Operational Mapping Technology For Generating High-Level Semantic Geo-Information From New Satellite Sensors

Dragos Bratasanu [1], Ion Nedelcu [1], Mihai Datcu [2]

[1] Romanian Space Agency ROSA, Mendeleev 21-25, Bucharest, Romania, 010362, (dragos.bratasanu, ion.nedelcu)@rosa.ro

[2] German Aerospace Center DLR, Oberpfaffenhofen D-82234, Wessling, Germany, mihai.datcu@dlr.de

Abstract – With the launch of the new imaging satellite sensors, content extraction methodologies for scene understanding require updates that aim at bridging the gap between science, operational activities and politics on various levels. In order to fully benefit from these spaceborne sensors, the new directions of research should integrate the language that all parties involved easily understand – *semantics*.

By following the new standards of cartography introduced by EO emergency centers, the automated technology presented in this paper authorizes simultaneous information retrieval, mapping and semantic annotation of the data. The first step is patch-level analysis of the image, this enabling automatic generation of map legends. Following an agreement between the learning model and the spatial resolution of the images, we exploit Latent Dirichlet Allocation model (LDA) to map heterogeneous pixels with no semantic meaning into thematic classes with high-level ontology attached directly by the users.

Keywords — Latent Dirichlet Allocation, semantic annotation, scene understanding, geoinformation

1. INTRODUCTION

The recently launched or planned imaging satellite sensors authorize innovatory methods of image analysis that support the growing number of applications and requirements coming from the users. The data provided by these instruments have increased in spatial, radiometric, spectral resolution faster than the human capacity of exploiting it to the maximum. The procedures that rely on remote sensing data still require the use of a common language to link scientists, operational planners and politicians together - i.e. *semantics*.

The existent automatic content-based retrieval systems solve the content extraction problem only partially, by building pixel-based classification and object-based segmentation models using spectral and textural features. The latent semantic meaning of the scene is usually described by the user with words from a vocabulary but most of the times they don't integrate the connecting link to the visual information in the image. The search engines make use of image descriptors, a set of data-driven features extracted from the image that may not always be directly connected to the objects the user queries for. *The user seeks similarity in semantics while the database can only provide similarity in image processing results* (Smeulders, 2000).

In this paper, the satellite image analysis is taken to the next level - *introduction of contextual information in automatic classification procedures to discover latent semantic knowledge hidden in the data*. The method allows scientists and operational analysts to generate semantic rules that bridge the gap of understanding between science, research, development, operations and politics of the environment. The approach presented in this paper enables simultaneous data mining and semantic annotation using semantic concepts directly applied by the user. While standard processing algorithms classify image pixels based on low-level features (i.e. spectral bands, texture), the method described here clusters image pixels in a latent semantic space, available only at the human level of understanding. By employing the statistical model Latent Dirichlet Allocation (LDA) adapted from the text domain (Blei, 2003), similar pixels that belong to the same information class are grouped together under a semantic label provided directly by the users. As a text document that contains many words can be regarded as belonging to a single topic of interest (e.g. politics, news, science), so too a region of the satellite image can contain multiple pixels that belong to a single class (e.g. Urban Areas). Grouping together pixels belonging to the same information class under the same high-level semantic label leads to discovery of the semantic rules that bridge the processing layers, from primitive features with no semantic meaning to high-level human-centered maps of information.

The paper is organized as follows: section 2 describes the concept and motivation of contextual modeling applied to sub-meter resolution optical satellite images, section 3 explains the LDA model and the correspondence between the text and image domains, section 4 is the experimental setup and evaluation of results on WorldView-2 and RapidEye data and section 5 ends the paper with conclusions.

2. CONTEXTUAL MODELING OF THE SCENE

The methodology described in this paper is based on *patch-level analysis*, capturing *contextual spectral information* of a limited spatial environment. Sub-meter resolution image areas interconnect complex structures covering many pixels with high diversity of spectral information. The use of contextual information in multimedia image retrieval has led to a substantial increase in the accuracy of results (Parikh, 2009) and several researchers have adopted similar approaches to satellite image understanding (Datcu, 2003). People automatically use contextual information to recognize objects in the scene (Torralba, 2007) and in some cases the objects are detected solely by using the contextual information, even though the appearance of the objects themselves is withheld. This effect is called blind recognition. The approach to discover hidden information from the surroundings relying on powerful cognitive background is used by experienced analysts in many military and geo-intelligence applications.

