

# LOGICAL QUERY COMPOSITION FROM LOCAL VISUAL FEATURE THESAURUS

*Julien FAUQUEUR and Nozha BOUJEMAA*

INRIA, IMEDIA Research Group, BP 105, F-78153 Le Chesnay, France

{Julien.Fauqueur, Nozha.Boujemaa}@inria.fr

<http://www-rocq.inria.fr/imedia/>

## ABSTRACT

We present a novel framework for intelligent search and retrieval by image region composition. Unlike traditional Query-by-Example paradigm, **no starting example image** is used : the user provides its mental representation of target image by means of a **region photometric thesaurus**. It gives an overview of the database content in the query interface.

Unsupervised generation of this thesaurus is based on grouping similar regions into “region categories”. The user can indicate the presence and absence of types of regions in searched images by a new **logical composition of region categories** query paradigm, which relates to that of text of retrieval. The sophisticated specification of the target image content introduces visual semantics in the search.

The implementation of this framework is simple and fully unsupervised. Image symbolic indexing coupled with logical query allows very powerful and fast search. Two test scenarios are investigated : a photo-agency scenario on Corel Photostock and a TV news scenario.

## 1. INTRODUCTION

The earliest Content-Based Image Retrieval approach is the global query-by-example approach, initially developed in these systems : QBIC [4], PHOTOBOOK [8] and others. This approach provides approximate results when the focus of the search is a specific object or part in an image. Partial query formulation allows the user to specify which part of the image is the target of his/her interest and leads to a higher user satisfaction.

Partial query systems based on image regions have been proposed to allow more specific queries. Among the few existing region-based query systems we can cite BLOBWORLD [2], NETRA [7] and more recently IKONA [3]. These systems simply perform an exhaustive search among regions in database from a single example region. The SIMPLICITY system [12], although involving region matching, actually performs global image retrieval. In VISUALSEEK [10], a

multiple region query was proposed by sketching rectangles of synthetic colors and synthetic textures.

Be they global, single or multiple region based, existing CBIR systems all rely on the same paradigm : query-by-example (example being an image or one or more regions) and retrieval by exhaustive search. This approach is well suited to perform visual comparison between a given example and entries in the database, i.e. to answer to such a query : “show me images/regions in the database similar to this image/regions”. Very often, however, the user does not have an example image to start the search. The target image is only in the user’s mind. In this case, the prior search of an example to perform the actual query by example is tedious, especially in the case of a multiple region query.

The new framework presented in this paper differs completely from this paradigm on both query and retrieval processes. After segmenting images in the database, all extracted regions are grouped into categories of visually similar regions. In the query interface, the categories provide an overview of the types of regions which constitute the images in the database. They can be viewed as a “region photometric thesaurus”. Images are simply indexed by the list of category labels and, as a consequence, the user can very quickly retrieve images from queries as complex as : “find images composed of regions of these types and no regions of those types”.

In section 2, we will briefly present the algorithm to group visual features. In section 3, the generation of region categories and their neighbor categories to allow range queries will be explained. Then, in section 4, we will detail the approach to achieve image retrieval by composition of region categories. We will present an adapted retrieval strategy to handle queries from complex compositions and then an efficient indexing and retrieval scheme. In section 5, we will present an original user interface along with the results which will be followed by discussions in section 6. In section 7 concluding remarks on the specificity of this new query by composition framework, related and future work will be addressed.

## 2. VISUAL FEATURE GROUPING METHOD

For region categorization, an efficient clustering scheme is required. The CA (for *Competitive Agglomeration*) clustering, originally presented in [5], was chosen because of its major advantage of determining automatically the number of clusters. Using notations from [5], we call  $\{x_j, \forall j = 1, \dots, N\}$  the set of data we want to cluster and  $C$  the number of clusters.  $\{\beta_i, \forall i = 1, \dots, C\}$  denote the prototypes to be determined.  $d(x_j, \beta_i)$  is the Mahalanobis distance between data  $x_j$  and prototype  $\beta_i$ . The CA-clustering is performed by minimizing the following objective function  $J$ :

$$J = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^2 d^2(x_j, \beta_i) - \alpha \sum_{i=1}^C [\sum_{j=1}^N u_{ij}]^2 \quad (1)$$

