

TOWARDS ADAPTIVE HIGH-RESOLUTION IMAGES RETRIEVAL SCHEMES

A. Kourgli^{a*}, H. Sebai^a, S. Bouteldja^a, Y. Oukil^b

^a USTHB, Faculté d'Electronique et d'Informatique, B.P. 32 El-Alia, Bab-Ezzouar, Alger, Algérie - (akourgli, hsebai)@usthb.dz

^b ENS, Dept. of Geography, 93 Rue Ali Remli, Bouzareah, Alger, Algérie - oukil.youce@gmail.com

Commission WG II/3 - Spatial Analysis and Data Mining

KEY WORDS: CBIR, High Resolution Satellites Images, Retrieval, Adaptive scheme

ABSTRACT:

Nowadays, content-based image-retrieval techniques constitute powerful tools for archiving and mining of large remote sensing image databases. High spatial resolution images are complex and differ widely in their content, even in the same category. All images are more or less textured and structured. During the last decade, different approaches for the retrieval of this type of images have been proposed. They differ mainly in the type of features extracted. As these features are supposed to efficiently represent the query image, they should be adapted to all kind of images contained in the database. However, if the image to recognize is somewhat or very structured, a shape feature will be somewhat or very effective. While if the image is composed of a single texture, a parameter reflecting the texture of the image will reveal more efficient. This yields to use adaptive schemes. For this purpose, we propose to investigate this idea to adapt the retrieval scheme to image nature. This is achieved by making some preliminary analysis so that indexing stage becomes supervised. First results obtained show that by this way, simple methods can give equal performances to those obtained using complex methods such as the ones based on the creation of bag of visual word using SIFT (Scale Invariant Feature Transform) descriptors and those based on multi scale features extraction using wavelets and steerable pyramids.

1. INTRODUCTION

With the steadily expanding demand for remote sensing images, many satellites have been launched, and thousands of high resolution satellite images (HRSI) are acquired every day. Therefore, retrieving useful images quickly and accurately from a huge image database has become a challenge. Given its importance, this problem has been drawing the attention of people and has received a lot of attention in the literature. As high spatial resolution images are complex and differ widely in their content, even in the same category, the main issue is to find relevant features according to colour, texture and shape information describing the image contents. Many approaches have been proposed to retrieve low and mid-satellite images using their content such as region level semantic features mining (Lu and al., 2012), Knowledge-driven information mining (KIM) (Daschiel and al., 2003), texture model (Aksoy and al., 2013) entropy-balanced bitmap (EBB) tree (Scott and al., 2011). High resolution satellite retrieval schemes use different features according to colour (spectral) features (Bag and Guo, 2004), texture features (Yang and Newsman, 2012) (Sebai and al. 2015) (Shao and al., 2014) and structure features (Yang and Newsman, 2012). Most of these approaches are expressed by visual examples in order to retrieve from the database all the HRSI that are similar to the examples and achieved a satisfactory success for some types of categories. Indeed, approaches based on global features are more adapted to mono thematic images, whereas techniques based on key points extraction are more suited to multi thematic images. So, their efficiency depends on the choice of the set of visual features and on the choice of the similarity metric that models user perception of similarity. Recent studies (Eptoula, 2014) (Sebai and al. 2015) (Shao and al., 2014) (Yang and Newsman, 2012) (Bouteldja and Kourgli, 2015) using a common dataset showed that multi scale feature are more adapted for the retrieval of HRSI.

Even if we can reach, using these methods, better precision values, some categories are not well retrieved especially built-up categories and classes containing several different objects. We still believe that a CBIR (Content Based Image Retrieval) scheme dedicated to HRSI should be adapted to all types contained in a typical dataset. So, our work is motivated by a need to develop an efficient content-based image retrieval scheme so that the recognition of arbitrarily oriented objects in a complex high resolution satellites image can be achieved. To this aim, a preliminary stage is added to the CBIR process to make it adaptive. Accordingly, the paper is organized as follows. Section 2 gives an explanation of how we intend to introduce preliminary stage in our CBIR scheme. Section 3 presents the experimental and discussed results; and Section 4 concludes.

2. SUPERVISED CBIR SCHEME

A Content Based Image Retrieval System (CBIR) is a system which analyzes the visual features such as colour, texture and shape of a query image and retrieves similar images from the image database on the basis of a similarity distance. What features and representations should be used in image indexing depend on the type of images to be retrieved. For remote sensing images the choice of indexing features depends on the type and resolution of the sensor. The main problem in all retrieval strategies is the optimal selection and combination of useful features that provide efficient similarity matching in large databases. To this end a number of relevance feedback mechanisms are currently adopted to refine image queries by modifying the feature space to improve the searching strategy. In a relevance-guided iterative retrieval process (Grana and al., 2008), the user feedback is specified through the identification of a set of relevant and irrelevant images, aiming to better approach the target that the user has in mind. In general, relevance feedback methods demand too much user effort to increase retrieval accuracy (Zhang and al.). Another solution is

* Corresponding author

to conceive a system that guides the user to search for an image on the basis of a first categorization based on image nature. To better explain the issue and the solution proposed, we prefer to introduce the dataset in this section.

2.1 Dataset

It is a manually constructed data set consisting of 21 image classes LULC (Land use Land cover), containing each 100 images of size 256×256 with spatial resolution of 30 cm (Yang and Newsman, 2012). it contains the following classes: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis court.

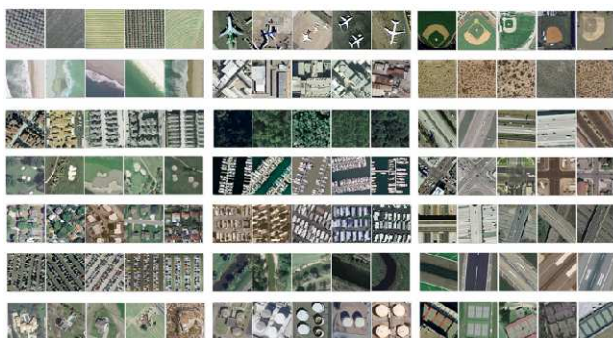


Figure 1. Image patches of the 21 land-use/land-cover classes.