Satellite images have high complexity of structures in the scenes and the higher the complexity the greater the likelihood of it benefitting from the context (Datcu, 2005).

Figure 1. Patch-level analysis of satellite images

Patch analysis is a balanced combination of spectral and spatial information and it is based on investigating the spectral signatures of objects in a limited spatial environment (figure 1). By using this approach we line up to the way users create cartographic products for multiple applications (e.g. maps for emergency response, geo-intelligence, forensics) (www.zki.dlr.de). Another motivation for the patch-level analysis of satellite images is the similarity of the method to the *quadrat analysis of maps*, widely used in GIS. The quadrat analysis is performed by dividing the area of interest into cells of equal size and by investigating the statistics of each patch. The size of the cells will influence the observed distribution (Mitchell, 2005).

3. LATENT DIRICHLET ALLOCATION

The problem of modeling heterogeneous pixels in a spectral map to describe the corresponding information classes following the one-to-one or many-to-one rules is performed using the Latent Dirichlet Model (LDA) described in (Blei, 2003), (Mochihashi, 2004).

3.1 Theoretical Background

LDA is a generative probabilistic model for collections of discrete data such as text corpora. Generative models are random sources that can generate infinite sequences of samples according to a probability distribution. LDA was created to describe large collections of digital text documents and recently applied in classification and semantic annotation methods of satellite images (Lienou, 2008), (Bratasanu, 2010). The LDA model is a three-level hierarchical Bayesian model, in which each document in the collection is modeled as a finite random mixture over a 'latent' set of topics. Each topic in turn is modeled as a probability distribution over a set of words in the vocabulary. In applying this text model to annotate satellite images, we need to define the analogy between the terminologies used. The text-based model is using the following three levels of data descriptors: text corpus (the collection the documents), documents and words in the vocabulary.

* Word = basic unit defined to be an item from a vocabulary V
* Document = a sequence of N words denoted by $d = (w_1,...w_N)$
* Corpus = collection of M documents denoted by $D = \{d_1,...d_M\}$

The documents are represented as a sequence of *N* words w_n from the vocabulary. The model discovers 'latent' topics (aspects) and assigns each word in the document to one of these inferred topics. Therefore, by introducing a new layer of information between the words-level and documents-level, each document is represented as a probability distribution over the set of topics and each topic in turn as a probability distribution over the words in the vocabulary. The number of topics to describe the documents is pre-defined by the user according to the application. Considering the vocabulary $W = \{w_1,..w_N\}$, LDA assumes the following generative process for each document in the corpus:

1. Choose a K-Dimensional Dirichlet random variable θ ~ Dir (α), where K is the number of topics in the collection.

2. For each of the word positions $n \in \{1 \ldots N\}$:

* choose a topic $z_n \sim Multinomial(\theta)$
* choose a word w_n from $p(w_n \mid z_n, \beta)$

The likelihood of document W with such a model is given by:

$$p(W \mid \alpha,\beta) = \int p(\theta \mid \alpha)\left(\prod_{n=1}^{N}\sum_{z_n} p(z_n \mid \theta) * p(w_n \mid z_n,\beta)\right)d\theta \qquad (1)$$

Finally we obtain the probability of a corpus:

$$p(D \mid \alpha,\beta) = \prod_{d=1}^{M}\int p(\theta_d)\left(\prod_{n=1}^{N_d}\sum_{z_{dn}} p(z_{dn} \mid \theta_d) p(w_{dn} \mid z_{dn},\beta)\right)d\theta_d \qquad (2)$$

LDA has the flexibility to assign probabilities to documents outside the training corpus thus being the basic tool for performing supervised classification over unknown documents. We use this property to classify and annotate all patches in the satellite image.