Subject to membership constraint :  $\sum_{i=1}^C u_{ij} = 1, \forall j = 1, \dots, N$ , where  $u_{ij}$  represents the membership degree of feature  $x_j$  to cluster  $i$ . Minimizing the first term is equivalent to perform a Fuzzy C-Means clustering which determines  $C$  optimal prototypes and the fuzzy partition  $U$ . Minimizing the second term guarantees the clusters validity. So minimizing  $J$  with an over-specified number of initial clusters optimizes simultaneously the data partition and the number of classes. CA is used in our framework for both similar regions grouping and segmentation (see [3]).

## 3. CATEGORIZATION AND RANGE QUERY IN THE REGIONS FEATURE SPACE

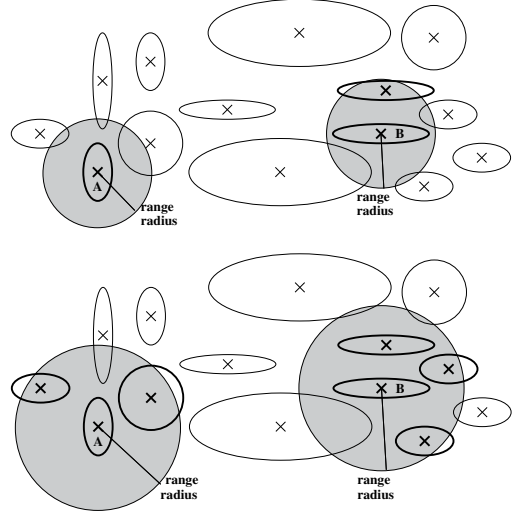
To extract regions, we adopt the image segmentation technique presented in [3]. It was developed specifically for image retrieval by regions. It is based on the CA-clustering of Local Distributions of Quantized Colors (*LDQC's*).

We define the **region categories** (denoted  $C_1, \dots, C_P$ ) as the clusters of regions which have similar visual features. They are the basis of the definition of similar regions in the retrieval phase. Here we choose to characterize regions with their mean color which means that regions from the same category will have similar mean color. It is important to note that here, other visual cues could be used or combined to color here such as color distribution, position, surface. Despite the straightforwardness of mean color description, we will see it is relevant to form generic categories.

Regions mean color is determined in the LUV space, which is chosen for its perceptual uniformity. In the color space, all the regions mean colors form a relatively compact and dense data set and no natural data grouping can be expected. We can not make a priori assumption concerning the well-definition of clusters of regions for any database. However what can always be guaranteed is an intra-category homogeneity by setting a fine clustering granularity. Region categories are formed by grouping with CA and a fine granularity the regions mean color features. For each region

category, its **representative region** is defined as the closest region to its prototype. Representative regions are only used to identify each category in the query interface.

Since similarity between regions will be defined, at a first level, as members of the same category, a fine clustering granularity will ensure the retrieval of very similar regions (hence high retrieval precision). At a second level, we also consider as similar regions which are in close categories (called “neighbor categories”) to also allow high recall. This key idea allows to achieve **range queries** in the regions feature space. The **neighbor category** of a category  $C_q$  of prototype  $p_q$  is defined as a category  $C_j$  whose prototype  $p_j$  satisfies  $\|p_q - p_j\|_{L^2} \leq \gamma$ , for a given range radius threshold  $\gamma$ . We call  $N^\gamma(C_q)$  the set of neighbor categories of a category  $C_q$ . By convention, we decide that a category  $C_q$  belongs to  $N^\gamma(C_q)$  as a neighbor of itself at distance zero. Range radius  $\gamma$  will be adjusted at retrieval phase. See figure 1 for an illustration of the definition of neighbors using the radius. Note that thanks to this range query scheme, the search is less dependent on the database partition into categories since all close categories are considered together.