Studying references (Eptoula, 2014) (Sebai and al. 2015) (Shao and al., 2014) (Yang and Newsman, 2012) (Bouteldja and Kourgli, 2015), one can observe that false images are retrieved when different categories share some common textures or/and structures such as those representing buildings, intersection, storage tanks, overpass and tennis court categories. This could be explained by the fact that these categories are more complex containing different structures with different shapes and textures. For example intersections images contain houses sharing some similarities with residential categories, while some areas of golf course category look like those of baseball diamond category (See Figure 1).



Figure 2. Some samples belonging to buildings, tennis court and storage tanks classes.

Moreover, for some categories, objects of interest are too different for the same class and the use of global as well as local features, do not permit to fully and exclusively describe those images (See Figure1). Thus, the feature vectors average the information and the heterogeneous images such as tennis court (surrounded by residences or trees) or storage tanks of different sizes in the middle of fields are not well represented even with SIFT descriptors based on key point description. Thus, the main information with these categories is embedded.

2.2 Pre-Analysis Step

To address both the confusion occurring between classes and differences intra-classes, we thought about pre-analysing the images constituting the base. This preliminary analysis aims, among others, to determine the categories that are mono-textured or representing the same structures from those containing specific objects. As, this pre-analysis is expected to be simple and fast, we tested some common features characterising texture, structure and shape. They are briefly presented below.

2.2.1 2nd Order statistical moments: In statistical texture analysis, texture features are computed from the statistical distribution of observed combinations of intensities at two specified positions relative to each other in the image. Although, there is a lot of parameter describing texture, we restricted our tests on some uncorrelated measures namely: correlation, local variance and local entropy. Our main motivation behind the use of such simple texture features based on statistics of texture is that they are known to provide less number of relevant and distinguishable features in comparison to existing methods such as those based on wavelet transformation or Gabor filters.

2.2.2 HOG (Histogram of Gradients): The HOG feature is widely use for object detection (Dalal and Triggs, 2005). It captures edge or gradient structure that is very characteristic of local shape. The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions. HOG decomposes an image into small squared cells, computes a local 1-D histogram of gradient directions or edge orientations in each cell normalizes the result using a block-wise pattern for improved accuracy, and returns a descriptor for each cell. The blocks can be overlapped with each other for performance improvement. By concatenating all the normalized histograms into a single vector, we get the global HOG feature.

2.3 Features Extraction

Once pre-analysis is realized, a label is given to each query image before features extraction and retrieval stage. Feature extraction is the basis of content-based image retrieval. It is carried out by computing a visual feature on colour images. In a broad sense, visual features include colour, texture, and structure/shape. Because of our image nature, some invariant rotation and translation features are required to characterize spatial colour distribution and thus integrate structure information. As our purpose is to test the pre-analysis stage, simple features are employed: Local variance, and a thresholded version of Local Binary Pattern (LBP). Both are computed using only luminance images.

2.3.1 Local Variance: Measures of local variance have been widely used in image processing for texture and spatial image structure measures. As, this parameter is invariant to illumination changes, we compute the average value of local variances (LV) estimated around each pixel according to each luminance image. The histogram of LV computed in this manner will constitute a rotation invariant feature that permits to identify localized intensity distributions. In this study, it is used both for characterization and for improving labelling (see section 3.2).

2.3.2 Thresholded LBP: The original LBP operator is defined in a rectangular 3×3 pixel neighbourhood. It operates with eight neighbouring pixels using the center as a threshold. The final LBP code is then produced by multiplying the threshold values by weights given by powers of two and adding the results (Ojala and al., 2002). This LBP is extended to a generalized greyscale and rotation invariant operator. However, conventional LBP is very sensitive to noise and a single difference in neighbourhood induces a significant change in the code generated. To avoid this issue, we modified the thresholding function that assigns a bit 1 or 0 according to the difference between neighbouring gray levels and the central pixel so that it depends on if this difference is less than a threshold T .

2.4 Supervising Indexing Scheme

In a query by example scheme, we are interested in retrieving several similar images and this requires comparing two descriptors to obtain a measure of similarity (or dissimilarity) between the two image patterns. Distance measures permit one to translate the similarity concept through a mathematical representation. The choice of distance is crucial and should be considered carefully. Generally, some experience is required in selecting an appropriate distance for a given application. Some common measures are tested in this paper. They are summarized in Table 1.

Manhattan distance	$(L1) = d(x, y) = \sum_{i=1}^n x_i - y_i $ (1)
Euclidian distance	$(L2) = d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ (2)
Chi-square distance	$d(x, y) = \sqrt{\sum_{i=1}^n \frac{(x_i - y_i)^2}{(x_i + y_i)}}$ (3)
Canberra distance	$d(x, y) = \sum_{i=1}^n \frac{ x_i - y_i }{ x_i + y_i }$ (4)
Squared chords distance	$d(x, y) = \sum_{i=1}^n (\sqrt{x_i} - \sqrt{y_i})^2$ (5)

Table 1. Common Similarity measures

To obtain the labelling, the parameters introduced in section 2.2 are computed for different non overlapping windows for each image of the database and similarity (defined by range or variance) between the different blocks is measured to derive a label related to the parameter employed.

Then, to make the process supervised, the feature vector corresponding to the query image with a label L_i is compared to all the features extracted from the dataset possessing also a label L_j . Then the similarity measure is weighted by α defined as:

$$d(x, y) = \alpha \cdot d(x, y) \quad (6)$$

Where α is given the value 1 for features providing of images whose have the same label. The more the labels differ, the more the value of α increases.

By this way, the query image is preferably compared to the ones possessing the same label i.e. sharing similar characteristics.

