Mental Image Search by Boolean Composition of Region Categories

Julien Fauqueur and Nozha Boujemaa
{Julien.Fauqueur[1] ,Nozha.Boujemaa}@inria.fr

Projet IMEDIA - INRIA, BP 105
78153 Le Chesnay Cedex - FRANCE

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

Abstract

Existing content-based image retrieval paradigms almost never address the problem of starting the search, when the user has no starting example image but rather a *mental image*.

We propose a new image retrieval system to allow the user to perform *mental image search* by formulating boolean composition of region categories. The query interface is a *region photometric thesaurus* which can be viewed as a visual summary of salient regions available in the database. It is generated from the unsupervised clustering of regions with similar visual content into categories. In this thesaurus, the user simply selects the types of regions which should and should not be present in the mental image (boolean composition). The natural use of inverted tables on the region category labels enables powerful boolean search and very fast retrieval in large image databases. The process of query and search of images relates to that of documents with Google. The indexing scheme is fully unsupervised and the query mode requires minimal user interaction (no example image to provide, no sketch to draw).

We demonstrate the feasibility of such a framework to reach the user mental target image with two applications : a photo-agency scenario on Corel Photostock and a TV news scenario. Perspectives will be proposed for this simple and innovative framework, which should motivate further development in various research areas.

[1] The first author is currently a Research Associate with the Signal Processing Group at the University of Cambridge (UK) and is supported by the Defence Technology Consortium on Data and Information Fusion. His current email is jf330@cam.ac.uk.

Preprint submitted to Multimedia Tools and Applications *4 August 2004*

1 Introduction

In Content-Based Image Retrieval (CBIR) context, the earliest and most common approach is the *global query-by-example* paradigm (GQbE). It consists in retrieving images whose visual appearance is globally similar to a selected example image. Initially proposed by Swain and Ballard [50], it was then adopted by a vast majority of CBIR systems [16,39,20,22,44,37]. However this paradigm has a narrow scope of usage (such as checking if a logo is in the database [6]). While it has helped prove the feasibility of CBIR at early years, it is now rather used for testing purpose to evaluate similarity measures and visual descriptors performance.
Partial query by example paradigms were later introduced. They allow the user to explicitly select an image component which is relevant for the query and retrieve images which contain a similar visual component. This approach proved to be more selective, hence more precise than GQbE. As reviewed in [13], image components have been defined either by fixed block subdivision [36,33], manual outline [8], selection of points of interest [19], histogram backprojection [50,47] or region segmentation [4,30,15].
To refine image search, the *relevance feedback* mechanism inspired from text retrieval was successfully applied to CBIR [55,35,6]. Among the retrieved images, the user specifies the ones which are relevant and nonrelevant and reiterates the search. By refining the similarity measure, the searched image can be reached more efficiently than with the GQbE alone, because it takes into account the subjective preference of the user.

These query paradigms consider the image retrieval problem as a matching problem between a visual example (image, group of images or regions) and visual entities in the database. However in practice the user rarely has a relevant example image to start the search, but rather a *mental image* [23]. The prior search by random browsing for the example itself can be tedious - if not impossible - as visual queries can be complex and image database content very heterogeneous. In such a context, when no starting example is available, how can the user reach his/her mental image?

This problem has been referred to as the *Page Zero Problem* by La Cascia et al. [5] and also raised by MacDonald and Tait who concluded from a user study [31] that an *entry point* is still missing in image search systems. A starting example obtained by random browsing is very likely to lead to unsatisfactory results. In other words existing CBIR techniques are successful only if the user has a relevant starting point. Alternatively visual browsing techniques (such as [43,21,25,42,26,51]) help providing an overview of database but make sense for image search only if the goal is vague [46]. Indeed, if the user has a target image in mind, these techniques still lack an entry point.

We define the **mental image search paradigm** as a search mode which allows the user to access a set of relevant images *directly* without using specific image examples. Only two approaches in the literature seem to implement this paradigm : target image search [6] and query by sketch [38,48,22,7]. The target image search process proposed in PicHunter [6] asks the user to iteratively select images which are similar to the target image. A simple Bayes's rule is used to predict the target image, given the user actions. The iterative display strategy is designed to maximize the information obtained from the user. Although this paradigm allows the user to reach his target image, it does not solve the page-zero problem. With the query-by-sketch paradigm, the user draws a sketch which resembles his/her mental image. The user does not have to provide an example image. A prototype image can be created ex nihilo where spatial layout of regions can be specified as well as their shape and color. It serves as an example image which is matched against the database.

We propose here a new approach to implement the mental image search paradigm : the **boolean composition of region categories**. It differs completely from existing frameworks on both query and retrieval processes. It provides a solution to the page zero problem. As it directly retrieves a set of relevant images it can be used alone for approximate image search but also as a database filtering tool to provide a good initial start for query by example techniques. The query interface consists of a visual summary of image regions available in the database which constitutes a *region photometric thesaurus* - it is the page zero. The user can directly specify the boolean composition of the target mental image by selecting the *types of regions* which should and should not appear. Types of regions correspond to visual concepts and are defined as categories of regions with similar visual content. Region categories are obtained by clustering the region visual descriptors. 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". To support these queries, a new symbolic indexing and querying approach is presented which is equivalent to well-known mechanisms in information retrieval. Simple and open, it can be easily extended to multimedia document retrieval indexed by physical content descriptors. Seminal work of this approach was published in [14].

As depicted in figure 1, the system workflow is the following : 1) image components are first detected by unsupervised segmentation into salient regions, 2) visual descriptors are then extracted for each region and grouped into visual categories by unsupervised clustering of all region descriptors, 3) indexing tables are built which associate images and category labels, 4) for each category a representative region is defined by its category prototype, and the visual thesaurus is constructed from all category representative regions, 5) the user formulates the query in the thesaurus by selecting relevant regions, 6) target images are determined by operations on the index tables, 7) the system determines images which satisfy the boolean query and displays them, 8) and finally, in the result interface, the user may choose to modify the boolean formulation to refine the search.

