Indexing and Retrieval in Large Satellite Image Databases

Henri Maître

GET/Télécom-Paris, Paris, France
CNES-DLR-ENST Competence Centre on *Information Extraction and Image Understanding for Earth Observation*

ABSTRACT

We are presenting a global scheme devoted to the management of very large databases of remote sensing images, with a specific attention to high resolution images. We show at first the specificities of indexing satellite images. Then we present the different stages of a global system, mostly based on textural feature detection and SVM classification, for indexing and retrieving parts of satellite images with given properties. We show the role of the semantic information extraction and we propose and experiment three different manners to introduce such a semantic information in the query system and to adapt the data to the user's demand.

Keywords: indexing, retrieval, image databases, remote sensing, satellite images, Earth observation, clustering

1. INTRODUCTION

A grat number of Earth Observation (EO) satellites have yet been launched. Even recently, CosmoSkymed I and II, TerraSar-X, were launched just the last 2 months before this paper is written ; they add their high capacity of image acquisition to the very large number of existing EO satellites which permanently orbit around the Earth and cover our world with images. The number of available images is growing day after day, but our capacity to process and exploit them is remaining almost constant because it mostly relies on human experts. Human expertise being the result of a long and expensive training, is at best stagnating in many countries.

Not only the number of observing satellites is increasing, but also the satellite agility makes them more efficient in acquiring images, and each image is richer in pixels because of the sensor integration progress, each pixel being coded with more bits. Therefore, the volume of image archives is growing in an unconcontrolled manner. One satellite only often delivers more than one terabyte of data a year. The storage problem itself is an important issue, but it is not addressed here. This paper is interested in the way to allow an efficient access to these images. In the most conventional way to access to EO image databases, the satellite is first chosen and the acquisition mode of the sensor. Then, most of the existing systems today are using as a reference to any image, two pieces of information only : the date/time of acquisition and the scene geographic reference. In some cases the image may also be selected on a few criteria like its quality (in terms of contrast, saturation, acquisition defects, etc.), or the cloud ratio. In order to make a larger use of stored images, it would be necessary to keep track of the image content. Browsing the image descriptions, the user may select those images which are of interest for his/her application. This problem, known as indexing, is becoming one of the central problems in managing databases, and especially image and multimedia databases. We will concentrate on this problem in this paper. At first we will underline the main differences between indexing multimedia databases (a very popular domain where an abundant literature exists) and indexing satellite images (far less addressed).

In Section 2 we present the problem of EO image indexing as depending both on the user's demand and on the very nature of satellite images ; then, in Section 3, the choice of low level primitives is discussed, as well as classification and evaluation of performances. We also address the important issue of resolution and time sequences. Then, in Section 4 the difficult problem of inferring high-level information from the low-level features extracted at the signal level is covered with some first results.

Further author information:
Henri Maître: E-mail: Henri.Maitre@enst.fr, Telephone: +33(0)145817655

2. IMAGE INDEXING AND RETRIEVING

2.1 Multimedia image retrieval vs. remote sensing image retrieval

A vast amount of literature has been devoted to the image indexing and retrieval topic in the recent years, mostly under the pressure of the multimedia demand. Literature may be distributed according to several different categories. Two ways to classify these works are to distinguish, firstly :

- **the exact retrieval** of a given item as as for instance "Mona Lisa" or "Fred Astaire", where exact may be mitigated by possible variations (rotation, scale, illumination, affine or perspective transforms, ...) ;
- **the categorical search**, where the user is interested to retrieve a member of a family as for instance the search for "smiling children" or "black dogs".

then :

- the search in **open databases** as the web or personal archives, where "open database" stands for an *a priori* unlimited archive with many different kinds of documents from uncontrolled origin of unknown format, size and quality ;
- the search in **closed thematic databases** as for instance medical or astronomical archives where there is no risk to find a face of a baby, or a photo from a football player among the documents of the collection.

In the case of remote sensing images we are dealing with "categorical search in a closed database". The exact search is not of great interest for remote sensing applications, since the unique position of one precise object is usually perfectly known and easily addressed by its coordinates on the Earth. The database may be considered as "closed" since all the image sources are perfectly known and usually well documented, allowing to discriminate most of the documents without opening the files. There is no confusion for the user on the type of archive which may bring the solution to his/her query. The kind of information brought by SAR images, by hyper spectral sensors, by meteorological or high definition satellites are so different that there is no ambiguity on the one which may provide the correct and best answer to any well formulated demand. Moreover, using the image references, the range of dates and the list of countries of interest may be very fast isolated, limiting even more the candidates.

Being "categorical" and addressing "closed archives", makes the query task rather different in Remote Sensing from its counterparts in multimedia for instance. These differences are reflected by the techniques which are employed, by the protocols to build the complete information management system, but also by the level of confidence which are demanded to this system. For instance, in EO, a research is made to obtain not only some samples of the category of interest but, as most as possible, all the instances.

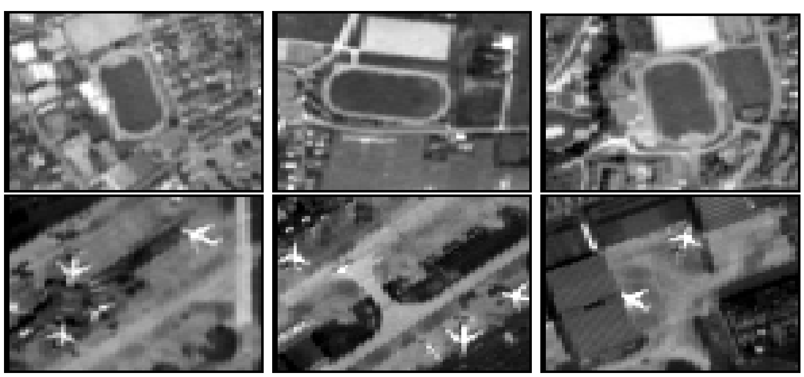

FIG. 1 – Some instances of two classes of images : sport-places (top) and planes on airports (bottom) as seen with SPOT 5 at 5 m resolution

2.2 What is a request in Remote Sensing ?

The Remote Sensing domain is supporting a large number of different applications with various possible demands : defense and security, forest and agriculture, land management, natural and man-made risk survey, ecology, meteorology, ...are typical fields where the demand for efficient information management systems is strong. What kind of questions may be asked ? Here are some instances which demonstrate the variety of possibilities :

- the oak forests degraded by a given parasite or disease,
- the parts of mega-cities suffering from an uncontrolled growth,
- the area where it would be possible to build a water purification station or a provisional landing terrain,
- the list of open-air cultures which have been replaced by green houses in the last year,
- the candidate implantations of a gas-pipe or a power-line,
- the localization of all the golf-greens, close from a marina and a beach.

The first and very important remark which has to be done, is that the field of possible demands is not limited and not known before-hand. The purpose of satellite image indexing is to allot to every acquired image this layer of indexes allowing an efficient future access to the data.

There are two different ways to ask a question in image data mining, which are being differently addressed by the information management system :

1. **query by example**, where the user is providing an instance of the kind of image he/she is looking for (for instance an agricultural region with green houses) ;
2. **query by language**, where the user expresses his/her demand with words. Language may be natural (if the user is allowed to write his request in his own words, with sentences and free vocabulary) or constrained, with constraints varying from a fix menu of chosen words to more complex systems where words may be qualified with attributes or associated together with syntactic rules.