3. TESTS AND RESULTS

Tests have intensively been conducted using 8 classes according to the following reference (Shao and al., 2014). These are 1: agricultural, 2: airplane, 3: beach, 4: buildings, 5: chaparral, 6: dense residential, 7: forest, and 8: harbor. Two kinds of investigations have been carried out. The first series of tests is designed to find the feature that permits to label the images and to show the effect of pre-analysis stage as well.

3.1 Labelling using single parameter

To build the pre-analysis stage, we tested some parameters known to characterize the structure and texture and others recognized to highlight the shape of objects. First, to make the distinction between mono-thematic images and those that are not, the images are divided into non overlapping blocks, and the parameters described in section 2.2 are computed. The parameters obtained for the different blocks according to each image are compared via a difference measure (range or variance) to obtain one single value.

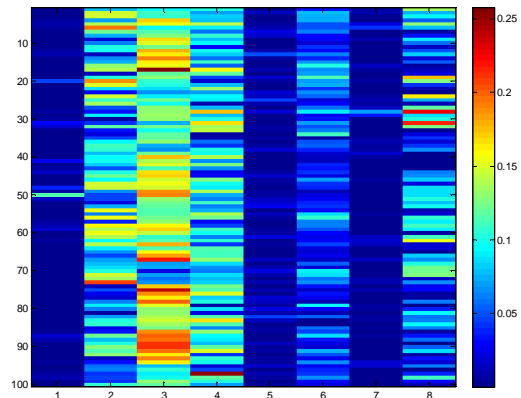


Figure 3. Variance of parameter correlation using 9 windows for each sample (1: agricultural, 2: airplane, 3: beach, 4: buildings, 5: chaparral, 6: dense residential, 7: forest, 8: harbor).

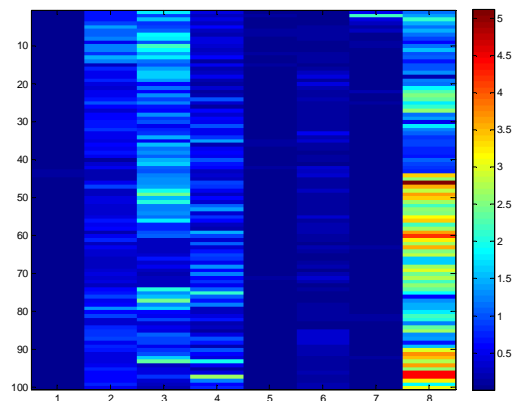


Figure 4. Variance of parameter entropy using 25 windows for each sample (1: agricultural, 2: airplane, 3: beach, 4: buildings, 5: chaparral, 6: dense residential, 7: forest, 8: harbor).

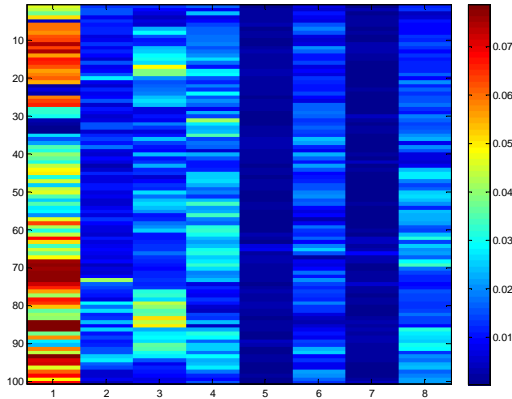


Figure 5. Variance of HOG histograms using 9 windows for each sample (1: agricultural, 2: airplane, 3: beach, 4: buildings, 5: chaparral, 6: dense residential, 7: forest, 8: harbor).

The above figures illustrate the values obtained in the form of images. Recall that we are using 8 classes, each one containing 100 samples. These figures show that the correlation parameter varies little (low value in blue) between different windows for agricultural, chaparral and forest classes that are mono-textured. For other classes, it is not very discriminator. As expected, entropy parameter, reflecting the degree of disorder, is almost constant between the different blocks of a single image composed of the same textures such as the three categories mentioned previously. However, for some samples of dense residential class, this value is also low, which generates confusion and does not permit to choose a fixed threshold.

The last parameter HOG related to the shape not only helps to differentiate images containing one thematic but permits also to distinguish between structured and non structured ones. Indeed, agricultural images are textured and most of them are also structured with predominant orientation.

This parameter permits to define three intervals [01 0.03 0.08]. Once the intervals determined, a label is given to each image of the database according to the interval it belongs to.

The next step is features extraction. As mentioned before, our aim is to study the impact of pre-analysis, so we consider simple features: global histogram of local variance computed on 16 bins and global histogram of thresholded uniform LBP constituted of 10 different codes.

Usually, CBIR performance is measured by precision and recall. Precision P as well as average precision and AP are given as:

$$P = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \quad (7)$$

$$AP = \frac{1}{Nq} \sum_{i=1}^{Nq} P(i) \quad (8)$$

where Nq represents the number of queries.

Similarly, recall R and average recall AR are given as :

$$P = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}} \quad (9)$$

$$AR = \frac{1}{Nq} \sum_{i=1}^{Nq} R(i) \quad (10)$$

We first give some indexing results using the chi-square distance with (Fig.6) and without (Fig. 7) pre-analysis stage.

The index of retrieved images are sorted from left to right and given according to their color.

One can observe from Fig. 6 that the images belonging to the first class (top left in dark blue) are very poorly indexed because the features employed are quite simple and not suitable for this category but the fact of going through a preliminary stage of analysis increases the number of correctly retrieved images (Fig. 7). This is true for other categories because the confusion between different classes diminishes. Thus the values of precision and recall increase (see Figure 8 and Table 2).

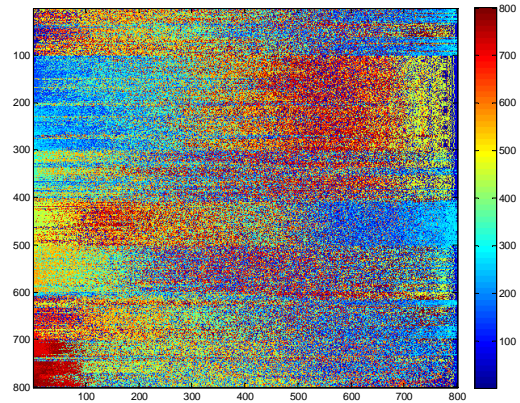


Figure 6. Retrieved image indexed by corresponding colour without employing pre-analysis stage.

