature presented in [6].

In order to perform a proper conversion from image data to the thematic information required by end users we propose an image processing chain that first extracts objects (pixel level analysis) and then understands configuration of regions (object level analysis) based on the relative positioning of objects in the configuration.

This paper aims to attach topological attributes to the land-cover classification performed using intelligent concepts of interactive learning and probabilistic retrieval of user-specific cover-types, such as the knowledge based image information mining (KIM) system [2], [7]. A signature describing the spatial interaction of objects inside configurations [6] and a k-means classification completes a high-level scene classification. Spectral and topological information are thus combined to bridge the semantic gap between already existing low-level features of the sensors and the desired ontology used by a human operator.

The proposed algorithm is detailed in Chapter 2. Chapter 3 illustrates the results on a very high resolution image. In the end, Chapter 4 will conclude and present further applications.

## 2. Semantic classification algorithm

For an automatic semantic annotation of very high resolution images, the proposed algorithm pursues a hierarchical structuring of the scene content. There are progressively extracted pixel, object and scene level features that characterize component parts of an image and the interactions between them.

Primary, image content can be described by a series of feature vectors. A feature vector normally represents one of the textures, color and shape characteristics computed based on coarseness, contrast, color distribution or directionality. Their number depends on the multitude of methods used to extract the feature vectors. These are considered the low level features of the image and they have no semantic meaning.

Further, by statistically combining these feature vectors through a Bayesian process, the KIM system is used to define regions in the image [2], [7]. A human operator is giving positive and negative examples to the machine in order to define basic semantics in remote sensing images. Therefore, single objects, described as a composition between proportions of different feature vectors are annotated like building, speedway, forest, lake or boat. The regions extracted are considered important by the user, according to the analyzed scene and the requirements of the application involved.

In order to extract real useful meanings from the content of an image and to add high level ontology to the scene it is necessary to define semantic rules inside the scene. All the defined objects are grouped two by two; for each possible pair we compute the spatial signature [6], [8] as a measure for the topology and spatial interaction between the regions included. This signature is based on a simplified algorithm of the histogram of forces and it encapsulates the information about the size, the shape and the distance between the analyzed objects. Moreover, the spatial signature is invariant to the translation, rotation and scaling of the pair examined. Two similar pairs can thus be characterized by similar signatures.

It is imperative to mention that natural landscapes are not composed from perfect geometrical objects. Besides, their regions receive even more irregular shapes due to the incapacity of clustering algorithms when extracting the real contours of regions. Shadows, saturation or just the imperfection of many methods are few of the reasons. Another issue is given, for instance, by the fact that buildings, even if designed the same, they will not form identical configurations. Therefore, in real imagery, the spatial signatures of similar configurations will not resemble perfectly, although they will be very much alike. A classification is thus required in order to identify analogues pairs of regions and assign semantic labels to them.

As the spatial signatures corresponding to pairs of objects are vectors, a k-means classification is the easier method and it provides a good accuracy. The number of classes is randomly chosen depending on the amount of regions extracted by the user, because it is not easy for the human eye to observe a certain configuration and distinguish it from the rest of the objects.

According to this classification, different semantic labels can be attached to every configuration. Residential area, industrial park, sea port are high-level classes that describe the content of remote sensed Earth’s surface.

The diagram in Fig. 2 presents the workflow for the described algorithm. The original scene is introduced in the KIM system. A human operator defines classes of objects considered of interest for a specific scenario case. These objects are grouped two by two. For every configuration, a spatial signature is computed.

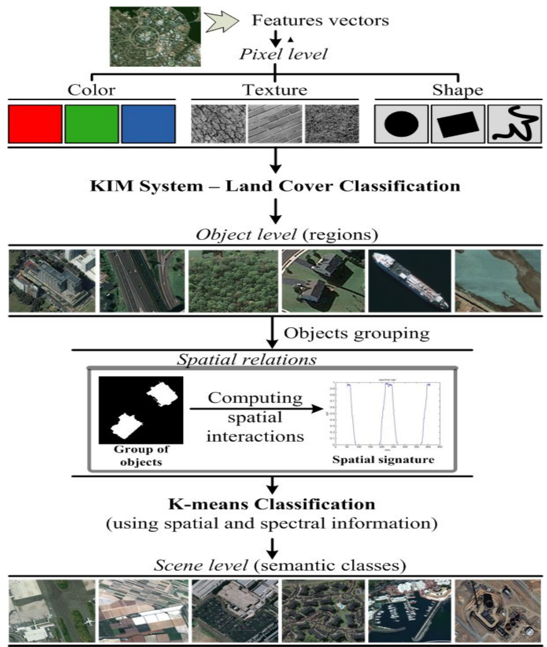


Fig. 2. The proposed algorithm for generating thematic maps with semantic classes based on color features and the spatial relationships between objects.