In the following, we will be interested in both strategies. The former allows to derive purely automatic methods without recurse to any kind of semantics. Most of the methods presented in Section 3 will be convenient for it. But it is obviously limited to the search for items which have at least one known example. "Query by example" allows the use of abstract descriptors which are selected by the machine as the most efficient but are not required to be understood by the user.

It is not the case when querying by language, at least in the most evoluted case where the user is allowed to formulate his request in his own words. To facilitate query by language, a simple solution is to make use of **annotations**, i.e. to add a set of keywords to describe the image content. Widely used in multimedia where each image only contains a few number of items of interest, annotation is inefficient in remote sensing because of the extremely large number of different objects which may be distinguished in the very large satellite scenes. For some applications, **captions** may extend the concept of annotation introducing richer descriptions than isolated words. Captions require some syntactical analysis of the text which makes their automatic interpretation less simple than annotation. Both annotations and captions are today mostly made by human interpreters and not automatically deduced from image analysis. Therefore, "querying by language" is much more complex and Section 4 will mainly devoted to provide possible solutions to solve it.

2.3 A channel coding approach to image indexing

Within the scope of the project here reported, the link between an image and the user, is considered as a communication channel.[2] The flow of information is dictated by Shannon theory : each step is a coding which is supposed to convert the incoming flow into another one with a smaller bit rate but almost the same information content. This idea was developed using a two stage scheme by Datcu[1] (Fig. 2). The two-stage process is introduced in order to take into account the two aspects of information : in a first stage, the best of the signal information is extracted and coded in an efficient way, in the second stage, the specific demand from the user is translated into requirements which drive the query. While the first stage is mostly image and signal processing, the second one is syntactical and semantical.

3. TOWARDS AN EFFICIENT LOW LEVEL IMAGE INDEXING

: In the sequel, we will concentrate on visible range satellite images with a main focusing on high to very high resolution images (i.e. better than 10 m).

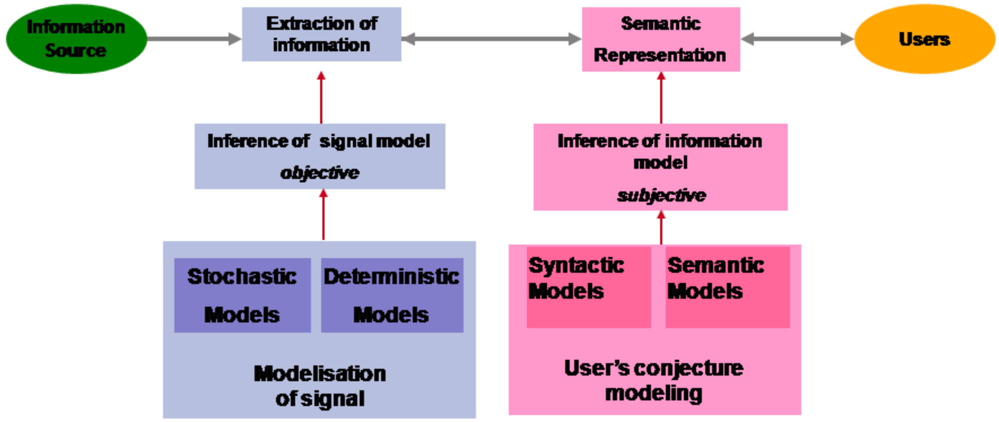

FIG. 2 – Image processing as a two-stage communication channel as proposed by M. Datcu. The first stage takes into account the low level image processing operations when the second translates the high level user's demand.[1]

3.1 Indexing features

The choice of indexing features received a great attention in the image processing literature. Features depend on the type and resolution of the sensor, as well as on the task we want to achieve.

In the domain of multimedia for instance, features like SIFT descriptors[3] associated with strong interest points[4] have been widely used[5] when looking for an exact retrieval. For categorical retrieval one often prefers shape and color descriptors issued from segmented regions.[6]

With multi-spectral or hyper-spectral low resolution sensors (Noaa, Landsat, Vegetation), efficient features are issued from the radiometry of pixels and often combine channels in complex indexes like Vegetation index (NDVI), Brightness index (BI) or Urban index (ISU), and even more complex when more channels are involved.

With high resolution pan-chromatic images (resolution better than 10 m per pixel, like SPOT), textural properties are known to be highly discriminating.[7,8] We obtained very good performances with features reflecting the image textural properties. Moving a window with fixed size (typical size of 64×64 pixels or 128×128) at regular places through the image (Fig. 3), a vector of textural descriptors is measured. Several features have been compared which behave almost the same :

- Haralick cooccurrence matrix descriptors,[9]
- Gabor filters,[10–12]
- Quadratic Mirror Filters (QMF).[13,14]

TAB. 1 – A typical classification result on the dataset of 6×600 high resolution images (SPOT 5), using textural features when using Gaussian kernel SVM with 80 % cross-validation and 145 features per vector.

	city	clouds	desert	fields	forest	sea
city	98.8			0.5		
clouds		99.3	0.2			
desert			99.0	0.3		
fields	0.5	0.2	0.8	98.1	0.3	0.4
forest		0.2			98.0	1.4
sea	0.7	0.3		1.	1.7	98.2

Used in conjunction with radiometric features, textural features provide excellent performances as may be seen for instance on Tab. 1.

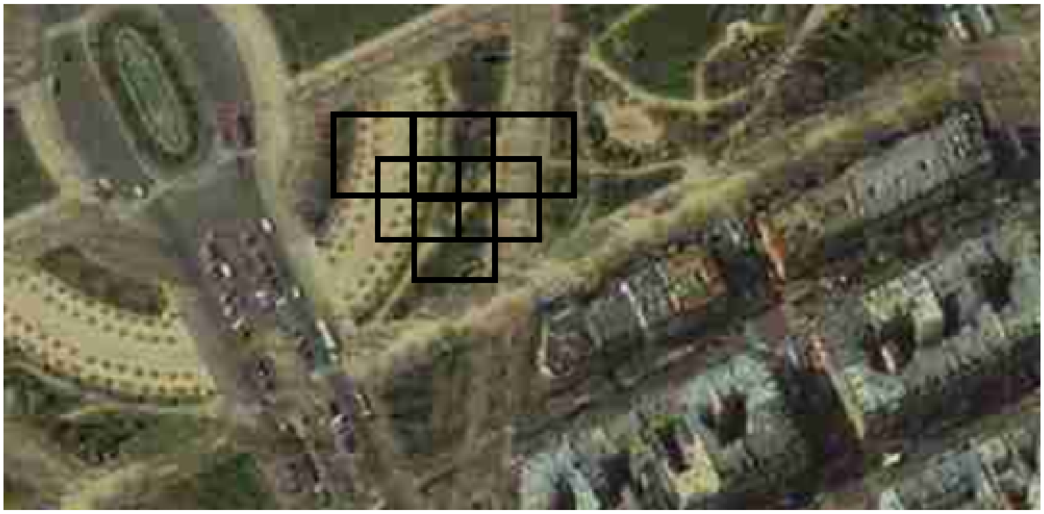

FIG. 3 – In the proposed scheme, the image is decomposed in overlapping windows (of typical dimension 64×64 or 128×128). From each window a vector of texture is measured which describes the image content. In this case a regular set of feature vectors is obtained.