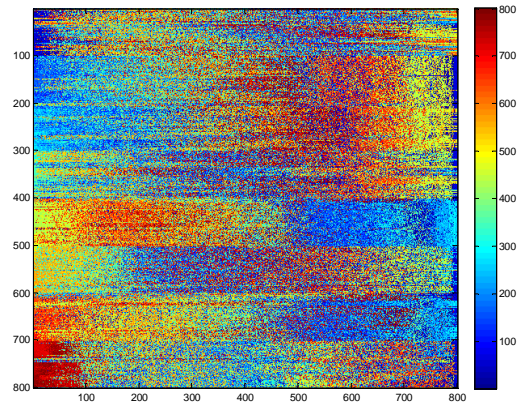


Figure 7. Retrieved image indexed by corresponding colour employing pre-analysis stage.

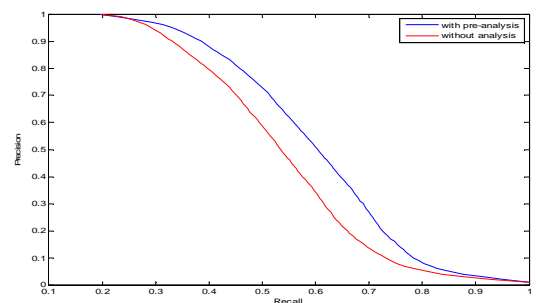


Figure 8. Precision-Recall curves obtained using the 800 samples with (in blue) and without (in red) pre-analysis stage.

Table 2 and Table 3 illustrate the overall recognition rate obtained for different precision values using the similarity measure of Chi-square for each category.

Cl	Pr% 10	Pr% 20	Pr% 30	Pr% 40	Pr% 50	Pr% 60	Pr% 70	Pr% 80	Pr% 90	Pr% 100	Avg
1	55.9	45.0	40.6	37.2	34.8	33.0	31.2	30.3	28.6	27.4	36.4
2	82.8	75.7	68.6	63.8	61.0	58.1	55.6	53	50.8	48.8	61.8
3	84.7	76.6	71.8	67.3	63.4	56	56.4	54.0	51.9	49.6	63.6
4	69.6	63.0	58	55.6	53	51.1	49.1	47.5	46.2	45	53.9
5	79.4	74.9	72	70.2	67.8	66.5	64.7	62.8	60.3	57.3	67.6
6	78.1	70.7	68.6	67.3	66.2	65.0	64.2	63.2	61.8	60.6	66.6
7	63.7	55.4	51.7	49.1	46.7	44.8	42.8	41.6	40.0	38.8	47.5
8	95.3	94.3	93.3	92.1	90.6	88.4	86.1	83.7	80.5	76.1	88

Table 2. Average precision values for each category.

Cl	Pr% 10	Pr% 20	Pr% 30	Pr% 40	Pr% 50	Pr% 60	Pr% 70	Pr% 80	Pr% 90	Pr% 100	Avg
1	79.8	74.0	69.8	64.1	58.9	54.3	50.2	46.9	44.2	41.6	58.4
2	84.5	78.8	72.9	68.5	65.5	62.6	59.6	56.8	54.6	52.2	65.6
3	82.8	75.4	70.3	65.9	62.6	58.5	55.1	51.9	49.2	47.0	61.9
4	70.6	62.0	57.6	53.7	51.1	48.7	46.9	45.3	43.9	42.7	52.2
5	80.7	76.5	74.0	72.4	70.2	69.0	67.7	66.5	64.3	62.2	70.3
6	77.5	71.0	68.9	67.9	67.4	66.4	65.6	64.5	63.5	62.4	67.5
7	76.3	69.8	67.1	65.5	63.8	61.5	59.5	57.6	55.6	53.8	63.1
8	95.2	94.2	93.3	92.1	90.7	88.5	86.4	83.8	80.9	76.4	88.2

Table 3. Average precision values for each category after pre-analysis stage and labelling.

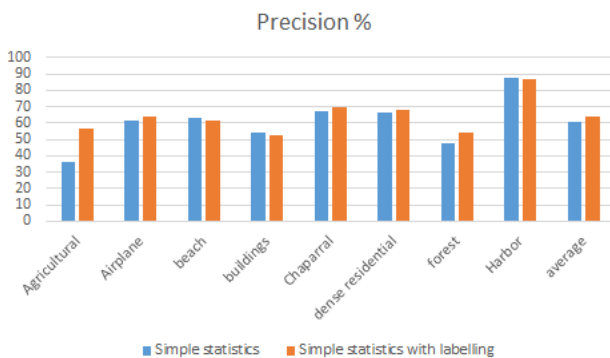


Figure 9. Average precision using simple statistics with an without labelling for each category.

One can notice that the addition of the pre-analysis stage allows the overall mean accuracy to increase from 59% to a precision of 65.9 i.e. a gain of 6%. To assess the effect of reanalysis and labelling, we compared the results obtained using more complex approaches. Thus, comparisons are limited to the basic version of these approaches and summarized in the following table

Precision/ Feature	Pr% 10	Pr% 20	Pr% 30	Pr% 40	Pr% 50	Pr% 60	Pr% 70	Pr% 80	Pr% 90	Pr% 100	Avg
CGOT	80	70	65	62	59	56	54	52	49	47	60
SP	87.5	79.9	75.0	71.0	67.6	64.8	61.9	59.3	56.8	54.3	67.8
SIFT	81.2	76.6	74	72.2	70	68	66.1	64.1	61.8	59.1	69.2
CT-DWT	85.9	77.3	71.7	67.3	63.7	60.5	57.6	54.9	52.3	49.8	64.1
Simple statistics	80.9	75.2	71.7	68.8	66.3	63.7	61.4	59.2	57.0	54.8	65.9

Table 4. Comparison with other methods using average precision values for each category.