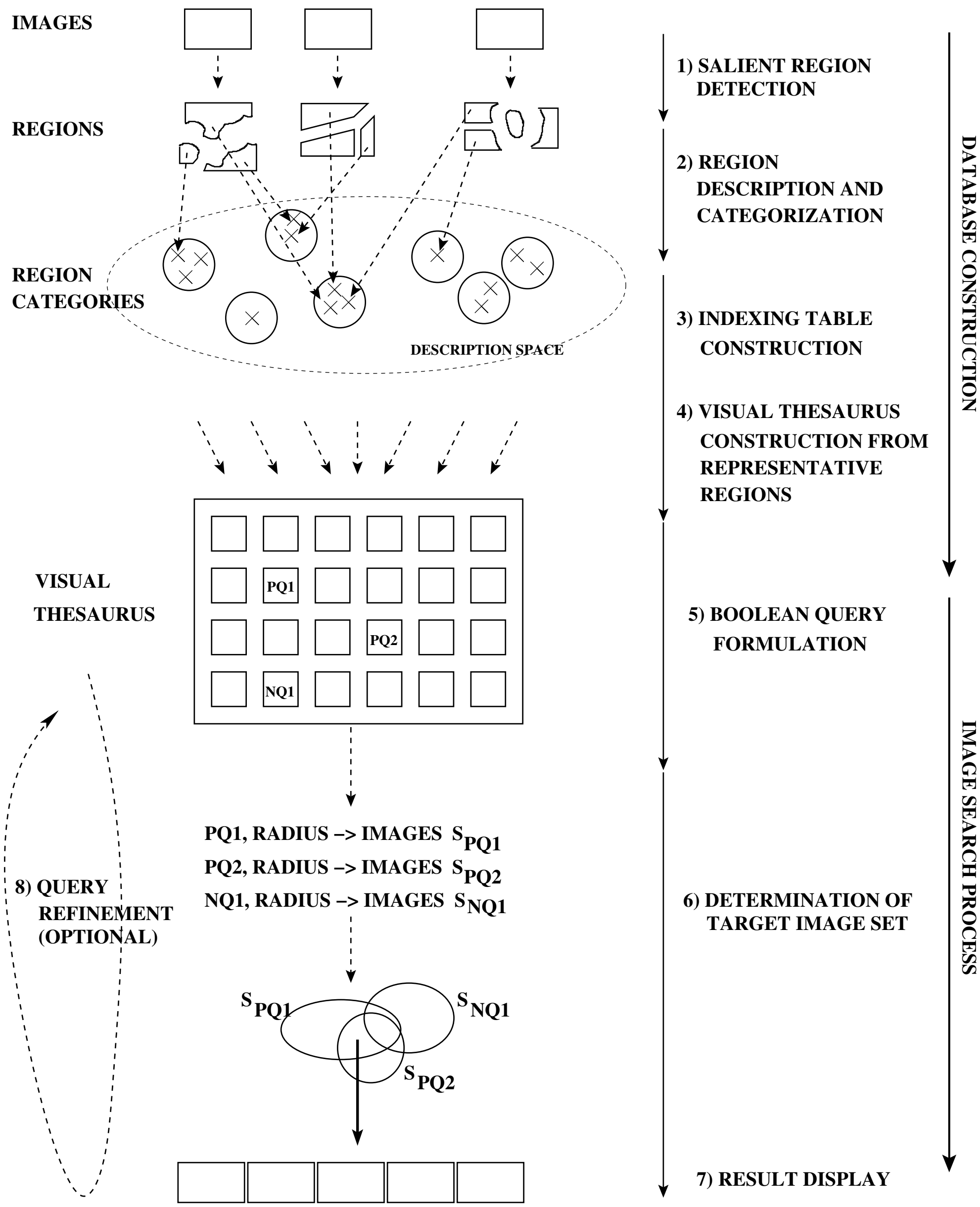
IMAGES
REGIONS
REGION CATEGORIES
DESCRIPTION SPACE
VISUAL THESAURUS
PQ1
PQ2
NQ1
PQ1, RADIUS –> IMAGES S_{PQ1}
PQ2, RADIUS –> IMAGES S_{PQ2}
NQ1, RADIUS –> IMAGES S_{NQ1}
S_{PQ1}
S_{NQ1}
S_{PQ2}
8) QUERY REFINEMENT (OPTIONAL)
1) SALIENT REGION DETECTION
2) REGION DESCRIPTION AND CATEGORIZATION
3) INDEXING TABLE CONSTRUCTION
4) VISUAL THESAURUS CONSTRUCTION FROM REPRESENTATIVE REGIONS
5) BOOLEAN QUERY FORMULATION
6) DETERMINATION OF TARGET IMAGE SET
7) RESULT DISPLAY
DATABASE CONSTRUCTION
IMAGE SEARCH PROCESS

Fig. 1. System workflow

In section 2, we will explain the thesaurus construction process. We will present in section 3 the symbolic indexing scheme and the use of neighbor categories. Then, in section 4, we will detail the retrieval scheme to match a boolean composition of region categories. In section 5, we will present the results on two application scenarios and discuss the evaluation issues of this approach, the “Image Google” aspect and how it is positioned with respect to existing techniques and notions. In section 6 various perspectives will be proposed and concluding remarks will be addressed.

2 Photometric thesaurus construction

In this section we describe the generation of the region photometric thesaurus which consists of three steps : salient region detection, region photometric description and categorization of regions by grouping their visual descriptor.

2.1 Salient region detection

An image is viewed as a composition of salient regions. We are not interested in visual details or tiny regions. Salient regions are detected by the coarse region segmentation algorithm proposed in [15]. It relies on the unsupervised clustering of a rich local color primitive, the local distribution of quantized colors. Regions are homogeneous with respect to this color variability primitive. This technique has been successfully applied to implement the partial query by example paradigm in the Ikona system [15]. It does not aim at performing object recognition nor a perfect semantic segmentation, but it rather detects coarse and visually salient regions which constitute intuitive search keys for the user.

2.2 Photometric region description

In the CBIR context various region photometric descriptors have been proposed in the literature [13] : mean color [22], color distributions [4][10][34][15], and texture [34][4][10]. For the proof of concept of our new framework the simple mean color descriptor proved to be sufficient and intuitive to produce an overall color thesaurus for database filtering. Region categories are hence formed of regions with similar mean color. The descriptor is the average of the region pixel values after transformation into the *Luv* space which is chosen for its perceptual uniformity. It is important for the reader to keep in mind that any other region visual descriptor can be used in the whole presented approach. The possible use of more specific descriptors will be discussed in section 6.