In the case of very high satellite images (Ikonos, Quickbird), texture features are not the most efficient. Features reflecting the structure of objects in images are more adequate to describe the image content, especially in case of man-made objects. Experiments are still under development for this problem, but promising results have been obtained with structural features reflecting for instance the road network density,[15] (see an instance in Fig. 4).

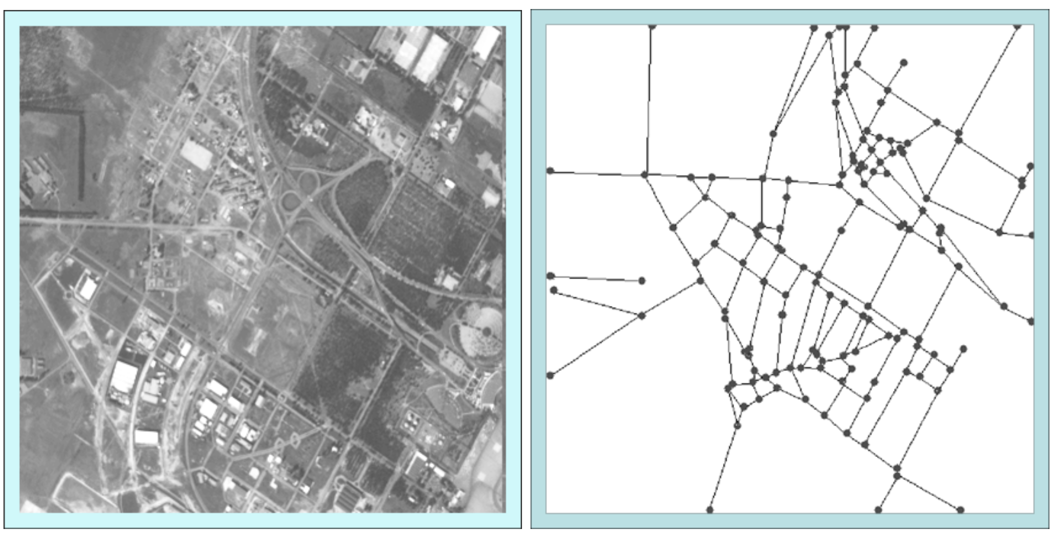

FIG. 4 – Structural features are very convenient to describe man-made sceneries for very high resolution images. Here a graph representation is used to reflect the road network. Contrary to the case of Fig. 3, an irregular description results from this detection. From.[15]

3.2 Classification of features

A description of the image content in satellite imaging makes use of a very large number of different concepts :

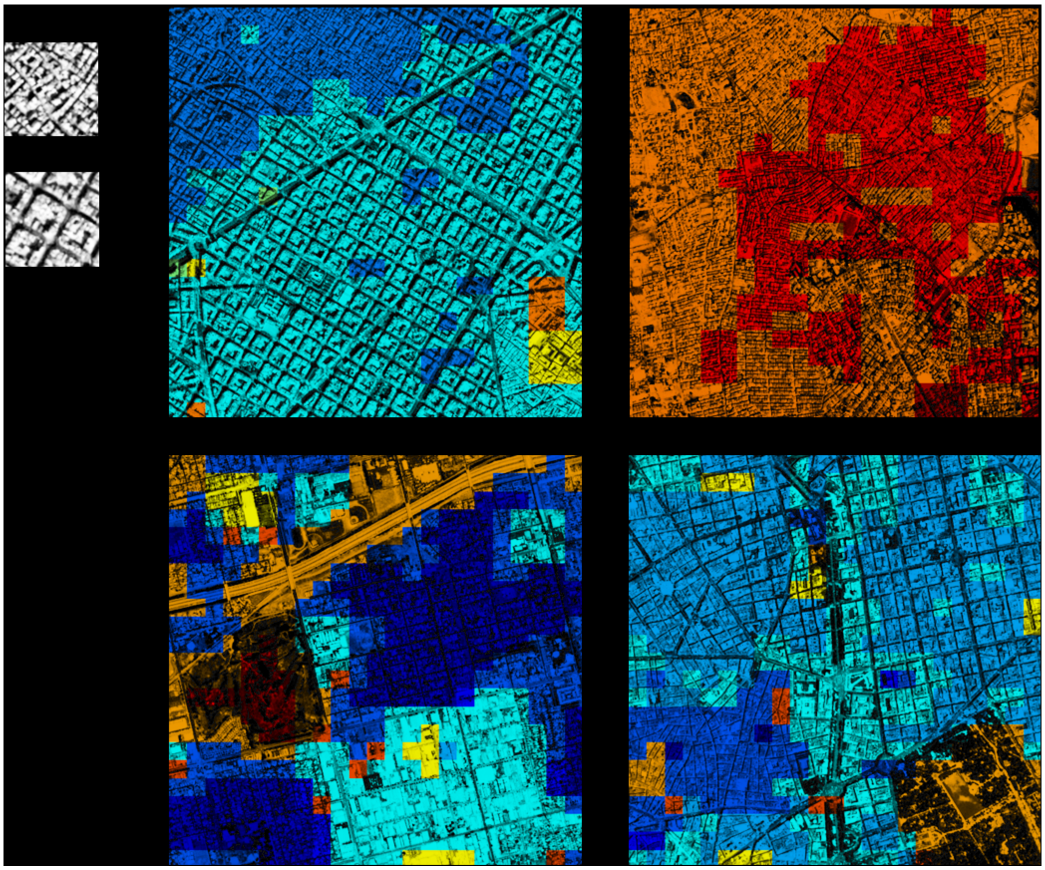

FIG. 5 – Consensual classification of cities. The original clustering algorithms were : K-Means, Kernel-K-Means,[16] Spectral K-Means,[17] Ward hierarchical clustering, AutoClass.[18] The four cities here presented are (from left to right and top to bottom) : Barcelona, Istanbul, Los Angeles and Madrid (images from SPOT 5 with a 5 m resolution - Copyright CNES). The 2 samples on the left present an enlargement of the two classes in blue, present for instance on Barcelona. From.[19]

– geomorphological landscapes like desert, savannah, mountains, water occupations, etc.,
– vegetation, either natural : various forests, meadows, bushes, or man-made : fields, cultures, vineyards, orchards, wastelands, ...
– various types of cities, villages and settlements, and man-made plants : factories, sport-camps, transportation networks, factories.

In such a hard context, it is difficult to propose an ideal set of classes and categories which would cover all these concepts. On the contrary, the description requires to use a more complex description as for instance provided by an ontology. The development of such an ontology for satellite image processing has been developed in some domains like for instance in the Cosem project*, but is not yet available for general purpose remote sensing.