An unsupervised k-means classification is performed. In the end these classes will be divided in semantic classes, according to the spectral labels of objects in the analyzed pair.

### 3. Experimental results

Test results were provided through an experiment performed on a WorldView2 image (4938x4938 pixels) scene acquired over Bucharest, Romania in august 2010 (Fig. 3). The image's high resolution (2 meters spatial resolution) provides an important amount of distinguishable objects which makes the data appropriate for the experiment.  The 8 spectral bands resolution of the remote sensing image brings a high degree of detail to this classification process.

Fig. 3. WorldView2 image, 2m resolution, 8 bands.

### *3.1. Object level analysis: KIM land cover classification*

The KIM system assigns meaning to the primitive features through a learning phase. Using the samples images (tiles of 500x500 pixel of the whole scene), the operator marks interest areas by giving positive and negative examples, refining the definition of derived feature through an iterative process. Great structures of interest will appear in red on gray scale panel visualization of the scene. Once this system training has been satisfactorily completed, the definition can be saved and used afterwards just by requesting images containing the derived features. Through the interactive learning, further positive and negative training samples bring a newly computed posterior map to the learning applet and show it to the user. When the training is completed, the system is able to export a map of the entire analyzed scene containing the cover type defined.

After the iterative learning process was repeated several times for different types of targets in the present experiment, the KIM system was able to retrieve 13 distinct and precisely defined cover types. The result of the KIM land cover classification is illustrated in Fig. 4.

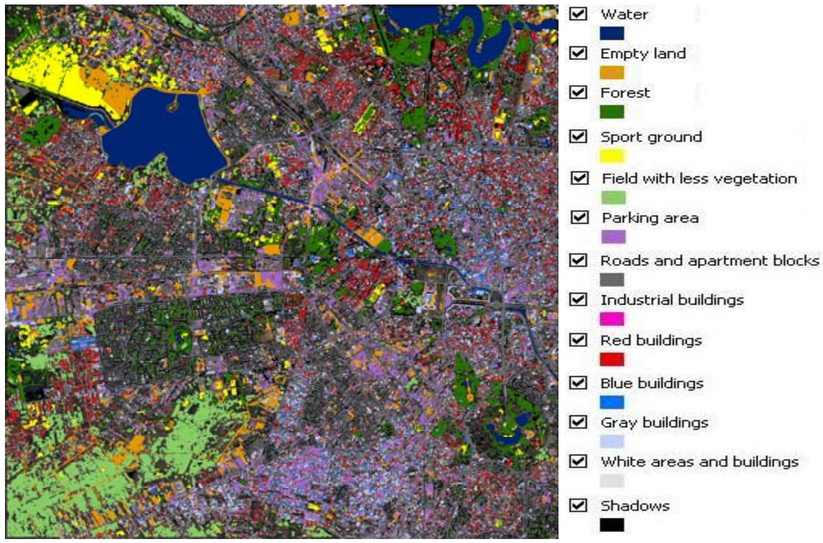


Fig. 4. KIM land cover classification (object level).

The classes defined using the KIM system covers the analysed area in proportion of 90% with an overall precision of approximately 75%. There are 54802 objects belonging to 13 classes automatically extracted with minimum human intervention during the training phase. Even if texture and shape information were also used during the process, the spectral information was prevalent. Therefore, we name these classes of objects spectral classes. Users label them during the training with a general definition. These classes of objects represent the basic level of semantics regarding image content. Their meaning is locally pointed, ignoring the environment. In order to get a general meaning for global understanding of the scene, neighbouring has to be taken into account.

For further evaluation, every user imagines a scenario relevant for his working domain. He chooses only the spectral classes considered to be important for him. For instance, this paper presents a general scenario (Fig. 5) contains 7 spectral classes: water, forest, blue buildings, red buildings, sport grounds, industrial buildings and parking areas and an amount of 470 objects.