2.3 Visual grouping and thesaurus generation

Once all regions in the database are detected and their visual descriptors extracted, regions are grouped into categories by unsupervised clustering of their descriptors.

We want the clustering algorithm to estimate automatically the number of categories so that the thesaurus reflects the database diversity. The CA (for *Competitive Agglomeration*) algorithm, originally presented in [17], is chosen because of its major advantage of determining automatically the number of categories. Note that in well known algorithms from the k-means family (k-means [32], Linde-Buzo-Gray / Generalized Lloyd Algorithm [28], Expectation-Maximisation [9], Fuzzy C-Means [2]), the number of clusters is supposed to be given. When the number of clusters has to be estimated, these algorithms need to be run several times for different numbers of clusters and a criterion, such as the minimum description length [41], is used to determine the optimal number. By requiring only one clustering pass, CA is much more computationnally efficient. We invite the reader to refer to [13] for a more detailed overview of these algorithms.

A brief description of CA algorithm is given below. We call $\{x_j, \forall j = 1, ..., N\}$ the set of region descriptors we want to cluster and P the number of categories. $\{p_i, \forall i = 1, ..., P\}$ denote the prototypes to be determined. $d(x_j, p_i)$ is the Mahalanobis distance between descriptor x_j and prototype p_i. The CA-clustering is performed by minimizing the following objective function J :

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

Subject to membership constraint : $\sum_{i=1}^{P} u_{ij} = 1, \forall j = 1, ..., N$, where u_{ij} represents the fuzzy membership of descriptor x_j to cluster i. As detailed in [17], the global minimum of first term is achieved when each cluster contains a single data point. The global minimum of the second term (including the negative sign) is achieved when all points are lumped in one cluster, and all other clusters are empty. When both components are combined the final partition will minimize the sum of intra-cluster distances, while partitioning the data set into the smallest possible number of clusters. Automatic estimation of the number of clusters is achieved by iteratively discarding spurious clusters.

At convergence the CA algorithm provides the following output : P the number of clusters, $\{p_1, ..., p_P\}$ the cluster prototypes (in the region description space) and $U = \{u_{ij}\}$ the fuzzy membership values between the data (the region descriptors) and the clusters (region categories). The P clusters define the **region categories** and are labelled $\{C_1, ..., C_P\}$.

For each region category C_i, we define its **representative region** r_i as the region whose prototype is the closest to its prototype p_i. The set $\{r_1, ..., r_P\}$ defines the **region photometric thesaurus** (RPT). It provides the user an overview and a visual summary of all regions available in the database. Note the RPT is both descriptor and database dependent. As we will see in section 5.1.1, the RPT will constitute the query interface from which the user will select the categories of regions which compose his/her mental image.

Two examples of region photometric thesaurus will be shown in section 5 on two different databases (figures 9 and 16).

3 Image symbolic indexing and neighbor categories

We present in this section an image indexing scheme which is similar to keyword document indexing in the classic Information Retrieval framework [1]. To comply with the specific nature of our visual data a new *range-query* mechanism will be integrated in the indexing scheme by means of *neighbor categories*.

As we will see in more details in section 4, a user query consists of the selection of region categories which should be present and absent in the retrieved images. To satisfy such a query we propose an image indexing scheme which solely relies on the labels of the P categories $\{C_1, ..., C_P\}$.

We first introduce two indexing tables $IC(C)$ and $CI(I)$ which provide association between images and categories. The table $CI(I)$ associates an image with categories which contain its regions. The table $IC(C)$ is constructed as the inverted table of $CI(I)$ to provide the reverse correspondence; it gives direct access to images which contain a region in a given category. Figures 2 and 3 illustrate examples of these tables for the Corel database which has 9,995 images (labeled from 0 to 9994) and 91 categories (labeled from 0 to 90). For example in the indexing table $CI(I)$ (figure 2), we have $CI(4) = \{53, 56, 58, 84, 90\}$ which means that image 4 has one of its regions in category 90. And conversely in the indexing table $IC(C)$ (figure 3), we have $IC(90) = \{4, 72, ..., 9458\}$ which shows the reverse correspondence : category 90 has a region which composes image 4. Given a set of selected regions, these two tables allow the system to directly know what images in the database are composed of regions from these categories. This is the inverted file mechanism well known in text retrieval in which documents are indexed by the keywords they contain. Inverted tables are a simple and efficient way to provide the list of documents which contain a given keyword.

This indexing scheme assumes that one region category represents the type of region (e.g. a patch of sky, building, face, background) which is in the user's mind. This holds only if the user has a precise idea of the target region (e.g. a particular shade of blue) because of the visual homogeneity of the selected category. The mental image can be precise in the case the user has already seen the target image and has a vivid memory of it. On the other hand the mental

images	categories
000000	36 47 62
000001	64 72 78 83
000002	5 30 38 56 56
000003	40 56
000004	53 56 58 84 90
:	
009993	2 13 15 45 56 69
009994	0 37 49 58 68 78 83

Fig. 2. Table $CI(I)$ associates categories to images. We have one row per image. For each image its corresponding category labels are stored.

categories	images
0	001082 001866 ... 009994
:	
90	000004 000072 ... 009458

Fig. 3. Table $IC(C)$ associates images to categories (it is the inverted table of $CI(I)$). We have one row per category. For each category its corresponding images are stored.

image can be vague if the user has seen it but has an approximate memory of it or if the search goal is intentionally broad (e.g. cityscapes). To enable a *user-dependent precision search* we propose to expand this indexing scheme by implementing a “range-query” mechanism. When the user will select a region category, extra similar categories (the *neighbor categories*) will also be matched. In this way more or less broad visual concepts can be specified in the query.

We define a **neighbor category** of a category C_q of prototype p_q as a category C_j whose prototype p_j satisfies $d(C_q, C_j) =\| p_q - p_j \|_{L^2} \leq \gamma$, for a given range radius threshold γ. 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 γ is selected by the user at retrieval phase depending on the required precision of search : for a given category A, increasing γ will increase the number of categories that are integrated in the search (as illustrated in figure 4). Note that neighbor categories are defined from distances between category prototypes, so their integration in the search process defines a range query mechanism in the description space.