The recurse to many different classifiers to exploit a vast amount of attributes issued from a very large set of samples is natural. Different classifier have different criteria to select clusters : some are looking for compact masses of samples, some other are looking for separation frontiers, some other are based on hierarchical description or exploit proximity links between clusters. When no statistical model is consistent to describe the set of data, we may expect that every classifier is making the best of a part of the existing information on the data. A nice track is to look for a consensus in the classifiers to reveal the most stable classes existing in the data. Such a project has been followed for instance by Fred and Jain[20] or Lange.[21] We followed the idea in[20] to make use of the coassociation matrix which represent in a single matrix the *probability* of a sample to be clustered with any other sample by any of the classifiers. But instead of using an information theoretic criterion to extract the most consensual clusters from this matrix, we have chosen to use a root mean square distance to an ideal classifier.[19] The benefit of such a criterion is the possibility to determine the optimal classification without storing the very large coassociation matrix in the case where we have millions of samples (as in the satellite imaging case). On Fig. 5 is presented the result of consensual classification of several urban centers in a study which was interested in extracting the most typical urban tissues from 10 different cities in the world.

For general purpose classification, several powerful tools have been investigated. However, SVM classifier[22] proved to be quite efficient to learn discriminant rules in large decisional spaces from very small numbers of samples. As our experiments deals with feature spaces of dimension 10 to 200, and as training is a quite tedious task, in the following, when no mention is made, classification is obtained using SVM with Gaussian kernels.

3.3 Performance testing

In order to test the efficiency of these descriptors and to optimize the choice of parameters, a database of samples with homogeneous and known content is built, from images issued from SPOT 5 satellite with resolution 5 and 2.5 m. This database has two components :

1. a large set (600) of images belonging to a small number (6) of broad and well distinct classes (i.e. *desert*, *fields*, *forest*, *clouds*, *city*, *sea*) ;
2. a small set (between 20 and 100) of images belonging to a large (about 50) set of narrow classes, providing typical instances of very characteristic sceneries (for instance : *commercial mall*, *parking lot*, *cemetery*, *green houses*, *stone quarries*, ...) (a typical sample set of such images is presented on Fig. 1).

The total database contains more than 7500 samples and it is regularly augmented. Performance testing results in detection error rate (a typical error rate of 1.4 % ± 0.4 % for the experiment reported in Tab. 1), confusion matrices (as in Tab. 1) or Receiver Operational Characteristic ROC curves (as in Fig. 9), depending on the need.

3.4 Feature selection

As said before, using in conjunction radiometric and textural features provides excellent results for data mining tasks. However, the obtained vector has about 150 dimensions, therefore storage is heavy and computation time consuming. In order to reduce the dimensionality one is used to proceed with feature reduction. Several techniques have been known as efficient for feature reduction, either as *wrappers* or as *filters*.[23,24] Filters are mostly based on the statistical properties of data, without reference to any classification, when wrappers are attached to the performances of a given classifier. While the second are often more efficient, the first ones present many advantages for satellite image indexing since they are unsupervised and able to deal with very large amounts

*COSEM : Semantics for the Coastal Zone - http ://cosem.erc.msstate.edu of data. We have developed a very efficient filter method, called KMeans-FS[19] which consists in grouping the features (and not the data) by a conventional K-means algorithm, then replacing a set of features contained in a cluster by its centroid. This method is easily extended to kernel K-means[25] (providing the KK-Means-FS method) and may also be used in conjunction with Support Vector Clustering (SVC) techniques.[26] A typical example of effect of the K-Means-FS method is presented on Fig. 6.

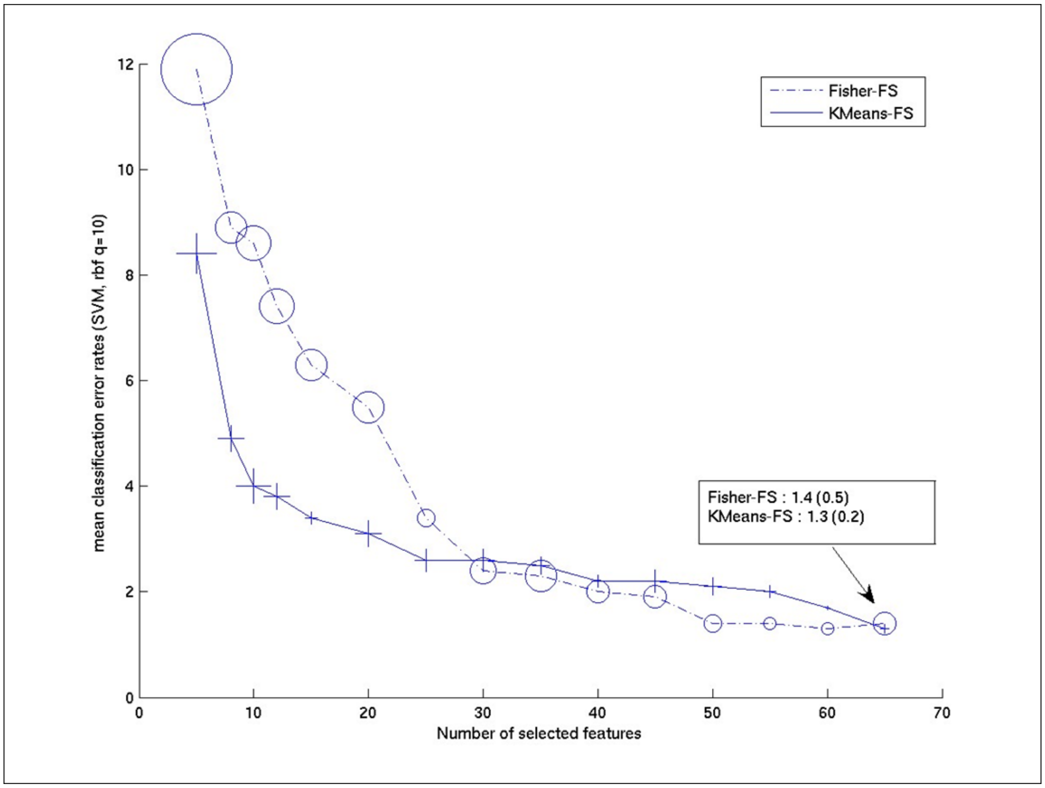

FIG. 6 – Experiment of feature reduction using the KMeans-FS method or the Fisher-FS method (based on Fisher discriminant analysis[27]). Starting with 68 features, KMeans-FS allows to reduce the number of feature to 10 without degrading too much the performances.

3.5 Role of scale in image indexing

3.5.1 Interpolating textural features

The many observation satellites orbiting around the Earth have different resolutions and even when high resolution or very high resolution images are concerned, the size on the ground of a pixel ranges from half a meter to ten meters. How is it possible in these conditions to retrieve a given type of landscape which may have been seen by different satellites ? How is it possible to train a classifier with images of different resolutions or, when training with one resolution only, to retrieve with similar performances images from another satellite ? Using textural features to index images implicitly involves the image resolution in the indexing vector, and therefore a solution to convert a feature measured at a resolution to the one which would have been seen at another.

Indeed it is possible to efficiently interpolate features[28] from one scale to another, making the retrieval almost as efficient as if the training were done at the final resolution. We have shown that, in order to get good performances, we have to know, not only the satellite resolution, but also the PSF of the camera (or at least should we have some good guess of this PSF). This point is well illustrated in Tab. 2, where the training stage was made with a linear SVM classifier on SPOT5-TMA Images, and the recognition stage transported on Quickbird (both panchromatic or multi-spectral images) or THR-SPOT5 images.