The table 4 shows that inserting a pre-analysis step in the CBIR scheme yields to interesting results compared to approaches found in recent references. Let one keeps in mind that for each channel, SIFT operator, computed on a vector of 128 elements (Lowe, 1999), needs the buildings of a bag of visual word, whereas CT-DWT (Kingsbury, 2000) and Steerable Pyramid produce a huge vector whose elements (histograms or statistical

moments) are derived from the sub images in different orientations at different levels. Also, CGOT vector is constituted of 80 Gabor texture features which give a long vector (Shao and al., 2014). While, we tested the proposed scheme with two features histogram of local variance computed using 16 bins and a modified uniform LBP that produce an histogram on 10 bins constituting after concatenation a feature vector whose length is 26.

3.2 Labelling using parameter combination

To increase the performances of labelling, other parameters are tested and combined to make labelling more efficient. Indeed, high resolution satellite images can broadly be divided into two types of images: The first type concerns the images representing objects of interest such as airplane, harbor and built-up areas, while with the second kind of images; the main information is texture such as those representing agricultural, beach, chaparral and forest. This second category of images can be, in turn, subdivided into 3 main classes: structured, mono textured, multi textured. Whereas, the first type gathers, also, three kind of images: those that are highly structured with heterogeneous areas between many objects to recognize such as those corresponding to built-up areas, those which correspond to few objects surrounded by homogenous textures (airplane class is an example of this type) and finally those containing also many objects with a mono textured back ground. This classification leads to the following labelling scheme.

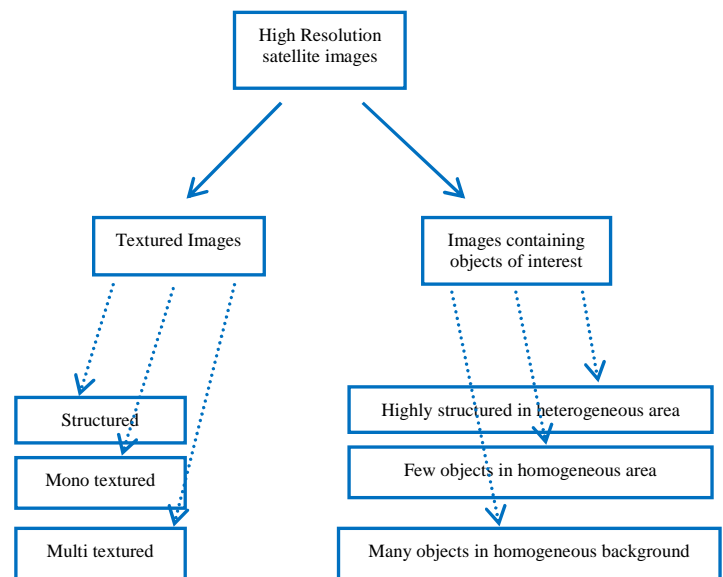


Figure 10. Adapting labelling to high resolution satellite images characteristics.

To establish the first distinction, the similarity between the different blocs of the image is measured via the local variance. Again, each image is decomposed into overlapping blocks (see Fig.11). For each block B_i , the range between local variances is estimated and the average of the range values is computed according to the following formulas:

$$RLV_{Bi} = range(var(x, y), x, y \in Bi) \quad (11)$$

$$Mean_{RLV} = \frac{1}{N} \sum_{i=1}^N RLV_{Bi} \quad (12)$$

Where N is a number of blocks.

For a textured image, the mean defined by equation (12) will be small because the range of local variances will be the same for each block B_i .

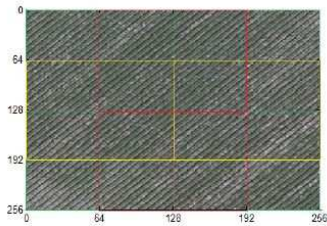


Figure 11. Example of nine overlapping blocks.

As it can be observed in the following figure (Fig.12), this parameter is low for textured images such as agricultural, beach, chaparral and forest (from blue to green)

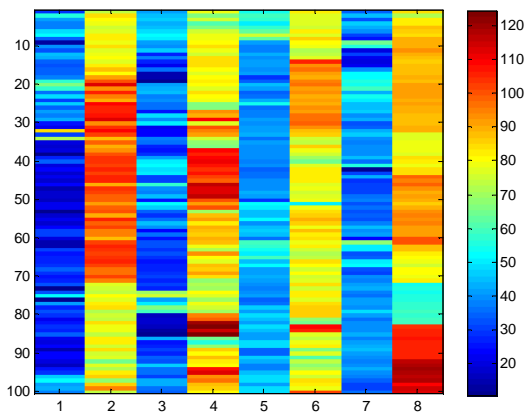


Figure 12. MRLV for each sample (1: agricultural, 2: airplane, 3: beach, 4: buildings, 5: chaparral, 6: dense residential, 7: forest, 8: harbor).

To distinguish if the textured images contains some structured areas as agricultural another parameter is computed, it is obtained by computing the correlation of HOG vector defined by:

$$Corr_{HOG}(l) = \frac{1}{L} \sum_{n=1}^{L-l} HOG(n)HOG(n+l) \quad (13)$$

Where L represents the size of HOG vector.

The use of correlation of HOG instead of HOG vector allows to discriminate between textures that are also structured (high value in red and yellow) in Fig. 13.

The distinction between mono and multi textured images is obtained through the use of homogeneity parameter derived from co-occurrence matrix. It is defined as:

$$Homogeneity = \sum_i \sum_j \frac{P(i,j)}{1+(i-j)^2} \quad (14)$$

Where $p(i,j)$ are the elements of co-occurrence matrix. As shown by Fig. 14, most of multi textured images such as beach exhibit a high value for homogeneity while mono textured ones such as agricultural, chaparral and forest show low values (in blue).