In addition to $IC(C)$ and $CI(I)$ a third indexing table $N(C)$ is built to integrate this range-query mechanism. For each category C_q, the ordered set $N(C_q)$ contains the list of distances to all categories sorted by increasing order of distance values, i.e. $N(C_q) = ((C_j, d(C_q, C_j)), \forall j = 1, ..., P)$ with $d(C_q, C_j) \leq d(C_q, C_{j+1})$, where $d(C_q, C_j) =\| p_q - p_j \|_{L^2}$. Figure 5 illustrates an example of this table for the Corel database.

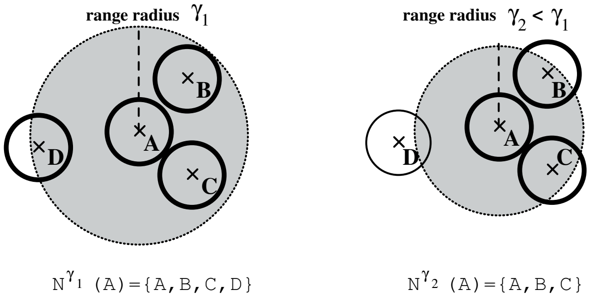
range radius γ_1
B
A
D
C
range radius $\gamma_2 < \gamma_1$
B
A
D
C
N^{γ_1} (A)={A,B,C,D}
N^{γ_2} (A)={A,B,C}

Fig. 4. Range radius and neighbor category selection : depending on the range radius (γ_1 or γ_2), category A has more or less neighbor categories ($N^{\gamma_1}(A)$ or $N^{\gamma_2}(A)$) which results in a more or less broad search.

categories	neighbor category and prototype distances
0	(0,0.00) (12,11.13) ... (9,164.47)
:	
90	(90,0.00) (89,15.06) ... (3,153.87)

Fig. 5. Table $N(C)$ associates neighbor categories to categories. We have one row per category. For each category its corresponding neighbor categories are stored as pairs (category label, prototype distance to the category).

As a summary, the indexing scheme relies on the three indexing tables $N(C)$, $CI(I)$ and $IC(C)$ which provide associations between images and categories and between categories and their neighbor categories (see illustration in figure 6).

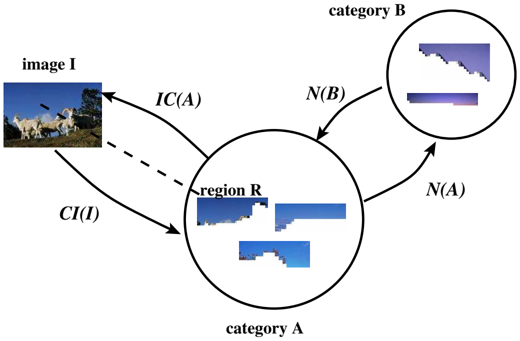

Fig. 6. Symbolic indexing relies on three tables of association (N, IC and CI) between images, categories and neighbor categories.

4 Query by boolean composition of region categories

We explain the boolean search process using the region categories indexing tables.

The region photometric thesaurus constitutes the query interface and allows the user to express a boolean query such as : “Find images composed of region of this type and this type but with no region of this type”. The user selects the set of region categories which must be present in the target image (which we call the *Positive Query Categories*) and the set of those which must be absent (the *Negative Query Categories*). The Positive Query Categories are referred to as **PQC**s and denoted as $\{C_{pq_1}, ..., C_{pq_M}\}$. The Negative Query Categories are referred to as **NQC**s and denoted as $\{C_{nq_1}, ..., C_{nq_R}\}$. The user chooses the range radius γ value which corresponds to the level of search precision he/she is expecting. A query is fully determined by the list of PQC labels $\{pq_1, ..., pq_M\}$, the list NQC labels $\{nq_1, ..., nq_R\}$ and the value of γ.

The translation of a query into a boolean expression is straightforward. The specification of presence of region categories is expressed by the AND operator between the PQCs and the absence by the NOT operator between NQCs. For a given range radius γ and each selected query category C (PQC or NQC), the range query will consist in searching images which have a region in C or in any of its neighbor categories in $N^\gamma(C)$. So the range query mechanism is expressed by the OR operator between each category and its neighbors. We can now express a full query as a boolean expression using AND, OR, NOT operators on the category labels :

$Q = (C_{pq_1}$ OR its neighbors) AND ... (C_{pq_M} OR its neighbors) AND NOT
(C_{nq_1} OR its neighbors) AND NOT ... (C_{nq_R} OR its neighbors)

The image retrieval scheme consists in finding images which satisfy this *boolean query composition*. In the information retrieval framework it is directly equivalent to retrieving documents which satisfy a boolean query of keywords. A major advantage of inverted files is that they make boolean search straightforward [54]. For our problem, the boolean query Q is processed as set operations on the inverted files $IC(C)$, which are image sets, and using the indexing tables $N(C)$, $CI(I)$.

For each given PQC and NQC query category C and a given range radius γ, the set $S_N^\gamma(C)$ of images which have a region in C or its neighbors is :

$$S_N^\gamma(C) = \bigcup_{C' \in N^\gamma(C)} IC(C') \qquad (2)$$

Note that $N^\gamma(C)$ is determined in the retrieval phase using the indexing table $N(C)$ which contains the prototype distances $d(C, C')$ (defined in the previous section) as follows :

$$N^\gamma(C) = \{C' \in N(C) \mid d(C, C') \leq \gamma\} \qquad (3)$$

Then the set S_Q of images which have a region in C_{pq_1} or its neighbors and

... a region in C_{pq_M} or its neighbors is :

$$S_Q = \bigcap_{i=1}^{M} S_N^{\gamma}(C_{pq_i}) \tag{4}$$

For the specification of absence of regions we introduce the set S_{NQ} of images which have a region in C_{nq_1} or its neighbors and ... a region in C_{nq_R} or its neighbors :

$$S_{NQ} = \bigcap_{i=1}^{R} S_N^{\gamma}(C_{nq_i}) \tag{5}$$

The final set S_{result} of relevant images, i.e. 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} :

$$S_{result} = S_Q \setminus S_{NQ} \tag{6}$$

To evaluate the expression of S_{result}, we use the fact that it is expressed as intersections and subtractions of image sets. S_{result} is initialised as one of the image sets. Then, to process intersections (respectively subtractions), we discard from it images which do not belong (resp. which 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} is determined through the following steps :

- initialize S_{result} as the set $S_N^{\gamma}(C_{pq_1})$.
- discard images in S_{result} which do not belong to any of the other sets $S_N^{\gamma}(C_{pq_i})$ for $i = 2, ..., M$ using the indexing table $CI(I)$. At this point, we have $S_{result} = S_Q$.
- to perform the subtraction of S_{NQ} from S_{result}, discard in S_{result} images which belong to any of the other sets $S_N^{\gamma}(C_{nq_i})$ for $i = 1, ..., R$ using the indexing table $CI(I)$. We get $S_{result} = S_Q \setminus S_{NQ}$.