TAB. 2 – Correct classification scores with wavelets or Haralick cooccurrence matrices. A linear SVM classifier is used with different satellites (QuickBird with Panchromatic or Multispectral images, or SPOT 5 with Very High Resolution or TMA images) on the same regions, after a training on SPOT 5 TMA images only. In the case where $p = 0$, we ignore the PSF of the satellite, in the case where $p = 0.5$, we consider as a PSF a Gaussian function with $\sigma = 0.5r$, where r is the ground resolution.

Image type resolution	QB-Pan 0.61 m	QB-Mul 2.44 m	SPOT5-THR 2.5 m	SPOT5-HMA 5 m (c.v.)
wavelets $p = 0.5$	84.1 %	83.0 %	92.0 %	96.4 %
wavelets $p = 0$	84.1 %	78.7 %	92.0 %	96.4 %
Haralick cooccur. matrix	71.4 %	72.3 %	88.0 %	95.8 %

3.5.2 Resolution independent characteristic scale

However, if it is possible to retrieve correctly objects of a given scale in satellite images whatever the satellite resolution, it is of course necessary that the satellite resolution be better than the dimension of the object of interest. The determination of the scale of the objects contained in an image is therefore an important problem in image indexing. When the objects have a unique and fixed size, this problem is not too difficult. More challenging is the case of many different objects with almost similar sizes, which give raise to the **characteristic scale** determination problem which provides a unique figure to characterize a homogeneous area. It is a problem of great interest to provide size information for instance for the houses in a suburb or the gardens or fields in countryside.

On this problem also we provided a very robust solution. Starting from the linear scale space determination method as proposed by Lindeberg,[29] we proposed to modify it by using a Total Variation norm[30] instead of $\mathcal{L}^2$ norm, and taking into account the optical PSF of the sensor.[31]

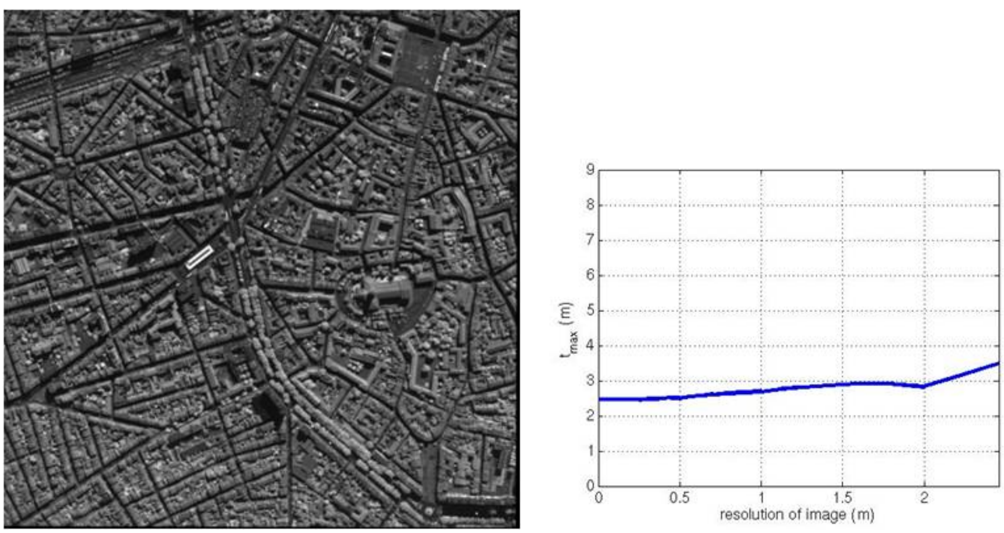

FIG. 7 – The city of Toulouse (as seen by SPOT 5 with 5m resolution Copyright CNES) and its :characteristic scale as measured from satellite images with resolution ranging from 1m to 8 m. Taking into account the satellite resolution **and** its PSF, the :characteristic scale appears almost the same for all the satellites.[31] A value of 2.4 corresponds to a characteristic scale of 15 m.

3.5.3 Map of scales

Characteristic scale is an average measure on a homogeneous area. We may also provide local measure of scale at each pixel position. What we call the **local scale** is the scale of the homogeneous region which contains this pixel. The proposed method is based on the decomposition of the image into its level set, using the fast level set transform,[32] and the estimation of the most significant object which contains each pixel.[33] A result of such a map is presented on Fig. 8.

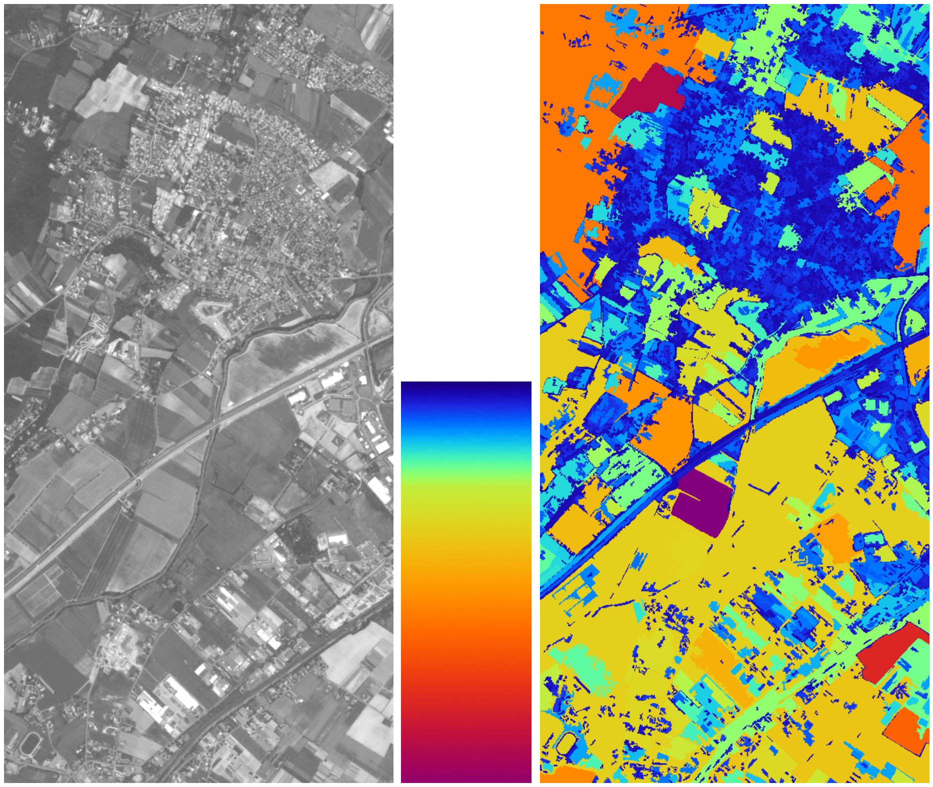

FIG. 8 – Around Toulouse (as seen by SPOT 5 with 5m resolution Copyright CNES) and its map of scales. The color chart ranges from 3.5 m (deep blue) to 50 m (red).