**Fig. 1. Range radius and neighbor categories.**  $A$  and  $B$  are two categories. Influence of range radius  $\gamma$  on definition of neighbor categories : a high radius (top) and a lower radius (bottom) integrate more or less neighbor categories to define the type of searched regions. The grey disks of radius  $\gamma$  cover the neighbor categories  $N^\gamma(A)$  and  $N^\gamma(B)$ . Neighbor categories are drawn with thicker contours. Prototypes are identified by crosses.

The combination of homogeneous region categories with the integration of neighbor categories is the key choice in the definition of the range query scheme.

#### 4. IMAGE RETRIEVAL BY COMPOSITION

From this point on, regions are not considered individually anymore but are totally identified to the category they belong to. With the help of all categories representative regions, the user will select Positive Query Categories (referred to as **PQCs**) and Negative Query Categories (**NQCs**). The PQCs correspond to the user-selected categories of regions which should appear in retrieved images. They are denoted as  $\{C_{pq_1}, \dots, C_{pq_M}\}$ . The NQCs correspond to the user-selected categories of regions which should *not* appear in retrieved images and are denoted as  $\{C_{nq_1}, \dots, C_{nq_R}\}$ .

In its most complex form, a **query composition** is defined as the formulation : “find images composed of regions in these PQCs and no region from those NQCs” and expressed as the list of PQC labels  $\{pq_1, \dots, pq_M\}$  and NQC labels  $\{nq_1, \dots, nq_R\}$ .

Performing a query composition first requires to retrieve images which contain a region from a single PQC category denoted  $C_{pq}$  say. For a given category  $C_{pq}$ , we define  $IC(C_{pq})$  to be the set of images containing at least one region belonging to category  $C_{pq}$ . To expand this search to a range query, we now take into account neighbor categories of  $C_{pq}$  by defining relevant images as those which have a region from category  $C_{pq}$  or from any of its neighbors :

$$\bigcup_{C \in N^\gamma(C_{pq})} IC(C) \quad (2)$$

Range radius threshold  $\gamma$  is set in the user interface.

To extend the query to all  $M$  PQCs :  $C_{pq_1}, \dots, C_{pq_M}$ , we search images which have a region in  $C_{pq_1}$  or its neighbors and ... and a region in  $C_{pq_M}$  or its neighbors. The set  $S_Q$  of images satisfying this multiple query is now written as :

$$S_Q = \bigcap_{i=1}^M \left[ \bigcup_{C \in N^\gamma(C_{pq_i})} IC(C) \right] \quad (3)$$

Then, to also satisfy the negative query we must determine images which contain a region from any of the  $R$  NQCs :  $C_{nq_1}, \dots, C_{nq_R}$ . As before, the neighbor categories should also be taken into account. So the set  $S_{NQ}$  of images containing the NQCs is written as :

$$S_{NQ} = \bigcap_{i=1}^R \left[ \bigcup_{C \in N^\gamma(C_{nq_i})} IC(C) \right] \quad (4)$$

So the set  $S_{result}$  of retrieved images which have regions in the different PQCs and which do not have regions in the NQCs is expressed as the set subtraction of  $S_Q$  and  $S_{NQ}$  :

$$S_{result} = S_Q \setminus S_{NQ} \quad (5)$$

This set  $S_{result}$  constitutes the set of relevant images.

We will see that unions, intersections and subtractions in the expression of  $S_{result}$  are directly equivalent to formulate the query with logical operators as illustrated in figure 4 : OR between the neighbors (see expression 2). AND between query categories (expression 3). ANDNOT for negative query categories (expression 4).

To evaluate the expression of  $S_{result}$  (expression 5), the brute force approach would consist in testing, for each image in the database, if it contains regions belonging to the PQCs (and their neighbors) but contains no region in any of the NQCs (and their neighbors).