So gradually, S_{result} is reduced from $S_N^{\gamma}(C_{pq_1})$ 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 boolean 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, as in the text retrieval context, by discarding some PQC or NQC.

We draw the reader's attention to the fact that the retrieval process involves no distance computation. It is simply based on accesses to the three indexing tables.

5 Application and Discussion

In this section two scenarios of application are investigated on two different databases : a photoagency scenario on Corel database and a TV news scenario. The thesaurus interface will be detailed in the Corel scenario. Then evaluation issues of our approach will be discussed in section 5.3. In section 5.4 we will explain why this approach can be considered as an “Image Google”. Other work involving related techniques will be detailed in section 5.5.

5.1 Photoagency application

We first tested our approach on 9,995 images of the Corel database. As samples show in figure 7 the content of this database is heterogeneous : landscapes, portraits, objects, flowers, cars, animals, kitchens, food, etc.

Fig. 7. Overview of Corel database.

After image segmentation (see section 2.1) 50,220 regions are automatically extracted from the 9,995 images. Clustering the 50,220 region mean color descriptors takes 150 seconds and produces 91 categories. Category populations range from 112 regions to 2,048 regions. Figure 8 illustrates two of these categories. As expected CA generates categories which are homogeneous with respect to the region mean color descriptor. The perceptual difference between regions within a category is due to regions which have similar mean color with different textures.

5.1.1 Query interaction with the thesaurus interface

The query interface is based on the region photometric thesaurus (see section 2). It is composed of the 91 category region representatives which provide an overview of the available types of regions in the Corel database (see figure

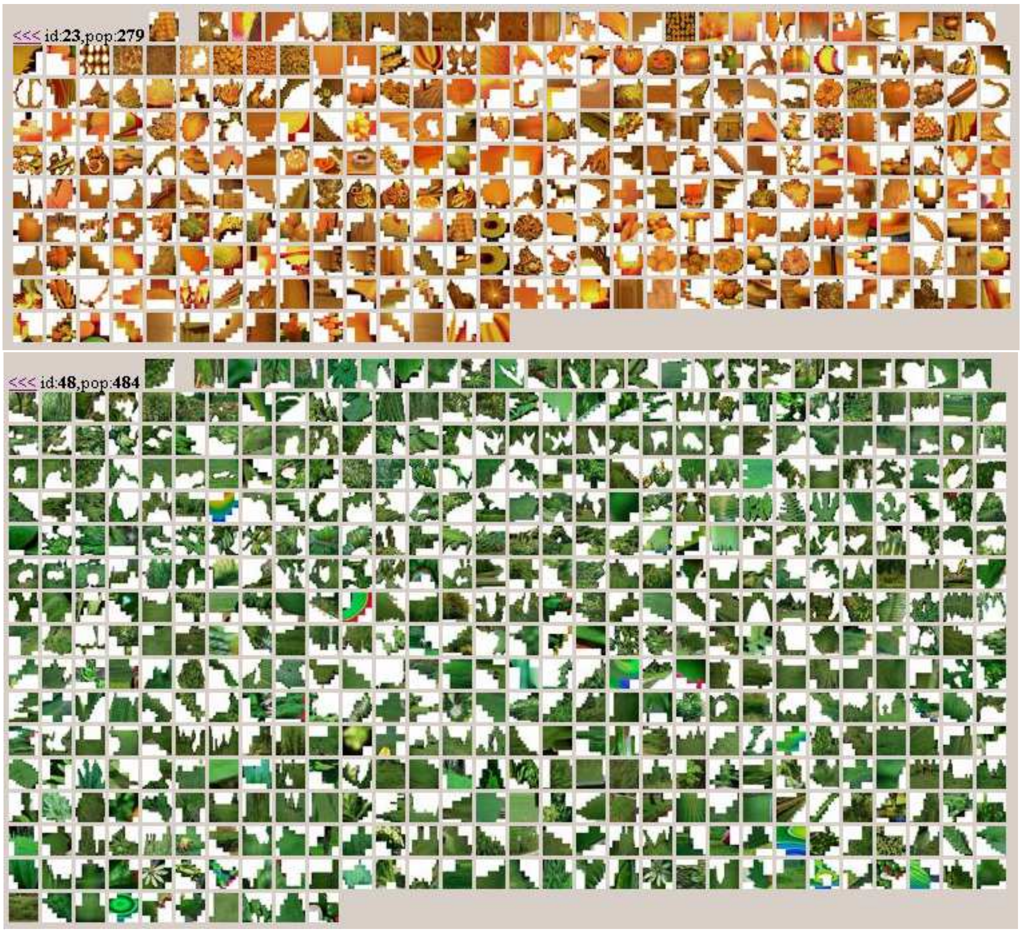

Fig. 8. Two among the 91 Corel region categories : category 23 (top) contains regions which have a similar orange mean color and category 48 (bottom) with regions of similar green mean color.

9). To facilitate the user selection of relevant categories in the thesaurus, representative regions are disposed in a grid such that similar categories lie near each other. This is achieved by re-arranging representative region descriptors in a bidimensional grid. Dimensionality reduction and topology preservation are obtained with a Kohonen map in its classical stochastic version [24]. Note that a similar approach was used in PicSOM [25] to arrange *images* (rather than regions) in a grid. The input data for Kohonen classification are the N category prototypes corresponding to the N representative regions. Output Kohonen map is a $p \times q$ grid, such that $p \times q = N$. The N category prototypes are trained with N Kohonen prototypes. Note that empty cells in the thesaurus are due to empty classes in the Kohonen classification and they shall be retained to preserve topology.

