TOOLS FOR VALUE ADDING BY SEMANTIC CODING: THE EARTH OBSERVATION IMAGE LIBRARIAN

M. Datcu [1,2], M.Costache [2], G. Schwarz [1], Amaia de Miguel [1]

[1] **German Aerospace Center (DLR)**
Oberpfaffenhofen, D-82234 Wessling, Germany
email: mihai.datcu@dlr.de
[2] **Paris Institute of Technology, Telecom Paris**
46 rue Barrault, F-75 013 Paris, France

ABSTRACT

The progress of Earth Observation (EO) meter and sub-meter resolution sensors, both for optical and Synthetic Aperture Radar (SAR) images, poses major problems for the optimal exploitation of the image information content. While, in the past decades the exploitation of the acquired EO images with resolution of tens of meters was limited to maximum of 5% from data acquired, with the advent of higher resolution sensors, this figure could decrease. Thus, we need a new generation of concepts and tools to add value to EO images, to facilitate the access to the image content, in directly understandable manner. The Image Librarian is a concept to enable interpretation of contents, to associate it with other information, to create semantic categories, to understand user inquiries, to conduct dialogues with the user to refine expression of needs, to comment and inform users about data content, and to suggest the most appropriate interpretation alternative. The Image Librarian is a proposal for the next generation of EO value adding intearctive tools.

High resolution images (0.5-2m), mainly of sites or objects in relation with human activities, are difficult to interpret since they make 3D structures visible, thus increasing the complexity. Their understanding is depending strongly on the spatial contextual information, and on the interpretation context.

Methods like physical parameter retrieval, image classification or segmentation, cannot any more support interpretation. Thus, the Image Librarian concept aims at elaboration of methods to help users to interactively, semi-automatically, and in a time effective manner, to interpret images, based on semantic learning and coding of image structures and objects. This is an interactive paradigm to decompose images into meaningful objects or structures using the user interaction for a limited number of examples and generalizing the results for large images, or image archives.

The proposed concept is based on the principle of semantic compositionality: *the meaning of a whole is a function of the meanings of its parts and their mode of syntactic combination.*

The semantics of images or image objects can be gained using a hierarchical process to link the data to the possible meaning of their contents:

1. An image is decomposed into atomic elements (visual words) from 3 perspectives: pixel oriented primitive attributes, attributes of regions obtained by image segmentation, and signs or localized patterns detected in the image.
2. All elementary components (visual words) can be grouped based on their attributes and their geometric and/or topological properties. At this stage, the image will be encoded in a set of configurations representing possible image structures or objects. The configurations are coding both descriptive and neighborhood information.
3. At the next level, these configurations will be analyzed to rank their relevance and generate categories based on similarity criteria.

For the last level, semantics will be induced in an interactive learning process by a user pointing to “interesting” structures in the image. The learning process will select and group existing image structures in configurations, considering their

categories, and generalize results over entire images. This step is a semantic inference resulting in a semantic coding of the image content.

The result of the first 2 processing steps is a set of semantic image categories and their associated classes, regions/patches and patterns, i.e. visual words. This new level of grouping, at level 3, is mainly spatial, i.e. topologic, but it should take into consideration all properties of salient features. The syntax rules of spatial grouping can only be partially extracted from the data properties. Hence, a supervised learning mechanism shall be implemented at this level.

Based on these concepts, two methods have been implemented.

In a first approach, Support Vector Machines (SVM) and Bayesian methods are used for category-based semantic learning. The visual semantic structures or objects in high resolution images requires in general huge amounts of training data. This will slow down the interpretation process and make it difficult. To learn faster, and from only few examples, it is suggested to take advantage of knowledge coming from previously learned similar and dissimilar categories. We propose a Bayesian implementation of this method in line with the proposed hierarchical concept. Image structures or object categories are represented by statistical models. A priori knowledge is acquired by learning the parameters of these models. The models describe the grouping of primitive features to semantically code image structures and to create categories for these learned structures. The categories are generated at an intermediate level, hidden to the user; however, the structures are learned based on the user interest and semantics.

In a second apparoach, the Latent Dirichlet Allocation (LDA) method was generalized for generation of semantic topics in EO images. LDA was originally introduced for text analysis. In its imaplemenation for image semantic coding the „words" are obtained by clustering primitive attributes of images, as as first procesing step. In the next 2 steps, a generative model enables the grouing of the words, for the inference of the semantic topics in the image. This apprach is beyond the classical image classification, since it generates „categories" by grouping entities dispersed in the feature space.

These concepts have been implemented in a series of tools dedicated to interactive operation and web services of the European Space Agency (ESA), as their Service Support Environment (SSE) and the the Knowledge Centered Earth Observation (KEO) systems.

The tools and feature extraction web services can operate on Landsat, SOPT, SPOT 5, Ikonos, Quick Bird and TerraSAR-X data. The primitive features used are the spectral signatures, total variation, texture parameters and for many applications only the mean and variance of the images.

The tools have been demonstrated for recognition of semantically complex scenes as urban, rural infrastructure and natural structures with applicabilty to rapid mapping and general image understanding tasks.

[1] D. M. Blei, A. Y. Ng and M. I. Jordan, (2003): Latent Dirichlet Allocation, *Journal of Machine Learning Research*, vol 3, pp. 993-1022

[2] Datcu, Mihai; Schwarz, Gottfried (2007): The Earth Observation Image Librarian. *Conference Proceedings PV 2007,* Oberpfaffenhofen, Germany.

[3] Costache, Mihai; Datcu, Mihai (2008): Semantic image content annotation based on Bayesian categorial learning and SVM classification. *ESA EUSC 2008: Image Information Mining*, Frascati, Italy.

[4] Lienou, Marie; Maitre, Henri; Datcu, Mihai (2008): Semantic Allocation of Satellite Images using Latent Dirichlet Allocation. ESA EUSC 2008: *Image Information Mining*, Frascati, Italy

[5] http://services.eoportal.org/

[6] http://earth.esa.int/rtd/Projects/KEO/