To reduce dramatically this number of tests in a simple way, we use the fact that  $S_{result}$  is expressed as intersections and subtractions of image sets. The idea is to initialize  $S_{result}$  with one of the image sets and then discard images which do not belong to the other image sets. This initialization avoids testing individually each image of the database and rather starts off with a set of potentially relevant images.  $S_{result}$  will be gradually reduced in the following order :

1. initialize  $S_{result}$  as the set  $\bigcup_{C \in N^\gamma(C_{pq_1})} IC(C)$ .
2. discard images in  $S_{result}$  which do not belong to any of the other union categories ( $i = 2, \dots, M$ ) to obtain the intersections of  $S_Q$  (see expression 2). At this point, we have  $S_{result} = S_Q$ .
3. to perform the subtraction of  $S_{NQ}$  from  $S_{result}$ , discard in  $S_{result}$  images which belong to any of the negative-query union categories ( $i = 1, \dots, R$ ) (see exp. 4). We get  $S_{result} = S_Q \setminus S_{NQ}$  (see exp. 5).

So gradually,  $S_{result}$  is reduced from  $\bigcup_{C \in N^\gamma(C_{pq_1})} IC(C)$  to  $S_Q \setminus S_{NQ}$ . By this approach, we will see in next section that a significant fraction of the database is not accessed at all. Note that some complex logical queries may yield an empty  $S_{result}$  set, i.e. no retrieved images. In this case, the user must loosen query constraints by either expanding the range radius values or discard some PQC or NQC, as he would do with a text retrieval system.

Indexing scheme is based on 3 hash tables. A table associates images with categories which contain its regions. An inverted table provides the reverse correspondence which gives direct access to relevant images given the labels query categories. And a third table lists neighbor categories of each category along with their distance.

The retrieval process is simply based on accesses to these 3 tables. It allows very fast search, first because no distance calculation between multidimensional features is involved, second because the inverted table provides direct access to relevant images given the labels query categories, and third because regions are not considered individually but as groups of similar regions.

## 5. RESULTS

Our system was tested with a 498 MHz Pentium PC and implemented within IKONA platform [1]. Two scenarios of application are investigated on two different databases : photoagency scenario on Corel database (9995 images) and a TV news scenario (910 images extracted from a TF1 video broadcast).

Evaluation of the precision of the retrieval by region categories composition can only be based on user satisfaction because query scenarios are too diverse in terms of range and composition and because it also depends on the user interaction in the query process especially if query refinement is involved.

The relevance of matched regions relies on the region extraction and grouping schemes. In retrieved images, regions from the positive query categories are salient. From the user point of view, false positives among matched regions are few and correspond to hard segmentation cases (complex composite natural images). In this case, a detected region may not be meaningful even if its mean color corresponds to a query category. Concerning the precision of composition matching in retrieved images, the simplicity of the indexing and retrieval scheme (comparison of category labels in image indexes) ensures a high user satisfaction. Indeed, in retrieved images, salient regions do satisfy the constraints of presence of regions from PQC's and the absence of regions of NQC's. Query can be refined by adjusting the range radius to widen or narrow the types of searched regions by considering more or less neighbor categories.

The image retrieval scheme is very fast first because it only involves accesses to hash tables (no feature vectors nor distance are involved) and second because it is *not region exhaustive* (in the case of the Corel scenario, we only deal with the 91 regions categories instead of the 50,220 regions) and third because it is *not image exhaustive* (only a fraction of the image database is accessed). On average on various query compositions, the fraction of accessed image entries is around 12%.

Retrieval process takes up to 0.03 second for the most complex queries on a 498MHz PC machine.

### 5.1. Photoagency application

50,220 regions were automatically extracted from the 9995 images. Clustering the 50,220 regions mean colors takes 150 seconds. 91 categories are automatically generated and their populations range from 112 regions to 2048 regions. Figure 2 illustrates two of these categories.