3.2 Semantic Knowledge Discovery In Satellite Image

In this section we classify and annotate satellite images into high-level information classes with semantic meaning provided by the user, both at pixel and patch level. For this purpose, the following connections between the text and image domains are required:

- A visual-word w_n is a cluster from an unsupervised k-means classification applied on the image features. Because information in the satellite image is usually represented on 11 to 16-bits, the vocabulary performs a simple reduction across the image features. A number of K=30 clusters is optimum for this task.
- A visual-document W is a patch of the image with variable size, according to the spatial resolution.
- The visual-corpus is the satellite image to be annotated. All the documents (patches) together yield the corpus (scene).

The LDA model works under the bag-of-words assumption, in which the order of words in the document is ignored and the image is represented as a random sequence of *N* visual words. Each visual-document is described by a distribution over visual-words in the form of frequency-count vector (histogram). LDA models each word in a document as a sample from a mixture model, where the mixture components can be viewed as representations of latent 'topics'. Each document is described as a probability distribution over latent topics and each topic in turn is a distribution over a fixed set of words from the vocabulary. Given a satellite image to be annotated using semantic concepts – target class and *'others'*, for each topic the user provides a set of patches, which will be used for learning. The annotation is the classification of unseen documents into the semantic classes.

4. EXPERIMENTAL SETUP

4.1 WorldView-2 Database Semantic Annotation

For the case studies presented in this section we created a small database of three WorldView-2 images but the set used may be increased with the level of accuracy. The data used is depicted in figure 2.

Figure 2. WorldView-2 test database

To perform simultaneous classification and semantic annotation of these three images in the base, *Water Bodies* is chosen the target class. Since the regions imaged in the data have suffered severe flooding in the last years, this method proves to have also valuable operational support.

The first step is to create a common visual-vocabulary for the database. To do this, we perform unsupervised k-means classification over the whole dataset, with K=35 clusters or visual words, depicted in figure 3. Applying the unsupervised classification over the whole data set establishes a common understanding of the spectral space – the same clusters represent the same visual-words.

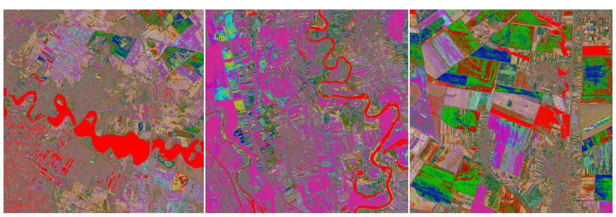

Figure 3. Common vocabulary of visual-words

All three images contain the information class *Water Bodies.* The next step is to train the LDA model with visual-documents representing the target class and documents labeled as *others. A very important aspect of this procedure is the fact that learning is performed on a single image in the set and the classifier discovers similar latent classes in all images available in the database.* Thus, we trained the LDA model with 5 documents sized 100 X 100 pixels chosen only from the first image and labeled *Water Bodie*s and another 30 documents labeled *others.*

Classification is performed over the entire data set, simultaneously at four different levels: pixel level and document level (25x25, 50x50 and 100x100 pixels), as shown in figures 4-7. All the pixels and documents in the output maps are labeled with the words defined by the user in the training step. The ontology thus used allows understanding across a wide variety of user communities, from scientists, researchers, to intervention teams and policy makers. Each pixel in the output pseudo-image is a visual topic described by a probability distribution over the visual-words in the vocabulary. The target class topic is depicted in black in the color map and its value is linked by an index to a semantic table.

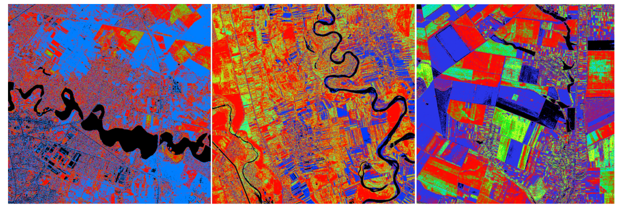

Figure 4. Pixel-level representation of visual-topics. *Water Bodies* topic depicted in black