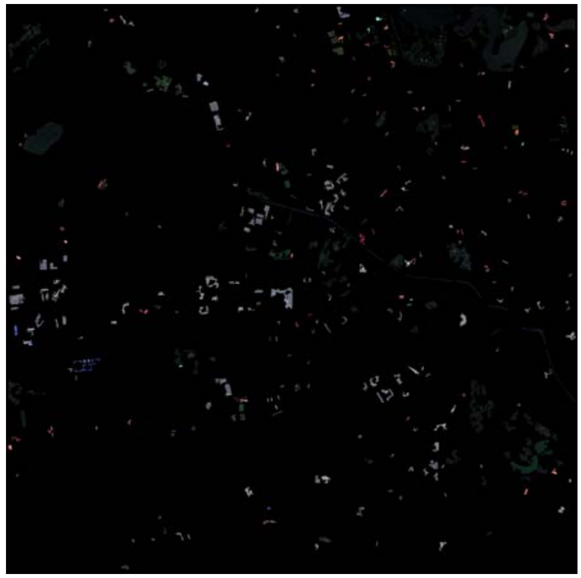

Fig. 5. The scenario described contains 470 objects illustrating 7 spectral classes: water, forest, blue buildings, red buildings, sport grounds, industrial buildings and parking areas.

### *3.2. Scene level analysis: semantic classification*

The first step in scene level analysis requires the grouping of the selected objects in two by two configurations. For every pair, a spatial signature is computed, as a measure of objects' relative positioning to each other. When grouping, in order to engender proper configurations, a threshold was established for the distance between the centroids of the two objects envisaged. It had to be less than one kilometre (500 pixels). Configurations of objects with higher distance then the threshold between them were not taken into consideration.

Therefore, considering these terms, 3891 spatial signatures were computed, as attributes for the topological information and spatial interaction contained inside the scene. They are scene level characteristics of the image content. The k-means classification defines the following topics: classes containing similar spatial configurations represented by similar spatial signatures. In the end these topics will be divided in semantic classes, according to the spectral labels of objects in the group. For instance, one semantic class is represented by all the groups of water and forest having a certain positioning.

*Tabel 1*

**Distribution of objects configurations over the semantic classes**

| | | | topic 1 | topic 2 | topic 3 | topic 4 | topic 5 | topic 6 | topic 7 | topic 8 | topic 9 | topic 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| parking areas | parking areas | group 1 | 9 | 35 | 2 | 4 | 0 | 26 | 1 | 17 | 18 | 8 |
| parking areas | industrial buildings | group 2 | 9 | 30 | 6 | 6 | 7 | 49 | 9 | 29 | 24 | 18 |
| parking areas | sport grounds | group 3 | 1 | 12 | 0 | 0 | 0 | 18 | 0 | 9 | 9 | 0 |
| parking areas | red building | group 4 | 2 | 29 | 1 | 10 | 2 | 56 | 0 | 43 | 16 | 11 |
| parking areas | blue building | group 5 | 7 | 28 | 1 | 11 | 6 | 54 | 3 | 38 | 16 | 8 |
| parking areas | forest | group 6 | 7 | 28 | 5 | 2 | 6 | 20 | 3 | 6 | 7 | 3 |
| parking areas | water | group 7 | 1 | 10 | 1 | 14 | 1 | 12 | 0 | 10 | 4 | 2 |
| industrial buildings | industrial buildings | group 8 | 4 | 22 | 2 | 6 | 0 | 28 | 9 | 13 | 28 | 12 |
| industrial buildings | sport grounds | group 9 | 5 | 30 | 0 | 3 | 1 | 45 | 1 | 9 | 14 | 6 |
| industrial buildings | red building | group 10 | 8 | 36 | 9 | 7 | 12 | 39 | 5 | 34 | 24 | 12 |
| industrial buildings | blue building | group 11 | 13 | 32 | 6 | 6 | 4 | 24 | 10 | 18 | 17 | 12 |
| industrial buildings | forest | group 12 | 3 | 32 | 2 | 7 | 1 | 31 | 3 | 7 | 19 | 13 |
| industrial buildings | water | group 13 | 4 | 9 | 0 | 25 | 3 | 19 | 2 | 8 | 4 | 2 |
| sport grounds | sport grounds | group 14 | 3 | 4 | 1 | 0 | 1 | 8 | 3 | 3 | 5 | 4 |
| sport grounds | red building | group 15 | 7 | 21 | 6 | 6 | 4 | 19 | 22 | 9 | 20 | 10 |
| sport grounds | blue building | group 16 | 4 | 12 | 5 | 1 | 3 | 9 | 6 | 7 | 15 | 14 |
| sport grounds | forest | group 17 | 8 | 22 | 5 | 5 | 5 | 31 | 4 | 8 | 12 | 9 |
| sport grounds | water | group 18 | 5 | 7 | 0 | 6 | 1 | 7 | 3 | 2 | 4 | 4 |
| red building | red building | group 19 | 4 | 30 | 1 | 4 | 1 | 52 | 1 | 156 | 7 | 10 |
| red building | blue building | group 20 | 2 | 35 | 0 | 0 | 1 | 72 | 0 | 148 | 12 | 5 |
| red building | forest | group 21 | 44 | 46 | 51 | 25 | 31 | 30 | 38 | 10 | 41 | 30 |
| red building | water | group 22 | 16 | 14 | 4 | 38 | 10 | 12 | 11 | 20 | 5 | 0 |
| blue building | blue building | group 23 | 2 | 31 | 0 | 3 | 2 | 34 | 1 | 61 | 30 | 16 |
| blue building | forest | group 24 | 33 | 21 | 23 | 29 | 19 | 10 | 17 | 2 | 11 | 23 |
| blue building | apa | group 25 | 7 | 6 | 1 | 7 | 7 | 4 | 2 | 4 | 1 | 1 |
| forest | forest | group 26 | 23 | 59 | 7 | 17 | 6 | 45 | 27 | 4 | 35 | 29 |
| forest | water | group 27 | 23 | 31 | 13 | 14 | 15 | 26 | 30 | 5 | 19 | 16 |
| water | water | group 28 | 2 | 1 | 2 | 6 | 2 | 10 | 4 | 14 | 0 | 1 |

