IMAGE SEMANTIC CODING USING OTB

Marine Campedel and Marie Linou

Institut TELECOM - TELECOM ParisTech - CNRS LTCI
46 rue Barrault - 75013 Paris - FRANCE
marine.campedel@telecom-paristech.fr

1. INTRODUCTION

Semantic coding is used to provide image representation able to make semantics emerge and to learn it. This process is based on a basic decomposition of the image content into "visual words"; using spatial relationships, these visual words are aggregated to form structures and (inter)active learning is used to associate semantics to the emerging structures. A complete OTB tool has been developed to illustrate the whole process; it allows feature extraction and "visual words" production, as well as LDA and SVM learning approachs.

2. VISUAL WORDS EXTRACTION

The classical way to characterize optical satellite images is to cut the image into overlapping patches. These patches are then used to compute features that are significant of their spectral, textural and geometrical content. In the context of visual words extraction, the idea is to find representants of the whole set of features that can be considered as atomic (like words in a text). In the presented tool we extract basic features like means and variances computed on a window around each pixel, in each spectral band. The produced clusters are then vectorially quantified using KMeans algorithm. The clusters labels are called "visual words". Each region of the image can be considered as a bag of visual words.

3. LEARNING STRATEGIES

3.1. learning machines

Some tools, initially developed for statistical text modelling in large document collections, have recently been used for semantic image annotation, but also object recognition, scene classification, and image retrieval. Actually, generative probabilistic models such as probabilistic Latent Semantic Analysis (pLSA) [1] and Latent Dirichlet Allocation (LDA) [2] have been exploited for this purpose. Generative probabilistic models are random sources that can generate infinite sequences of samples according to a probability distribution. Using the *bag-of-words* assumption that the order of words in a document can be neglected, probabilistic Latent Semantic Analysis was one of the first methods that provided a probabilistic approach towards modelling text documents as mixture of intermediate (hidden) topics (or aspects). Latent Dirichlet Allocation extends pLSA model by treating the topic mixtures parameters as variables drawn from a Dirichlet distribution, thus defining a complete generative model which overcomes some limitations of pLSA. Such approachs have already been used to annotate whole natural images. We propose to extend this application to satellite imagery.

3.2. Interactive learning

Relevance Feedback (RF) is now well-known [3] to reduce the semantic gap and help the user to precise its query. RF approaches are based on three main tools: i) a learning tool (for example a 2-classes classifier like SVM, or a LDA classifier to learn the target class), ii) a selection tool and iii) an interactive interface. The selection tool decides what should be presented to the user while the interface module chooses how to present it, in order to help the user annotate examples (imagelets in

Thanks for funding to CNES/DLR/TELECOM ParisTech Competence Center in the field of Information Extraction and Image Understanding for Earth Observation.

our case). In the context of satellite images, imagelets are components of a bigger image that cannot be displayed with full resolution on screen. Hence we defined a strategy allowing active learning on a reduced portion of the image and extending this learning to the whole image.

Support Vector Machines (SVM) are very powerfull learning machines [4], used in many classification and regression tasks. These methods are kernel-based learning machines and they have the ability of performing high accuracy classification for different type of data starting only from small number of examples for each class. The definition of a kernel is equivalent to the definition of a feature transform, allowing linear separation in the newly defined feature space.In [5], Ferecatu applied successfully classical CBIR with SVM-based approach to satellite images. He put in evidence the interest of the SVM output to identify 3 categories of data: those which are classified as positive, those classified as negative and those which are ambiguous. Annotation of ambiguous as well as erroneously classified data should be encouraged.

In our OTB tool we propose to compare a SVM active learning approach (based on tf-idf signatures of the visual words inside the imagelets) to a LDA one. In the first case, the user has to annotate examples belonging to the target class (called positives examples) as well as ”rest-of-the-word” examples (negatives examples). In the second one, only positives examples are required. Similar performances have been obtained on experiments designed to illustrate the ability of these approaches to learn ”complex classes” (urban area, industrial area, golf terrain, ...).

4. OTB TOOL

The proposed tool is called cocSemanticCoding. It is available at the following address: www.tsi.enst.fr/ campedel/SemanticCoding. As mentioned before it is able to compute basic features and estimate visual words from KMeans clustering. It largely benefits from the vectorial image format defined in OTB, as well as filters like KMeans and SVM. Based on the ”visual words” representation, active learning procedures can be used (SVM or LDA-based) to annotate the image using a target class. The active learning is first performed on a reduced portion of the image using ImageViewer capabilities and then extended to the whole image when local satisfaction is reached.

5. CONCLUSION AND PERSPECTIVES

Semantic coding is able to provide a synthetic and efficient representation of satellite image content. Modelling of pixel neighborhoods as ”bags of words” enable the use of very basic features and well-known learning procedures (like SVM and LDA) originally applied in the context of textual documents retrieval. Of course it is not able to learn any kind of semantics and we plan to study knowledge-based approach to better integrate spatial relationships in the representation.

6. REFERENCES

[1] T. Hofmann, “Unsupervised learning by probabilistic latent semantic analysis,” *Journal of Machine Learning Research*, vol. 42, pp. 177–196, Jan. 2001.

[2] A. Y. Ng D. M. Blei and M. I. Jordan, “Latent dirichlet allocation,” *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[3] D. Djordjevic, M. Grzegorzek, K. Chandramouli, T. Piatrik, F. Hopfgartner, J. Nemrava, and M. Campedel, “Report specification of proposed implementation of implicit and explicit user relevance feedback,” nov. 2007.

[4] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20(3), pp. 273–297, 1995.

[5] M. Ferecatu and N. Boujemaa, “Interactive remote-sensing image retrieval using active relevance feedback,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 4, pp. 818–826, April 2007.