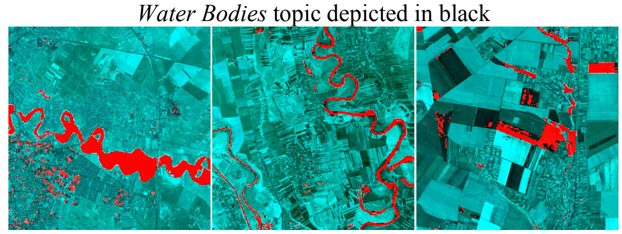

Figure 5. Document-level representation of visual-topics 25x25 pixels

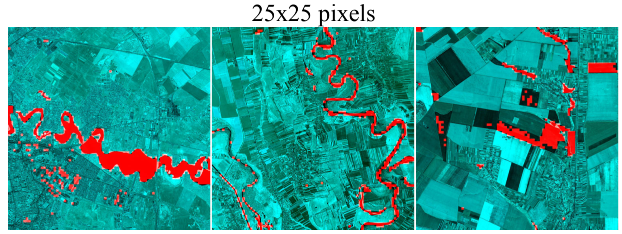

Figure 6. Document-level representation of visual-topics 50x50 pixels

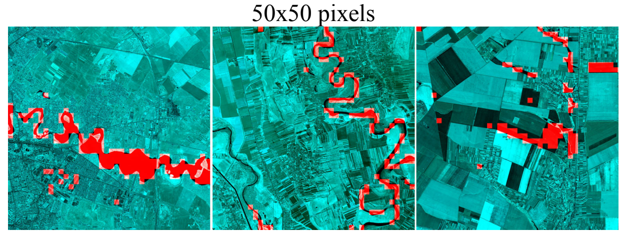

Figure 7. Document-level representation of visual-topics 100x100 pixels

The above results present an interactive approach to simultaneous classification and semantic annotation of VHR satellite image using concepts defined by the user. Latent Dirichlet Allocation model discovers latent semantic knowledge in previously unseen images. The key point in obtaining high performance is feature selection to ensure high dissimilarity between the target class and the others. The accuracy of this workflow reaches 94% for the first image, 93% for the second image and 75% for the last image. The commission errors in the first and the last maps emerge due to the high similarity in spectral signatures for the clusters in the target class and in other regions. These errors do not appear in the annotation step (LDA) but when the visual-vocabulary is defined. Different semantic classes may yield high spectral similarity and k-means groups them in the same cluster even though they are different in the semantic space.

4.2 Time Series Analysis Using RapidEye Data

The simultaneous classification and semantic annotation using the LDA model can be also applied to time series to map changes in a specific class and discover latent knowledge in previously unseen images. Learning the target class is performed only on the first available image and then the annotation is extended across the whole data set. In this case study the target class is *Forest Areas.* The first step is to automatically create the vocabulary by ingesting all images in the system and clustering them using k-means (k=35). The user trains the LDA model with 5 patches sized 200 X 200 pixels representing the target class, selected from the first image only, and labels them with the required semantic ontology. Classification and annotation is performed over the entire dataset and masks are extracted as depicted in figure 8.

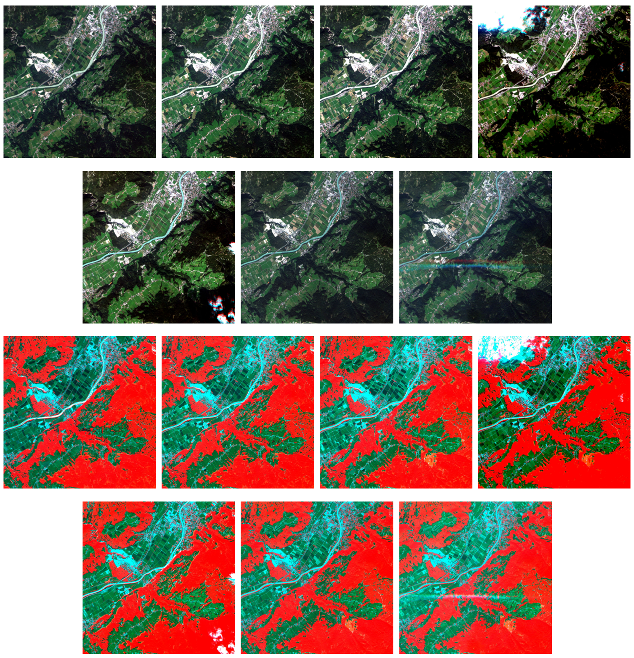

Figure 8. Upper rows: RapidEye images dated 29.07, 16.08, 20.08, 05.09, 22.09, 30.09, 03.10.2009
Lower rows: Forest Areas Map for all the images.