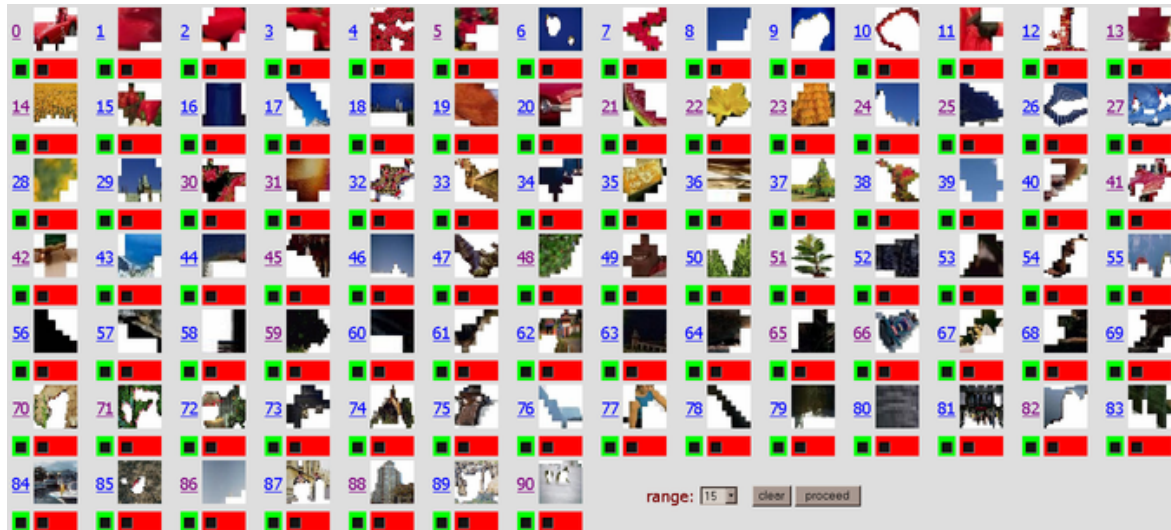
Since data are relatively dense in the color space and the granularity is fine, the CA algorithm provides categories homogeneous in mean color as expected. Intra-category variability is mainly due to the different textures which have the same mean color. All discriminant mean colors of the



**Fig. 2.** Region categories are homogeneous with respect to the feature used in categorization, i.e. mean color here. 2 region categories in Corel database : category 23 (top) contains regions which have a similar orange mean color and category 48 (bottom) corresponds to green mean color.



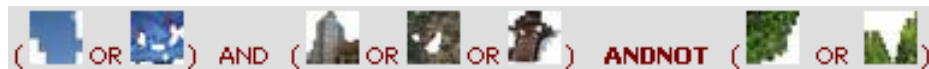
**Fig. 7.** Images rejected from the "cityscape" query : these images are rejected by the system due to the presence of a green region.



**Fig. 3.** Query interface : the 91 categories constitute the “region photometric thesaurus” of Corel database. Each category can be selected to form the query. No starting image or region is required. The content of each category can be seen by clicking on its corresponding representative.



**Fig. 4.** Example of query to retrieve “cityscapes” : range radius is set to default value, categories 39 (blue) and 88 (grey) are selected as Positive Query Categories and 48 (green) as a Negative Query Category. This query can be expressed as “blue region and grey region and not green regions”.



**Fig. 5.** Full expression of the logical composition of region categories : this expression is formulated by the system from the user query. Neighbor categories are shown and separated by disjunctions.



**Fig. 6.** Results : retrieved images satisfy the query composition (order of images is random). We can observe that the composition of presence of grey regions and blue regions and the absence of green regions is mostly matched by cityscapes and also by views of monuments or ruins.

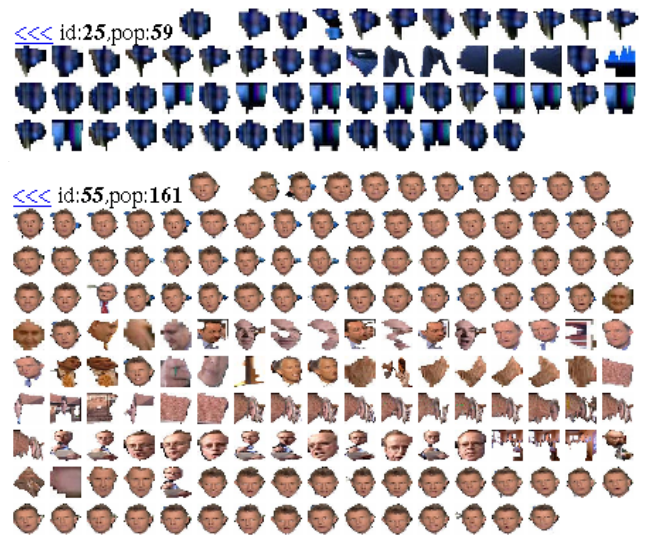
database are represented by at least one category : colors with high or low saturations but also grey levels. Some pairs of categories provided by the fine clustering granularity yield close representatives, but even when very close, two categories always differ in chrominance. Selecting any close category in this interface does not matter since the range radius integrates by default very close categories.