To select the set of PQCs (respectively NQCs) categories the user ticks the green box (resp. red box) below the corresponding representative regions. The value of range radius γ is selected from the range box. Once the relevant PQC and NQCs and the γ value are selected, the query is submitted by pressing

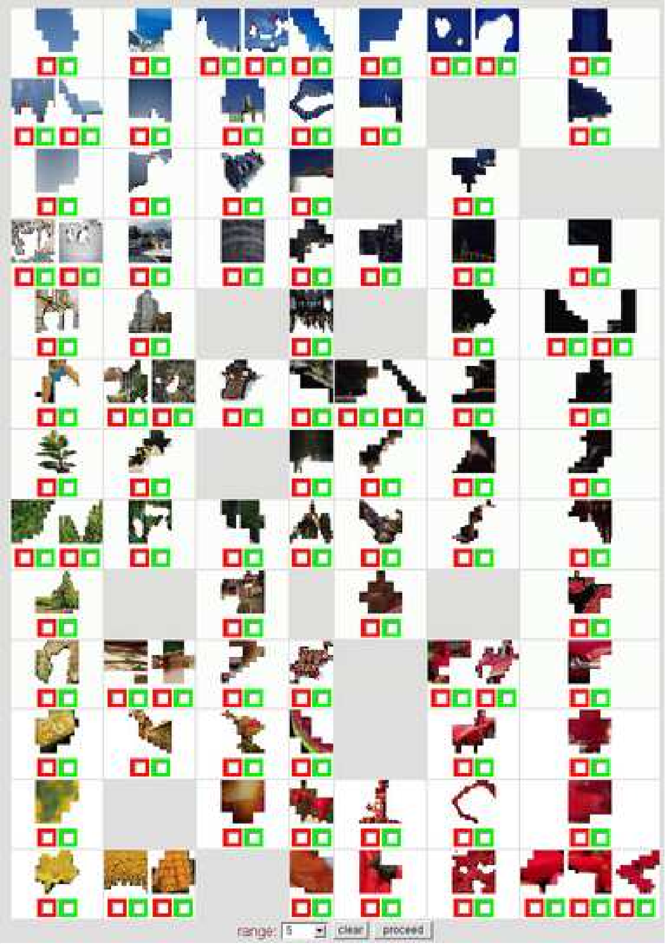

Fig. 9. Query interface : the 91 categories constitute the region photometric thesaurus of Corel database. Each category can be selected to form the query. The content of each category can be seen by clicking on its corresponding representative for region browsing.

the "proceed" button. A *region browsing feature* is also provided : by clicking a representative region in the thesaurus, the full content of the corresponding category is displayed the user as in figure 8.

In the thesaurus some region representatives look very similar although they are pairwise different. Indeed, by construction categories do not overlap in the region description space, because at the end of CA iterations regions are assigned to the closest class prototype. The user will choose γ depending on the level of precision of his/her mental image as explained in section 3. From expression 3, if γ is set to 0 only the very selected PQCs and NQCs will be matched. For non-zero low values of γ (i.e. precise visual concept) categories which look very similar in the thesaurus will be also integrated in the query. As γ is increased (i.e. broader visual concepts), additional categories which are less similar will also be integrated.

5.1.2 *Query scenario*

Let us now consider a query composition scenario. To find cityscapes in a photo agency context, the user may want to search images with a building, some sky, and no vegetation. In the region photometric thesaurus this query can be expressed by the following composition : "grey region *and* blue region *and no* green region" (see figure 10). Given the range value, the system determines the possible neighbors of each query category (grey, blue, green) and translates the query into a boolean composition query (fig. 11).

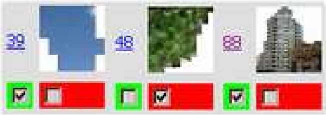

Fig. 10. Example of query to retrieve "cityscapes" : categories 39 (blue-like) and 88 (grey-like) are selected as PQCs and 48 (green-like) as a NQC.

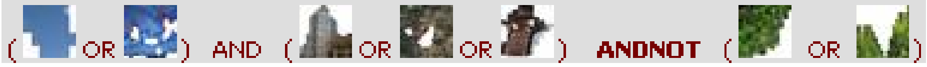

Fig. 11. Expression of the boolean query composition of region categories.

Figure 12 shows the set of relevant images retrieved for this query. Note that retrieved images are displayed in *random order*. From a visual point of view all retrieved images are relevant to the boolean query since they do contain a grey and a blue region and no green region. From the semantic point of view retrieved images contain many cityscapes but also pictures of ruins, monuments. False positives correspond to scenes which match the visual composition but are semantically irrelevant such as a painting or seascapes. Such false positives can be easily rejected by using extra features such as texture and spatial information. This mental search is successful since it filtered out the database to provide a first set of cityscape images. If a more precise search on cityscapes is needed the retrieved images constitute a satisfactory starting point for a query by example or relevance feedback search. Figure 13 shows the images which were rejected for this query due to the presence of a green region (in addition to a blue and grey region). It is interesting to observe that almost all these rejected images depict natural landscapes which are semantically opposite to the query for "cityscapes".

We draw the reader's attention to the fact that each single category defines a visual concept (a type of grey for instance) and hence is likely to contain regions with heterogeneous semantics : e.g. the grey categories for "building" also contain shade areas or rock regions; the blue categories for "sky" also contains cloth, car or swimming pool regions; the green categories also contains plant or vegetable close-ups or sweets. In spite of the semantic heterogeneity of categories, we observe that many regions in the retrieved images have a semantic relevance to the cityscape query. We observe that the boolean constraint in the user query composition helps the implicit selection of regions which are semantically relevant for the query within each query category.