3.6 Time series

In the Earth Observation literature most of the events of interest are issued from the analysis of successive images in time : evaluation of earthquake damages, pollution survey, city growth planing, agricultural planning, etc. Therefore, a great interest emerges from indexing, not only still images, but also time series as obtained, for instance, by repetitive satellite pass over a region. A great attention was paid to the tracking of region in time, to the modeling of their evolution and the detection of the most characteristic events which appeared along the successive observations.[34–36] Here again an information theoretic approach, combined with a spatio-temporal analysis of well registered images allows to determine the most efficient coding as a compromise between the fidelity to information content and the volume of representation.

4. FROM LOW LEVEL TO HIGH LEVEL : ACCESS TO SEMANTICS

:

Low level satellite image processing is rather efficient. It allows the extraction of features of interest for indexing and retrieval, but as said in Section 1, it hardly fits the user's demand to access to data at a natural language level. To improve the information adaptation to a human expert, we should progress in the identification of the extracted signal with known and used concepts as manipulated by this expert, and to translate most of the detected elements into their semantic counterparts. There exist several ways to achieve this evolution :

1. one is to improve the refinement of the categorization of each detected object by using more powerful recognition tools (using context dependency and specific knowledge) ; in this way we may go from *a contrasted rectangular blob of 10* m^2 to a *vehicle on a road* and finally to a *private car.*
2. another is to create new concepts, not detected at low level, by grouping local pieces of information to address descriptions at a larger scale ; for instance an *airport* is made of detected objects like *planes*, *barracks*, *runway*, etc. which may be locally detected by low-level features.
3. and of course a mix of both approaches.

When indexing satellite images it is necessary to gain in two directions : **spatial grouping** is necessary for obvious efficiency reasons (a modern satellite image size is typically 20 000 $\times$ 20 000 pixels and its description must cover very large areas with a few words to allow efficient queries), **semantic coding** is necessary because of the appropriateness of words to cover rich and subtle concepts in good agreement with the user's demand (a *fishing port* is definitively different from a *sailing port* but just differs in one word which carries all the differences for any expert).

The access to the semantic level requires "the man to enter in the loop", since meaning (in the sense we are interested in) is only the result of man's activity. We propose below 3 different ways to access to the semantic level, where both refinement and contextual grouping are used to access at higher level descriptions.

4.1 Relevant feedback

Relevant feedback is a classical way to benefit from the user's experience in the training stage of supervised recognition. It often starts from presenting some samples of the kind of landscape or objects we want to retrieve. At this level, the user may explicitly introduce some form of semantic (for instance the name of the objects he is looking for, its relation with existing objects or categories of objects, etc.). The system builds a first classifier from its own knowledge of what exists in the database and presents some results which are supposed to answer the demand. The user is asked to confirm (relevant positive answer) or infirm (irrelevant negative answer) the correctness of each sample classification. According to an internal strategy, the machine corrects its classification and presents new results. The training stage is repeated until the user is satisfied with the results. The final classification is done with the last optimized parameters.

Within this project an efficient relevance feedback system has been developed,[37] based on SVM and Bayesian classification and making use of a memory to keep track of the yet recognized classes. Every new object is modeled in the multi-dimensional feature space as a mixture of Gaussians, the number of Gaussians is decided by MDL[38] (Minimum Description Length) or Kernel MDL[39] and the parameters of each Gaussian are optimized by EM (Expectation Maximization).[40] Fig. 9 presents a typical ROC curve for such a learning engine.

4.2 Semantic information hiring from existing maps

A lot of semantic information is also contained in maps which provide interpreted elements available to enrich an indexing system as soon as a bijection is made between each element on the map and its existing counterpart in the image. Cartographers have established rules and norms to attribute words to represent most of the natural or man-made objects which appear in satellite images. Moreover they extracted the network of concepts which allows efficient reasoning with these concepts. In our experiment, we benefited from this approach by using the Corine LandCover database[41,42] as a source of semantic knowledge,[43] Fig.10.

4.3 Inference of high level information from examples

The last and most difficult way to derive semantic level information from low level information is through examples. This track has been explored for remote sensing applications thanks to the KIM system by a specification of the clustering.[1,7] Following some approaches which emerged some years ago in the text processing context,[44] it is also possible to infer semantic information by spatial reasoning, grouping local low level features in larger concepts.[45] In such an approach, instances of high level concepts (as for instance *industrial suburbs* or

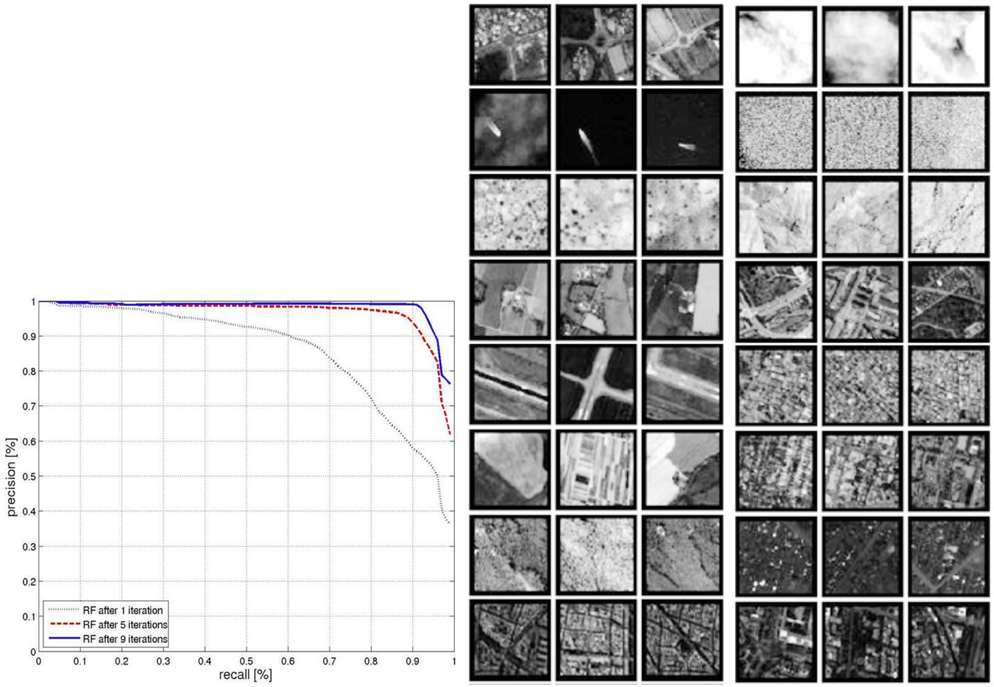

FIG. 9 – Left : Receiver Operational Characteristic for the relevance feedback system,[37] we see the improvement of the classification after 1, 5 and 9 iterations. The result is the average for a classification problem on 16 classes with 100 samples per class, each sample being used once as a prototype. 3 samples of each class are presented on the right. From top to bottom, middle columns : Traffic turnaround, boats, Savannah, village, airport, fields, forest, right columns : clouds, sea, desert, city (Copenhagen), city (Istanbul), city (Los Angeles), city (La Paz), city (Madrid), city (Paris).

residential suburbs are given to the machine which determines the most relevant low-level features, their probability, their spatial distribution (by learning model parameters) by EM mechanisms. After having learned the most frequent models, the system is able to describe an unknown scene as a mosaïc of models each model being the most adequate to fit the local properties and to propagate through a greedy algorithm. The arbitration between models is made using an information theoretic criterion, combined with MDL. An instance of such a classification is presented on Fig. 11.