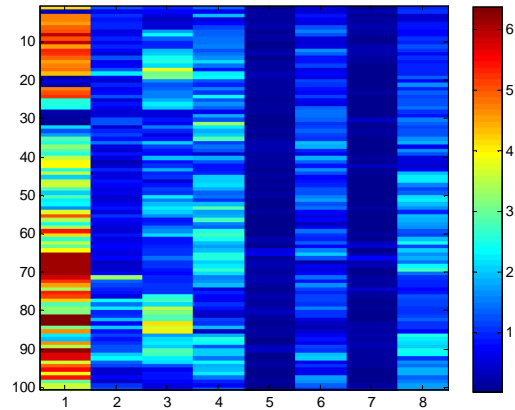


Figure 13. HOG correlation for each sample (1: agricultural, 2: airplane, 3: beach, 4: buildings, 5: chaparral, 6: dense residential, 7: forest, 8: harbor)

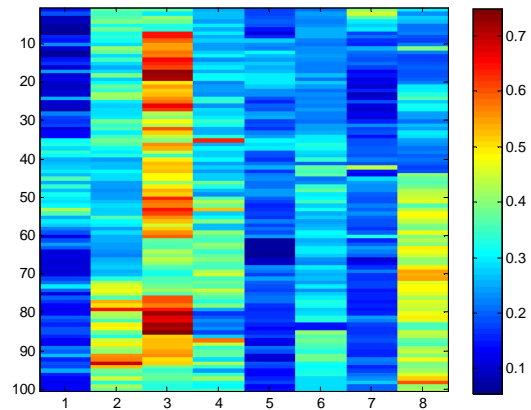


Figure 14. Homogeneity of co-occurrence matrix for each sample (1: agricultural, 2: airplane, 3: beach, 4: buildings, 5: chaparral, 6: dense residential, 7: forest, 8: harbor).

For the second category of images : two parameter are used to distinguish between those containing few objects in homogeneous area, those containing many objects and finally those containing many objects in heterogeneous area. These parameters are entropy and dissimilarity. While entropy is locally estimated for the whole image, the dissimilarity is computed from co-occurrence matrices. It is defined below:

$$Dissimilarity = P(i,j)\sqrt{i^2 - j^2} \quad (15)$$

Again, an image shown in Fig. 15 depicts the result for dissimilarity measure while for entropy, we refer to Fig. 4. The combination of the two parameters show that most of images containing man-made structures with homogeneous background exhibit a high value of entropy (see fig.4)

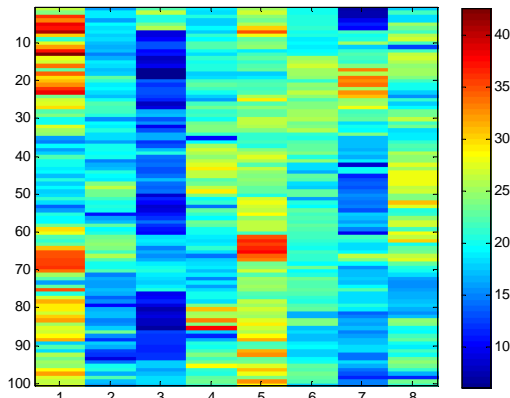


Figure 15. Dissimilarity parameter derived from co-occurrence matrix for each sample (1: agricultural, 2: airplane, 3: beach, 4: buildings, 5: chaparral, 6: dense residential, 7: forest, 8: harbor).

These different parameters permit to define the six labels contributing in the supervised indexing process.

We first give some indexing results using simple statistics and the chi-square distance with pre-analysis stage and the improved labelling obtained through parameter combination.

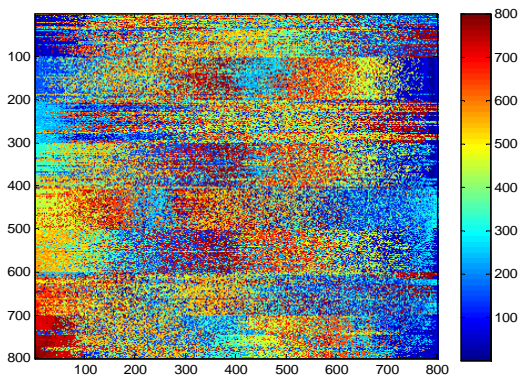


Figure 16. Retrieved image indexed by corresponding colour with improved labelling using simple statistics.

Compared to Fig. 7, Fig.16 shows that refining the labelling permits to gain more precision as confirmed by table 4 and Fig. 17. The gain in terms of precision is almost 10%.

Precision/ Feature	Pr% 10	Pr% 20	Pr% 30	Pr% 40	Pr% 50	Pr% 60	Pr% 70	Pr% 80	Pr% 90	Pr% 100	Avg
Simple labelling	76,2	69,5	65,7	62,9	60,5	58,4	56,3	54,5	52,5	50,5	60,7
Improv. labelling	83,5	78,7	75,8	73,6	71,1	68,7	66,3	63,8	61,2	58,6	70,1
CGOT	80	70	65	62	59	56	54	52	49	47	60
SP	87,5	79,9	75,0	71,0	67,6	64,8	61,9	59,3	56,8	54,3	67,8
SIFT	81,2	76,6	74	72,2	70	68	66,1	64,1	61,8	59,1	69,2
CT-DWT	85,9	77,3	71,7	67,3	63,7	60,5	57,6	54,9	52,3	49,8	64,1

Table 5. Comparison using average precision values for each category for improved labelling using simple statistics.

Fig. 17 permits to compare simple statistics to CT-DWT, one can observe that complex and heterogeneous images such as built-up areas (buildings, dense residential) and harbor that are not usually well retrieved because of their complexity gain better retrieving scores when the images are pre-analysed.