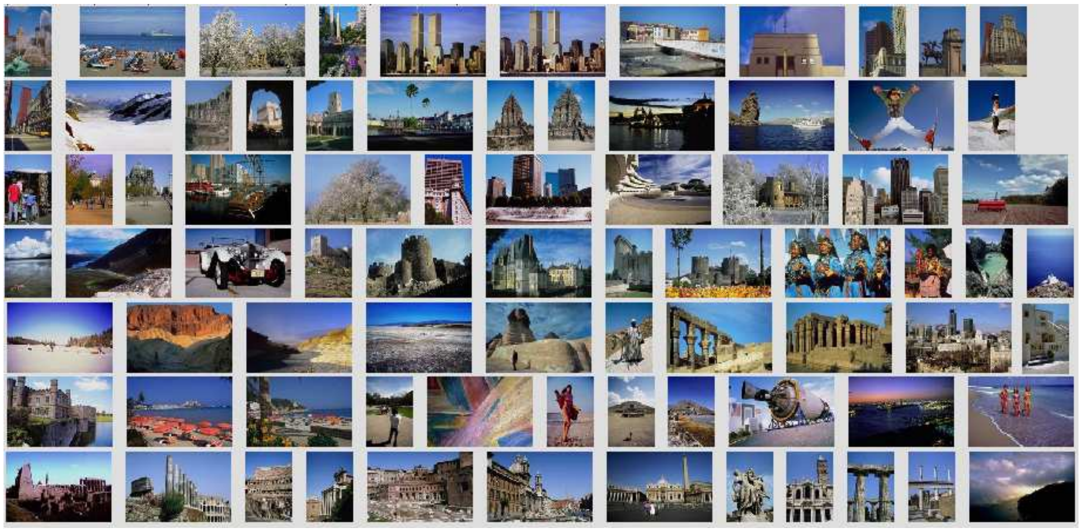

Fig. 12. Results for "cityscape" query (display order is random) : many images which matched the composition of presence of grey regions and blue regions and the absence of green regions depict cityscapes and also monuments or ruins.

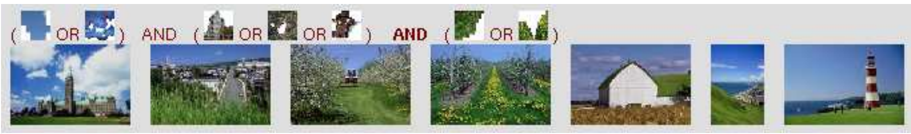

Fig. 13. Images automatically rejected by the "cityscape" query : these images are rejected due to the presence of a green region. They turn out to correspond to landscapes, i.e. semantically opposite to "cityscapes".

All our tests were performed on a 498 MHz Pentium PC and implemented within Ikona platform [3]. The image retrieval scheme is very fast : up to 0.03 second for complex boolean queries (with up to five query categories and a non-zero range radius). On average on various query compositions, the fraction of accessed image entries is around 12%. The storage cost is about one megabyte for all three indexing tables $IC(C)$, $CI(I)$ and $N(C)$ for the Corel database.

5.2 TV news video application

Our second scenario is related to the search in a database of video frames from a TV news broadcast (3 minutes, 910 frames extracted) from the French TV channel TF1. A sample of these video keyframes is shown in figure 14. The TF1 thesaurus is generated in the same way as for Corel database. $6,362$ regions are extracted from the 910 images and 65 categories are generated. Two categories are illustrated in figure 15. The region photometric thesaurus (see figure 16) contains less categories with saturated colors compared to the Corel thesaurus, but more with black or blue categories which are characteristic from

Fig. 14. Overview of TF1 database.

the TF1 news graphic chart. A look at the individual content of categories shows that they contain subgroups of identifiable parts : 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 pinkish category (shown in figure 15), different parts of inlays in categories corresponding to different shades of a saturated blue (shown in figure 15). Parts of categories can have a specific meaning in the news scenario, corresponding to elements of the specific graphic chart of TV news program.

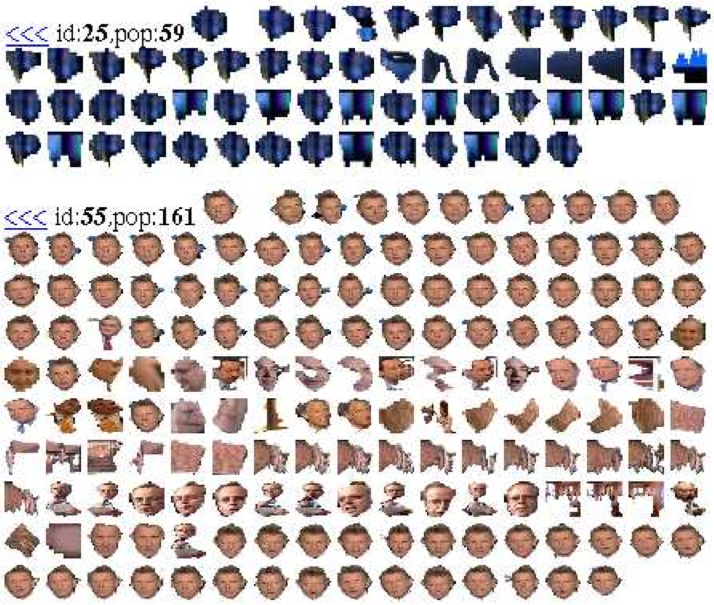

Fig. 15. Two of the 65 TF1 region categories : category 25 (top) corresponds to a dark blue mean color and category 55 (bottom) corresponds to a pinkish mean color.

We present two query scenarios which correspond to two practical problems expressed by archivists at TF1 : host detection and inlay detection. These queries assume 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. To retrieve host frames, the query is formulated as conjunctions of PQCs (see figure 17) which may match a face region and the host background which has a specific dark blue color in the TF1 graphic chart. The selected pinkish PQC contains a majority of faces. The two others dark blue PQCs contain background patches of the characteristic host background. The content of two of these PQCs is shown in figure 15. Figure 17 shows that this query retrieves exclusively the host frames although each selected PQC also contains regions which are not semantically relevant for the query, i.e. non-faces in the pinkish category and its neighbors and non-host background patches in the dark blue category and their neighbors. This supports the observation made in the Corel scenario, that the boolean query composition helps the selection of semantically relevant regions. In the inlay retrieval scenario, the query consists of one category (and its neighbors) of a bright blue color which is characteristic of the TF1 graphic chart. The results

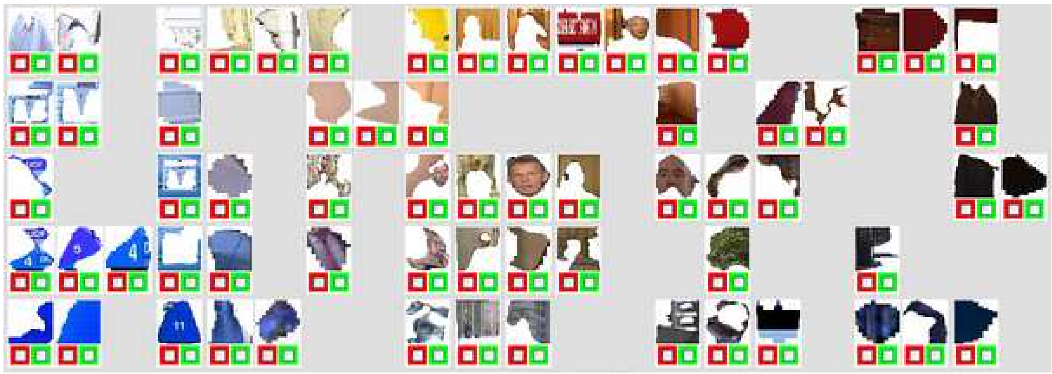

Fig. 16. Query interface : the region photometric thesaurus for the TF1 database is composed of 65 categories.