5. CONCLUSION

Satellite image indexing is a vast and difficult problem. It requires a series of cascaded algorithms based firstly on a clever determination of where information lies, and secondly on an efficient use of the extracted information to meet the user's requirements. Unfortunately, satellite images, because of their number and volume will not be indexed several times in their life-time, and certainly not at the user's demand. Indexing must be performed when the image is received and should be adapted to a question which will only be formulated 5 or 10 years later. Therefore, indexing is not prepared for a specific query but it should be open to a broad type of questions.

The proposed scheme here presented is made of hierarchical layers of information. The lowest layer is very close from the image, regularly distributed, dense and abstract. Additional layers are introduced to gather spatially neighboring low-level features forming by the way areas with a more informative content, where semantic may be added to help the user searching in the database. Above this intermediary level a global description of the image is necessary, which could be used for a query in natural language. The articulation between the top levels is to be made on the ground of concepts linked together as words in an ontology representing the observed world. This difficult task is under development.

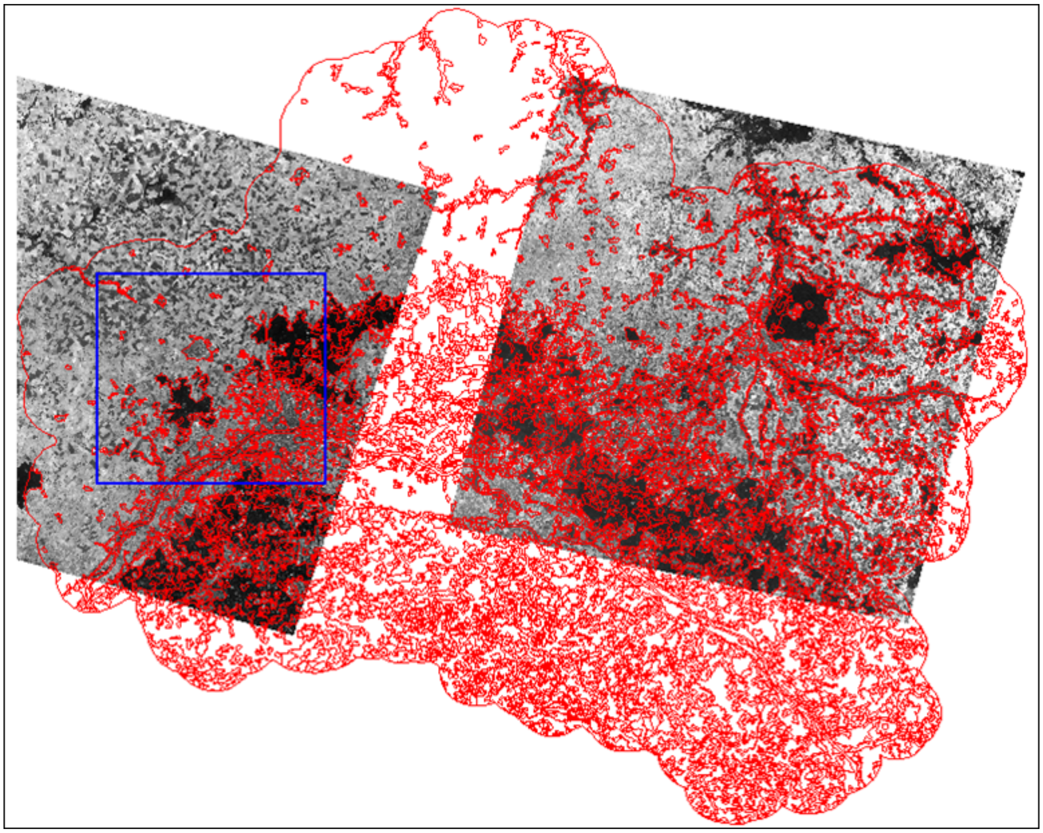

FIG. 10 – Super imposition of two SPOT 2 scenes with a Corine LandCover map (in red) of the French department of Loiret.[43] The training is made on the blue region and extended to the whole images.

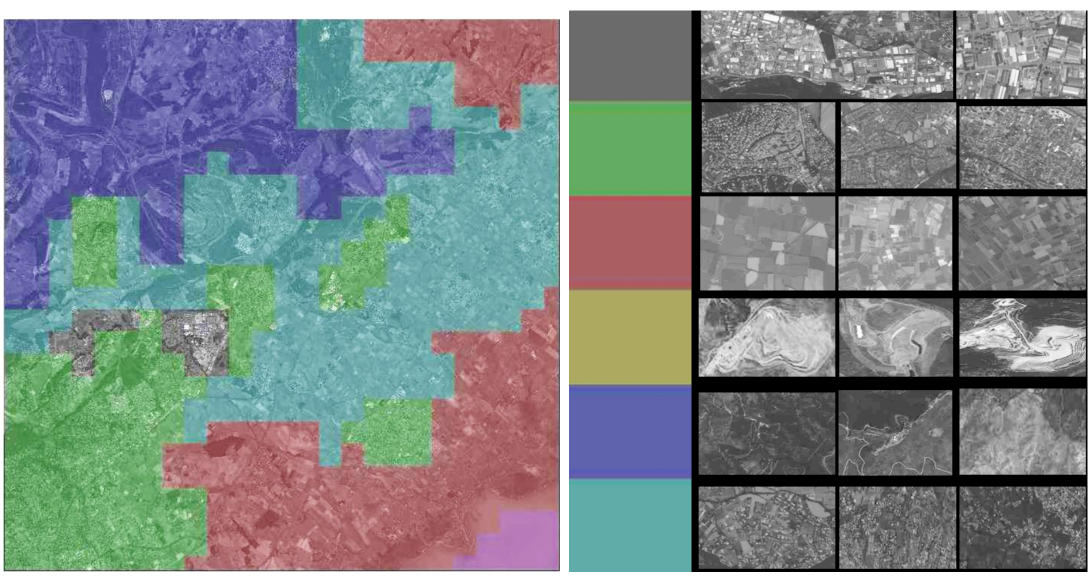

FIG. 11 – SPOT 5 image over the region of Marseilles (Copyright CNES) - Inference of high level semantics information from the low level. The recognized regions are : Industrial areas (grey), Residential areas (green), Fields (red), Quarry and Vast lands (yellow), Mountains (dark blue), Rural areas (light blue). The images on the right present some instances of each class. From.[45]

ACKNOWLEDGMENTS

This document has only been possible because of the help of many persons in the Cnes-Dlr-Enst Competence Centre (among which M. Datcu, A. Giros, M. Campedel, M. Roux, Y. Gousseau, Luo B., J.B. Bordes, I. Kyrgyzov, M. Costache, M. Liénou, C. LeMen, A. Bhattacharya, L. Gueguen) and the financial support of Cnes and Dlr[†].