In Table 1 is presented a quantitative assessment of the semantic classification. It reveals 10 spatial arrangements (topics) and 28 spectral grouping, namely 280 semantic classes corresponding to 7 spectral classes and 470 objects. A qualitative evaluation for this classification is illustrated in Fig. 6 and Fig. 7. The two examples present two semantic classes containing industrial building with a certain relative positioning related to a parking area and industrial building with a forest nearby.

Although the number of topics is randomly chosen, because it is not easy for the human eye to observe a certain configuration and distinguish it from the rest of the objects, the overall precision in this case is high, about 95%.

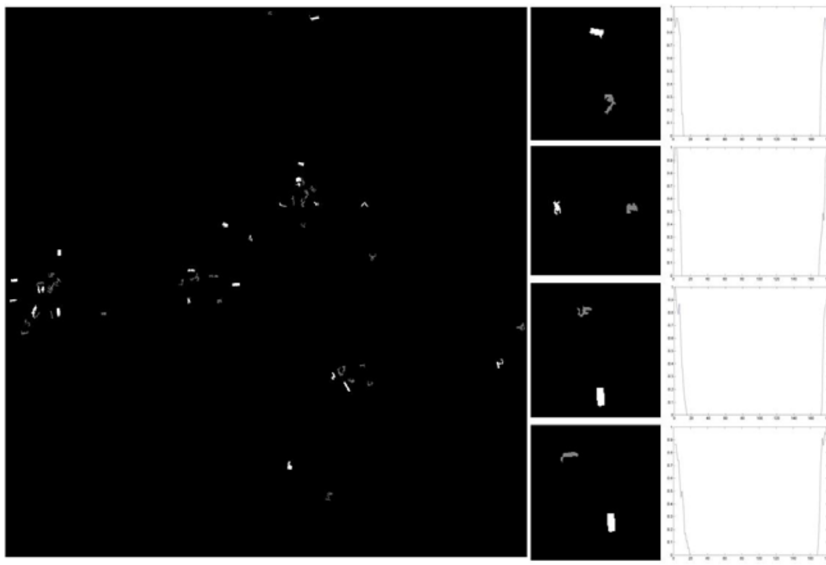

Fig. 6. The grouping “parking areas (in gray) - industrial buildings (in white)” with topic 6 spatial arrangement. Examples of configurations contained in this class and their spatial signatures.