As illustrated in figure 3, the query interface presents to the user the 91 category representatives of the categories. Categories are sorted by increasing population. They provide an overview of the available types of regions in the database, in other words the “region photometric thesaurus” of this database. Each representative can be selected to specify that the corresponding category constitutes a PQC or NQC. No browsing is required to find any example image, unlike with existing CBIR systems. Any combination of query composition can be expressed from this interface. In the query window, the range box allows to adjust the range radius  $\gamma$  which will define interactively the neighbor categories.

Figure 4 then illustrates the query composition : “find images composed of sky-like and building-like regions but with no grass-like region”. In a photo agency context, the user may want to find, for instance, cityscapes. This type of scene can be retrieved by the following composition : “grey region and blue region and not green regions”. Given the range value, the system determines the possible neighbors of each query category and translates the query into a logical composition query (fig. 5). Figure 6 shows the set of relevant images retrieved for this query. In these images, this composition shows that grey regions mostly correspond to regions of buildings, monuments or rocks and blue regions to sky. The set of images rejected by the system due to the presence of a green region, contains mostly many nature landscapes as illustrated in figure 7.

## 5.2. TV news video application

The second scenario is related to the search in database of video frames extracted from a TV news broadcast (3 minutes, 910 frames extracted) from the French TV channel TF1. 6362 regions are extracted from the 910 images and 65 clusters are generated. Compared to the Corel database, the “region photometric thesaurus” contains less categories with saturated colors but more with black or blue categories which are characteristic from the TF1 news graphic chart. Two categories are illustrated in figure 8. A look at the individual content of categories shows that they contain homogeneous groups of identifiable parts such as : vegetation in the green category, black suit halves (left or right parts of hosts or interviewed people) or dark backgrounds in a black category, faces in the “pink-flesh” category, different parts of inlays in categories corresponding to different shades of a saturated blue. Categories can have a specific meaning



**Fig. 8.** 2 region categories from TF1 database : category 25 (top) corresponds to a dark blue mean color and contains mostly parts of the host background. Category 55 (bottom) corresponds to a pink-flesh mean color and contains mostly faces.

in the news scenario, corresponding to elements of the specific graphic chart of TV news program. Indeed, in figure 10, the categories containing faces is selected and also two other categories which refer to parts of the host background which has a characteristic mix of blue and black. And in the query in figure 11, the three neighbor categories depict shades of a saturated blue which is visually characteristic of parts of inlays.

The two types of presented queries correspond to a practical problems for archivists at TF1 : host detection (fig. 10) and inlay detection (fig. 11). These queries rely on a certain knowledge of the visual specificity of the domain. On that database, they illustrate that host frames or inlays can be retrieved from typical compositions.