REFERENCES

1. M. Datcu, H. Dashiel, A. Pelizzari, A. G. M. Quartulli, A. Colapicchioni, M. Pastori, K. Seidel, P. Marchetti, and S. D'Elia, "Information mining in remote sensing image archives : system concepts," *IEEE Transactions on Geoscience and Remote Sensing* **41**(12-1), pp. 2923– 2936, 2003.
2. U. Grenander and M. Miller, *Pattern Theory : From Representation to Inference*, Oxford University Press, Oxford, 2007.
3. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision* **60**(2), pp. 91–110, 2004.
4. C. Harris and M. Stephens, "A combined corner and edge detector," *Proceedings of The Fourth Alvey Vision Conference* , pp. 147–151, 1988.
5. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(10), pp. 1615–1630, 2005.
6. S. Boughorbel, J. Tarel, and N. Boujemaa, "Generalized histogram intersection kernel for image recognition," in *IEEE Int. Conf on Image Processing*, **3**, pp. 161–164, sept 2005.
7. M. Datcu, K. Seidel, and M. Walessa, "Spatial information retrieval from remote-sensing images. I. Information theoretical perspective," *IEEE Transactions on Geoscience and Remote Sensing* **36**(5-1), pp. 1431–1445, 1998.
8. M. Schroder, H. Rehrauer, K.Seidel, and M. Datcu, "Spatial information retrieval from remote-sensing images. II. Gibbs-Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing* **36**(5-1), pp. 1446–1455, 1998.
9. R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems Man and Cybernetics* **SMC-3(6)**, 1973.
10. I. Fogel and D. Sagi, "Gabor filters as texture discriminators," *Biol. Cyber.* **61**, 1989.
11. S. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of textures features based on Gabor filters," *IEEE Trans. on Image Processing* **11**, pp. 1160–1167, October 2002.
12. A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using gabor filters," **24**(12), pp. 1167–1186, 1991.
13. N. Sebe and M. Lew., "Wavelet based texture classification.," in *In International Conference on Pattern Recognition*, **3**, 2000.
14. M. Vetterli, "Filter banks allowing perfect reconstruction," *Signal Processing* **10**(3), pp. 219–244, 1986.
15. A. Bhattacharya, I. H. Jermyn, X. Descombes, and J. Zerubia, "Computing Statistics from a Graph Representation of Road Networks in Satellite Images for Indexing and Retrieval," in *CompIMAGE - Computational Modelling of Objects Represented in Images : Fundamentals, Methods and Applications*, (Coimbra, Portugal), oct 2006.
16. J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
17. A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering : Analysis and an algorithm," in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, eds., pp. 849–856, MIT Press, (Cambridge, MA), 2002.
18. P. Cheeseman and J. Stutz, "Bayesian classification (AUTOCLASS) : Theory and results," in *Advances in Knowledge Discovery and Data Mining*, U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds., pp. 153–180, AAAI Press/MIT Press, 1996.
19. I. Kyrgyzov, H. Maître, and M. Campedel, "A method of clustering combination applied to satellite image analysis," in *International Conference on Image Analysis and Processing ICIAP 2007*, (Modena, Italy), sep 2007.
20. A. L. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(6), pp. 835–850, 2005.

[†]CNES-DLR-ENST Competence Centre on *Information Extraction and Image Understanding for Earth Observation* : http ://www.coc.enst.fr/

21. T. Lange and J. M. Buhmann, "Combining partitions by probabilistic label aggregation," in *KDD '05 : Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 147–156, ACM Press, (New York, NY, USA), 2005.
22. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20(3)**, pp. 273–297, 1995.
23. R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artiff. Intelligence* **97**(1-2), pp. 273–324, 1997.
24. I. Guyon and A. Elisseeff, "An introduction to feature and variable selection," *J. Machine Learning Research* (3), pp. 1157–1182, 2003.
25. B. Schoelkopf and A. Smola, *Learning with Kernels*, MIT Press, 2002.
26. A. Ben-Hur, D. Horn, H. Siegelmann, and V. Vapnik, "Support vector clustering," *J. Machine Learning Research* (2), pp. 125–137, 2001.
27. J. Weston, A. Elisseeff, G. Schoekopf, and M. Tipping, "Use of the zero-norm with linear models and kernel methods," *J. Machine Learning Research* (3), pp. 1439–1461, 2003.
28. B. Luo, J. Aujol, Y. Gousseau, and S. Ladjal, "Extrapolation of wavelet features for satellite images with different resolutions," in *IEEE IGARSS-06*, (Denver, Colorado), jul 2006.
29. T. Lindeberg, *Scale-Space Theory in Computer Vision*, Norwell, MA, USA, Kluwer Academic Publishers, 1994.
30. L. Alvarez, Y. Gousseau, and J. Morel, "Scales in natural images and a consequence on their bounded variation norm," in *Scale-Space '99*, *Lecture Notes in Computer Science* **1682**, 1999.
31. B. Luo, J. Aujol, Y. Gousseau, S. Ladjal, and H. Maître, "Characteristic scale in satellite images," in *IEEE ICASSP-06*, **2**, pp. II 809– II 812, (Toulouse, France), may 2006.
32. P. Monasse and F. Guichard, "Scale-space from a level lines tree," in *Scale-Space Theories in Computer Vision*, pp. 175–186, 1999.
33. B. Luo, J.-F. Aujol, Y. Gousseau, S. Ladjal, and H. Maître, "Resolution independant characteristic scale with application to satellite images," tech. rep., ENST 2006C003, jul 2006.
34. P. Heas and M. Datcu, "Modelling trajectory of dynamic clusters in image time-series for spatio-temporal reasoning," *IEEE Transactions on Geoscience and Remote Sensing* **43**, pp. 1635– 1647, nov 2005.
35. L. Gueguen and M. Datcu, "Image time-series data mining based on the information-bottleneck principle," *IEEE Transactions on Geoscience and Remote Sensing* **45**, apr 2007.
36. L. Gueguen, C. L. Men, and M. Datcu, "Analysis of Satellite Image Time Series based on Information Bottleneck," in *MaxEnt 2006*, pp. 367–374, (Paris), jul 2006.
37. M. Costache, H. Maître, and M. Datcu, "Categorization based relevance feedback search engine for Earth observation images repositories," in *IEEE IGARSS 2006*, (Denver (Colorado)), jul 2006.
38. J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Trans. Inform. Theory* **30**(4), pp. 629–636, 1984.
39. I. O. Kyrgyzov, O. O. Kyrgyzov, H. Maître, and M. Campedel, "Kernel mdl to determine the number of clusters," in *International Conference on Machine Learning and Data Mining MLDM 2007*, (Leipzig, Germany), jul 2007.
40. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)* **39**(1), pp. 1 – 38, 1977.
41. A. D. Gregorio and L. J. M. Jansen, "Land Cover Classification System (LCCS) : Classification concepts and user manual," tech. rep., Food and Agriculture Organisation of the United Nations, Rome, May 1998.
42. M. Bossard, J. Feranec, and J. Otahel, "Corine land cover technical guide - addendum 2000," Tech. Rep. 40, European Environment Agency, May 2000.
43. M. Lienou, H. Maître, and M. Datcu, "Is it possible to automatically produce a CORINE land cover map from a single SPOT image ?," in *ESA EUSC*, (Madrid, Espagne), nov 2006.
44. D. M. Blei, A. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *The Journal of Machine Learning Research* **3**, pp. 993 – 1022, 2003.
45. J.-B. Bordes and H. Maître, "Semantic annotation of satellite images," in *MLDM 5th Int. Conf. on Machine Learning and Data Mining in Pattern Recognition*, (Leipzig, Germany), July 2007.