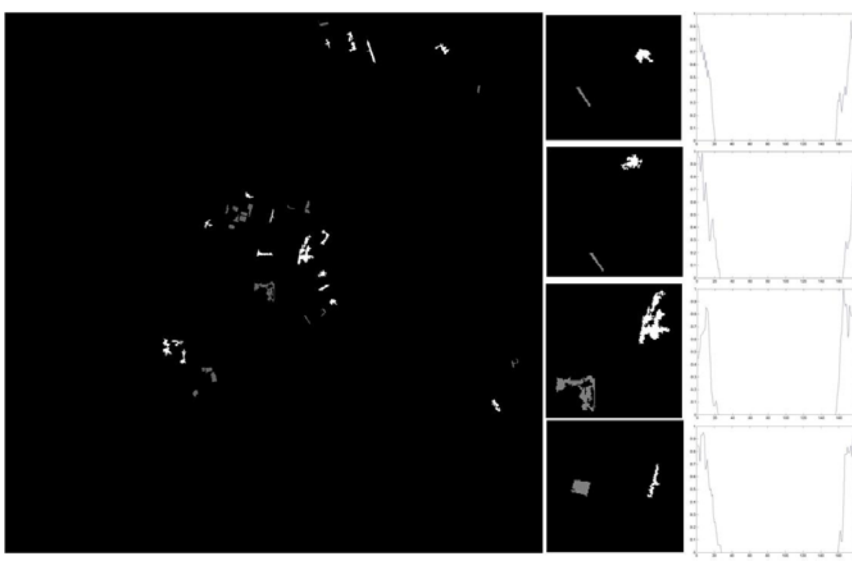

Fig. 7. The grouping " industrial buildings (in gray) – forest (in white)" with topic 2 spatial arrangement. Examples of configurations contained in this class and their spatial signatures.

## 4. Conclusions

Based on a hierarchically structuring of image content, this paper proposes a semi-automatic approach for a semantic classification of a scene for high resolution images. With minimum human interaction, this algorithm can be successfully applied in various applications for a scene-level analysis. It provides complex classification scenarios that can't be obtained with traditional pixel or region level approaches. High level semantic labels can be attached to the image in order to define real meaning of the regions inside the scene. It is possible that one object to be included in several configurations having different semantic class, thereby different meaning. An image can thus receive many semantic labels, according to its content. Based on the meaning of these labels, images can be referred to various application domains (e.g. geography, weather, mining, mapping, defence, etc).

## Acknowledgements

The WorldView2 images were provided by Digital Globe in the frame of the "8 Bands Research Challenge". This research has been developed in the frame of the ESA-PECS project ROKEO – Romanian Centred Knowledge Earth Observation. This work has been supported by the Sectoral Operational Program Human Resources Development (SOP HRD) 2007-2013, financed from the European Social Fund and by the Romanian Government under the contract number and POSDRU/6/1.5/S/16.

## R E F E R E N C E S

[1] *David Landgrebe*, Signal Theory Methods in Multispectral Remote Sensing in John Wiley and Sons, 508 pages, a CD, 2003.

[2] *M. Datcu, K. Seidel*, Human-Centered Concepts for Exploration and Understanding of Earth Observation Images in IEEE Transactions on geosciences and remote sensing, **vol. 43**, no. 3, March 2005, pp.601-609.

[3] *C. R. Shyu, M. Klaric, G. J. Scott, A. S. Barb, C. H. Davis, K. Palaniappan*, GeoIRIS: Geospatial Information Retrieval and Indexing System—Content Mining, Semantics Modelling, and Complex Queries in IEEE Transactions and remote sensing, **vol. 45**, no. 4, April 2007, pp. 839-852.

[4] *S. Aksoy, K. Koperski, C. Tusk,G. Marchisio,J.C. Tilton*, Learning Bayesian classifiers for scene classification with a visual grammar in IEEE Transaction on geosciences and remote sensing, **vol. 43**, no. 3, March 2005, pp. 581-589.

[5] *P. Matsakis, J. M. Keller, L. Wendling, J. Marjamaa, O. Sjahputera*, Linguistic description of relative position in images in IEEE Trans. Systems, Man, and Cybernetics (Part B), **vol. 31**, no. 4, 2001, pp. 573-588.

[6] *C. Vaduva, D. Faur, I. Gavat, M. Datcu*, Data mining and spatial reasoning for satellite image characterization in The 8th Communications International Conference, in Bucharest, Romania, COMM 2010.
[7] *H. Daschiel, M. Datcu*, Design and evaluation of human machine communication for image information mining in IEEE Transactions on Multimedia, **vol. 7**, no. 6, December 2005.
[8] *C. Vaduva, D. Faur, I. Gavat, M. Datcu*, Data Mining and Spatial Reasoning for Satellite Image Understanding in ESA-EUSC 2009, 3-5 November, Madrid, Spain.