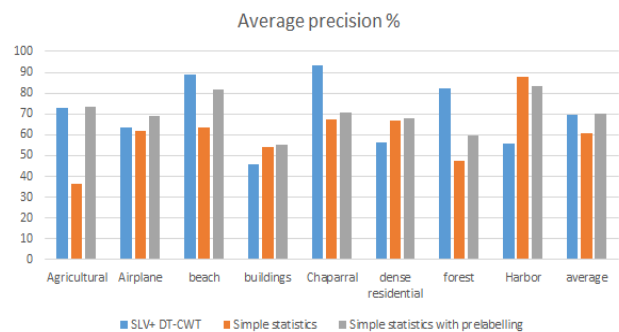


Figure 17. Average precision for each category.

Compared to basic versions of more complex descriptors such as those based on multi scale analysis, simple statistics using luminance images boosted by pre-analysis stage perform better. Precision versus recall curves depicted by Fig. 18 prove that simple statistics (in light blue) associated to labelling stage is as good as the best descriptor (i.e. SIFT descriptor) (in green).

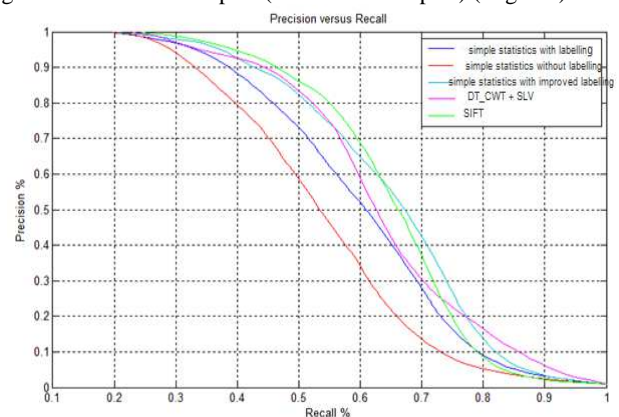


Figure 18. Precision-Recall curves obtained using the 800 samples.

To confirm the advantages of introducing pre -analysis step in any indexing scheme, other tests have been conducted using CT-DWT and SIFT descriptors.

3.3 Applying labelling to CT-DWT and SIFT descriptors

Multi scale representation of features descriptors is used to better reflect the objects of different sizes and shapes present in HRS images (Sebai and Kourgli, 2015). Multi resolution DT-CWT technique (Kingsbury, 2000) is widely used since it allows analysis that is localized in both space and frequency. It calculates the complex transform of a signal using two separate DWT decompositions (two trees). While SIFT (Scale invariant feature transform) descriptor developed by David Lowe (Lowe, 1999) permits to detect and describe local features in images. It is extracted from image patches around the interest point and then associated to a bag of visual word representation. Again, we associated pre-analysis stage to both indexing schemes. Results are summarized in table 6. Comparing average precisions reported in table 5 and table 6, one can notice a gain of approximately of 7% for indexing scheme based on CT-

DWT, while for the one based on SIFT representation, the gain reaches 4%

Precision/ Feature	Pr% 10	Pr% 20	Pr% 30	Pr% 40	Pr% 50	Pr% 60	Pr% 70	Pr% 80	Pr% 90	Pr% 100	Avg
Simple statistics	83,5	78,7	75,8	73,6	71,1	68,7	66,3	63,8	61,2	58,6	70,1
SIFT	85,4	81,5	79,4	77,1	75,1	72,4	69,8	67	64,0	60,8	73,2
CT-DWT	89	83,7	79,5	76,4	73,7	71,2	68,7	65,6	62,4	58,8	72,9

Table 6. Applying improved labelling to the other methods (SIFT, CT-DWT).

Once more, precision versus recall curves (Fig.19) permit to illustrate the advantage of incorporating a pre-analysis stage on different types of indexing schemes. In all case, it boost the CBIR performances.

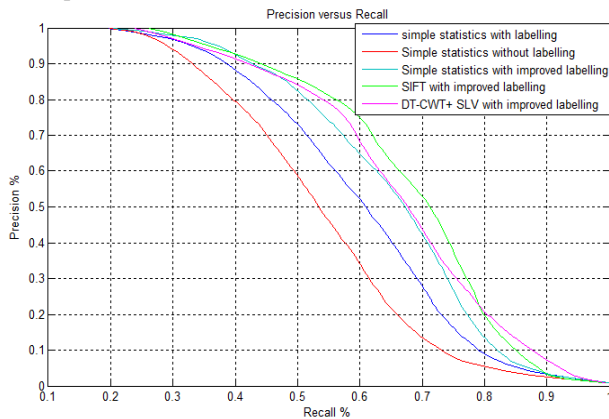


Figure 19. Precision-Recall curves obtained using the 800 samples.

To better evaluate the effect of adding pre-analysis stage, one query image has been chosen from the two categories that are the most difficult to retrieve i.e.: buildings and dense residential. Both contain man-made objects and share some similarities whether for shape or for structure.



Figure 20. Examples of images retrieved using simple statistics without labelling (buildings:7/18, dense residential: 13/18).

The retrieval results are presented in Fig. 21. One can observe that less confusions since labelling is employed whatever scheme is employed. Moreover, simple statistics, that are based on the concatenation of two global histogram yielding to a vector of 26 elements, give interesting performances. Indeed, using simple statistics, one gain not only, in terms of precision but also in terms of speed.