The accuracy of results is analyzed by Precision-Recall metrics, and presented in table A below.

Table A. Precision-Recall Accuracy Analysis For RapidEye

Image	Precision	Recall	False Positive Rate	True Positive Rate
29.07.2009	0.95	0.96	0.08	0.96
16.08.2009	0.90	0.95	0.10	0.95
20.08.2009	0.90	0.95	0.10	0.95
05.09.2009	0.90	0.95	0.10	0.95
22.09.2009	0.90	0.95	0.10	0.95
30.09.2009	0.90	0.95	0.10	0.95
03.10.2010	0.90	0.95	0.10	0.95

5. CONCLUSIONS

This paper concludes that the LDA model may successfully be employed to discover semantic rules inside satellite image databases and used for automatic scene understanding and content retrieval. A flexible generative probabilistic model for collections of discrete data, LDA is an efficient model for simultaneous semantic classifications and annotation on both the patch and scene levels. This paper presents experiments on WorldView-2 and RapidEye and exhaustive studies have been performed on multiple optical sensors (Landsat, SPOT, Quickbird, Ikonos, GeoEye-1, Formosat) with highly accurate results. The only varying parameter between experiments on different sensors is the size of the patch, as a function directly linked to the spatial resolution of the image. As the planned spaceborne imaging sensors will be launched soon (ESA Sentinels, SPOT 6 and 7, Pleiades), this tool can easily be adapted (in some case without any change) to operate on this data.

ACKNOWLEDGEMENT

The research presented in this paper is in the frame of ESA-PECS Romanian Knowledge Based Earth Observation (RoKEO) project. The authors would like to thank Digital Globe for the data provided for this research.

REFERENCES

D. Blei, A. Yang and M. Jordan. Latent Dirichlet Allocation. Journal of Machine Learning Research 3. pp. 993-1022. 2003

D. Bratasanu, I. Nedelcu, M. Datcu "Bridging the semantic gap for satellite image annotation and automatic mapping applications", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2010

M. Datcu and K. Siedel, "Human-Centered Concepts for Exploration and Understanding of Earth Observation Images", IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no.3, pp.601-609, 2005

M. Datcu, H. Daschiel, A.Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M.Pastori, K.Seidel, P.G. Marchetti, S. D'Elia, "Information Mining in Remote Sensing Image Archives – Part A: System Concepts", IEEE Transactions of Geoscience and Remote Sensing, Vol. 41, No. 12, Pp. 2923-2936, 2003

M. Lienou, H. Maitre and M. Datcu, "Semantic Annotation of Satellite Images using Latent Dirichlet Allocation", Proceedings of Fifth ESA-EUSC Conference on Image Information Mining. 2008. http://earth.esa.int/rtd/Events/ESA-EUSC_2008/

Any Mitchell, "The ESRI Guide to GIS Analysis. Volume 2: Spatial Measurements and Statistics", ESRI Press, 2005

D. Mochihashi. "LDA: a Latent Dirichlet Allocation Package". ATR Spoken Language Communication Research Laboratories, Kyoto, Japan. http://chasen.org/~daiti-m/dist/lda/, 2004

D. Parikh, "Modeling Context For Image Understanding", PhD thesis submitted to Carnegie Mellon University, 2009

A. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain. "Content-Based Image Retrieval at the End of the Early Years". IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.2. No. 12. Pp.1349-1380. 2000.

A. Torralba, R. Fergus, W. Freeman, "Tiny images", MIT Technical Report, 2007

www.zki.dlr.de