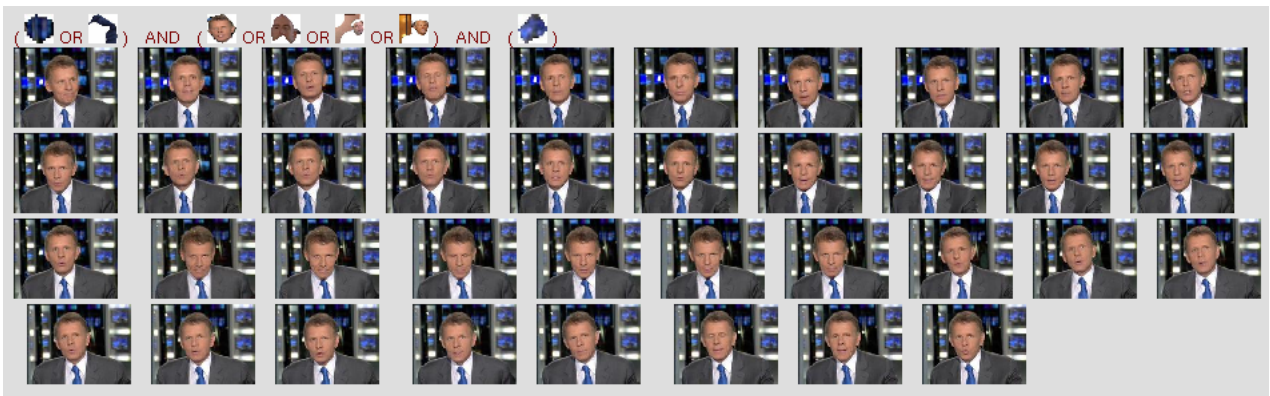
## 6. DISCUSSIONS AND FUTURE WORK

**Low complexity :** with existing region-based systems, the same query possibilities would be much heavier in terms of both user interaction and retrieval complexity. Finding by random browsing a set of example regions is tedious and comparing each example region to all regions in the database is computationally expensive. The complexity of our system is low because regions are not considered individually, but as groups.

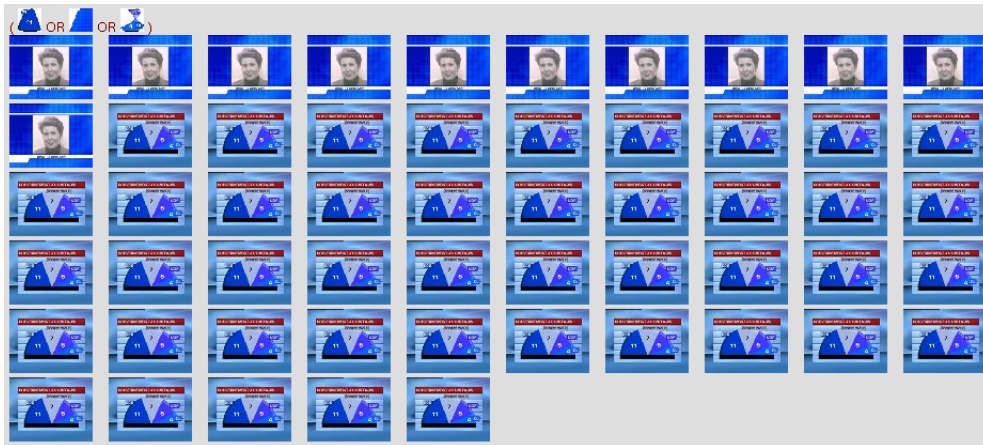
**Analogy with Text Retrieval :** when viewing categories as groups of similar regions, which are constituting units of images, this indexing can be considered as *symbolic* rather



**Fig. 9.** Query interface : the 65 categories constitute the “region photometric thesaurus” of TF1 database.



**Fig. 10.** Expression of logical query and corresponding results to retrieve host frames. 3 categories were selected : 2 categories corresponding to dark blue mean color contain typical elements of the background. And one category corresponding to “pink-flesh” mean color contains faces. It is interesting to note that semantically irrelevant regions (i.e. not faces) from the “pink-flesh” category are naturally rejected.



**Fig. 11.** Expression of logical query and corresponding results to retrieve inlays. Query consists in a bright blue category (and its neighbors), which is characteristic of inlays.