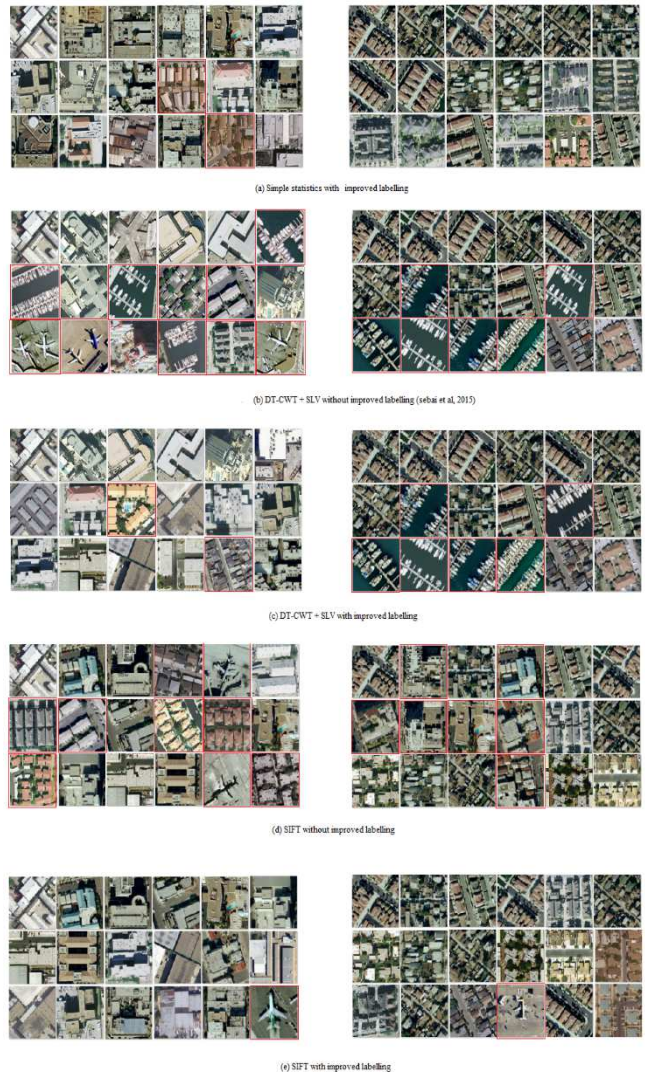


Figure 21. Examples of images retrieved a) simple statistics with labelling: buildings:16/18, dense residential: 18/18. b) CT-DWT without labelling: buildings:8/18, dense residential: 12/18. c) CT-DWT with labelling: buildings:16/18, dense residential: 12/18. d) SIFT without labelling:10/18, dense residential: 11/18. e) SIFT with labelling: 17/18, dense residential: 17/18.

4.CONCLUSION

Rapid growth of remote sensed information generates a new research challenges in processing, transferring, archiving, and retrieving of these huge amounts of data. Existing methods share some common issues; the main important is that they are not adapted to all types of categories. In this paper, we have proposed to modify CBIR scheme by adding a new stage that permits to label the image to be retrieved according to its inherent characteristics. We quantitatively analyzed the efficiency of the weighting of distance measure. From the tests, it appeared that this stage can greatly improve the retrieval process by boosting its performances even for basic features. Thus, associated to multi scale representation of colour features descriptors, this will permit to better take into account objects of different sizes and shapes present in HRS images. Future work includes extending the investigation to consider the 21 categories and better combine efficient parameters in the pre-

analysis to obtain a robust labelling and thus further improve the retrieval performance as well as to bridge the semantic gap.

REFERENCES

- Aksoy, S., Yalniz, I.Z., Nick, J., Tasdemir, K.: Automatic Detection and Segmentation of Orchards Using Very High-resolution Imagery. *IEEE Transactions on Geosciences and Remote Sensing* 50, 3117–3131 (2012)
- Bao, Q., Guo, P.: Comparative studies on similarity measures for remote sensing image retrieval. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 1112–1116 (2004)
- Bouteldja, S., Kourgli, A. Multiscale Texture Features For The Retrieval Of High Resolution Satellite Images. IWSSIP 2015, London, pp. 170-173.
- Daschiel, H., Pelizzar, A., Quartulli, M.: Information mining in remote sensing image archives: system concepts. *IEEE Transactions on Geosciences and Remote Sensing* 41, 2923 – 2936 (2003)
- Eptoula, E.. Remote Sensing Image Retrieval With Global Morphological Texture Descriptors. *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 52, No. 5, May 2014.
- Grana, M., Veganzones, M.A., Maldonado, J. O. “On content based image retrieval systems for hyper spectral remote sensing images,” *Computational Intelligence for Remote Sensing*, vol. 133, pp. 125–144, 2008.
- Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893 (2005)
- Kingsbury, N.G.: The dual-tree complex wavelet transform with improved orthogonality and symmetry properties. In: IEEE International Conference on Image Processing, pp. 375–378 (2000)
- Liu, T., Zhang, L., Li, P., Lin, H.I.: Remotely sensed image retrieval based on region-level semantic mining. *EURASIP Journal on Image and Video Processing*, 1687–5281 (2012)
- Lowe, D.G.: Object recognition from local scale-invariant features. In: ICCV, pp. 1150–1157 (1999)
- Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution grayscale and rotation invariant texture classification with local binary patterns. *Transactions on Pattern analysis and Machine Intelligence* 24, 971–987 (2002)
- Scott, G., Klaric, M., Davis, C., Shyu, C.R.: Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases. *IEEE Transactions on Geosciences and Remote Sensing* 49, 1603–1616 (2011)
- Sebai, H., Kourgli, A.: Dual-tree complex wavelet transform applied on colour descriptors for remote-sensed images retrieval. *Journal of Applied Remote Sensing (JRS)* 9(1) (2015)
- Sebai H., Kourgli, A. :Improving high resolution satellite images retrieval using color component features. 18th International Conference on Image Analysis and Processing ICIAP 2015, Part II of Lecture Notes in Computer Science Volume 9280, 2015, pp 264-275. Genoa, Italy, September 7-11.
- Shao, Z., Zhou, W., Zhang, L., Hou, J.: Improved colour texture descriptors for remote sensing image retrieval. *Journal of Applied Remote Sensing (JRS)* 8(1) (2014)
- Yang, Y., Newman, S.: Geographic image retrieval using local invariant features. *IEEE Transactions on Geosciences and Remote Sensing*, 818–832 (2012)
- Zhang Y. Li, W. Mo, Z., Zhao, T., An adaptive-weight hybrid Relevance feedback approach for content based image retrieval, IEEE International Conference on Image Processing (ICIP), pp. 3977 – 3981, 2013.