**Fig. 12.** Query refinement to ignore the diagrams by rejecting images with red regions.

than numerical. The following comparison with text retrieval can be made :

- image  $\rightarrow$  document
- region  $\rightarrow$  term
- region category  $\rightarrow$  concept
- neighbor categories  $\rightarrow$  synonymous concepts
- set of region categories  $\rightarrow$  thesaurus
- query by logical composition  $\rightarrow$  Google<sup>1</sup>-like query

**Perspectives :** simple and new, this framework can benefit from development from different fields such as visual description, text retrieval, unsupervised clustering, spatial indexing, database browsing. Our future work include : (1) integration of additional region description in the generation of categories to refine grouping, using ADCS [3], position, surface; (2) hierarchical clustering for the category generation. Levels in the hierarchy may correspond to different granularities of feature grouping or to groupings with respect to different features; (3) intuitive selection of  $\gamma$  radius and a more perceptual arrangement of the region photometric thesaurus in the query interface; (4) investigation of proven text-retrieval techniques to further meet users' need with the above mentioned text-retrieval analogy.

**Visual Thesauri in the literature :** some approaches introduced an idea of *Visual Thesaurus* of image blocks or regions [9, 6, 11]. Although different from one another, they all rely on a *supervised* learning process of visual features, either to represent user-driven visual groupings [9], or to learn domain-dependent visual similarity [6], or to learn visual description of predefined semantic classes [11]. On contrary, our approach is fully unsupervised and does not aim at finding the ideal or semantic similarity at the region level. We rather focus on the visual semantics which arises at the level of logical composition query (which is not addressed in those approaches) from the user expression.

## 7. CONCLUDING REMARKS

We have presented a framework to retrieve images based on logical composition of region categories. The system allows retrieving images by query composition like : “find images composed of regions of these types and not of those types”. The originality of this approach relies on the unsupervised grouping of similar regions into categories. In the interface, the set of categories forms a “region photometric thesaurus” from which the user can query images as he would query textual documents from logical expression of keywords. Simple and open to other fields, this framework is an interesting ground for further developments. It has the following specificities :

- image query by logical composition of region categories allows a specific search on image content
- no example region required to start query
- symbolic search and retrieval of images
- natural region range query by interactive definition of neighbor categories
- efficient visual indexing which results in very fast image retrieval

The constraint of composition in retrieved images expresses an underlying “visual semantics” in images by means of the user interaction. Various directions were proposed to further develop this new framework in terms of visual description, retrieval scheme, interface, category generation.

## 8. REFERENCES

- [1] N. Boujemaa, J. Fauqueur, M. Ferecatu, F. Fleuret, V. Gouet, B. Le Saux, and H. Sahbi. Ikona: Interactive generic and specific image retrieval. *International workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR'2001)*, Rocquencourt, France, 2001.
- [2] C. Carson and al. Blobworld: A system for region-based image indexing and retrieval. *Proc. of International Conference on Visual Information System*, 1999.
- [3] J. Fauqueur and N. Boujemaa. Region-based retrieval: Coarse segmentation with fine signature. *IEEE International Conference on Image Processing (ICIP)*, 2002.
- [4] M. Flickner and al. Query by image and video content: the qbic system. *IEEE Computer*, 28(9):23–32, 1995.
- [5] H. Frigui and R. Krishnapuram. Clustering by competitive agglomeration. *Pattern Recognition*, 30(7):1109–1119, 1997.
- [6] W. Y. Ma and B. S. Manjunath. A texture thesaurus for browsing large aerial photographs. *Journal of the American Society of Information Science*, 49(7):633–648, 1998.
- [7] W. Y. Ma and B. S. Manjunath. Netra: A toolbox for navigating large image databases. *Multimedia Systems*, 7(3):184–198, 1999.
- [8] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. In *SPIE Storage and Retrieval for Image and Video Databases*, II(2185), Feb. 1994.
- [9] R. W. Picard. Toward a visual thesaurus. *MIT Technical Report TR358*, 1995.
- [10] J. R. Smith and S. F. Chang. Visualeek: A fully automated content-based image query system. In *ACM Multimedia Conference, Boston, MA, USA*, 1996.
- [11] C. Town and D. Sinclair. Content based image retrieval using semantic visual categories. *ATT Technical Report*, 2001.
- [12] J. Z. Wang, Jia Li, and Gio Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2001.

<sup>1</sup>Google: <http://www.google.com>