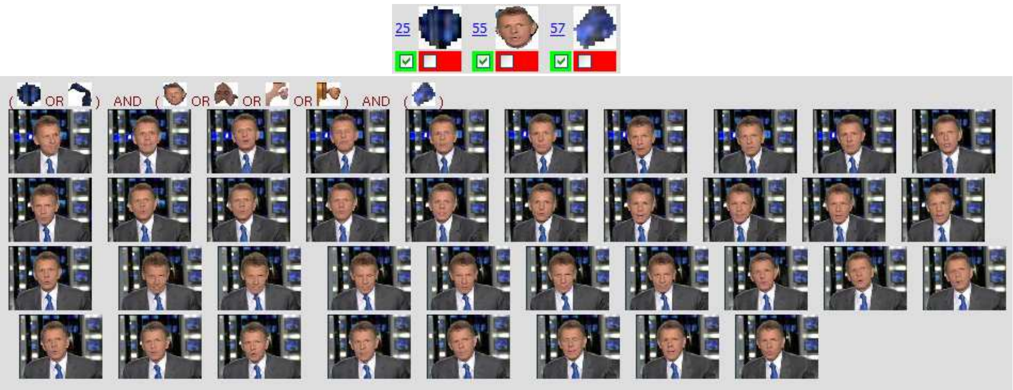

Fig. 17. Host frame retrieval scenario. Boolean query composition and corresponding results. Three categories were selected : two categories corresponding to dark blue mean color contain typical elements of the background, and one category corresponding to pinkish mean color contains faces.

in figure 18 show that two types of inlays are retrieved : a mugshot inlay and a pie diagram inlay. To focus on the mugshots alone the query is refined by adding a constraint of absence of red regions. Pie diagrams have been filtered out and the results for the new query contain only the mugshot inlay (figure 19).

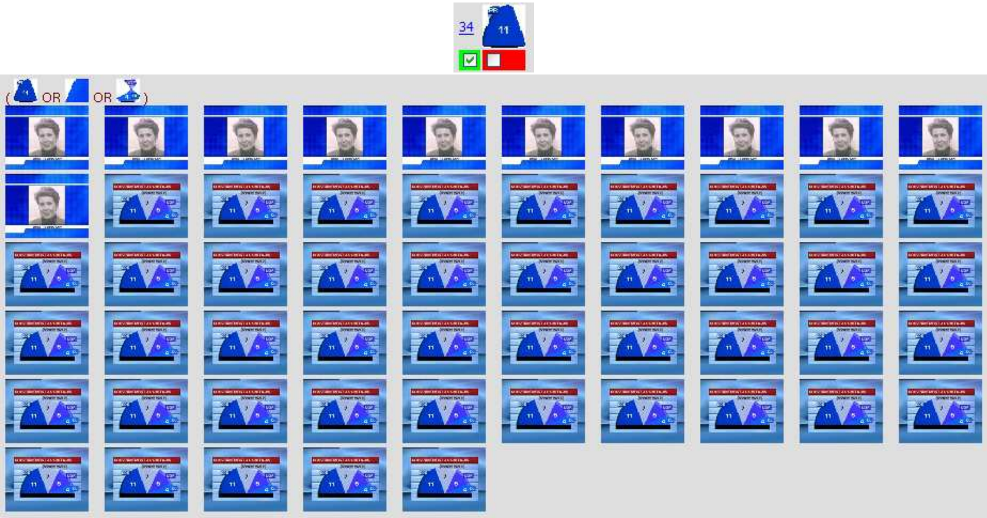

Fig. 18. Inlay retrieval scenario. Expression of boolean query and corresponding results. Query consists in a bright blue category (and its neighbors), which is characteristic of inlays.

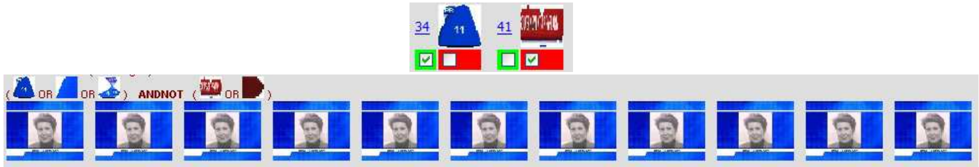

Fig. 19. Query refinement for inlay retrieval. To discard pie diagrams from previous results, query is refined to reject images with red regions.

5.3 Evaluation issues

Unlike query by example systems, the goal of our approach is to provide a relevant set of images which match the user's mental image. We are not in the context of precise search against a given example, but in the context of approximate search for database filtering. We discuss the four factors which participate to the successful retrieval of the user's mental image : the segmentation technique, the clustering algorithm, the boolean composition matching scheme and the query interface.

Segmentation : from the user perspective, false positives among matched regions are few and correspond to hard segmentation cases in complex composite natural images. In this case, a detected region may not be meaningful even if its mean color does correspond to a query category.

Clustering : the requirement for the clustering algorithm, CA in our case, is to produce homogeneous clusters of region descriptors which result in intuitive categories for the user. The clustering performance has a direct influence on the retrieval performance : if a selected query category contains incoherent regions, i.e. which are not similar to the majority of regions within the class, then the retrieved images which are composed of such incoherent regions will be considered as false positive. Thus, categories must be homogeneous; however they must not be too numerous, otherwise the resulting visual thesaurus may become “overloaded” and hard for the user to interact with. The CA algorithm proved to be a good choice since it produced homogeneous categories while keeping their number low.

Composition matching : given a user query (consisting of the selected PQCs and NQCs and γ), the boolean composition matching scheme relies on set operations (union, intersection and subtraction). The exact nature of these operations ensures that no false composition match can be done.

Query interface : the query interface is more sophisticated than query-by-example CBIR system interfaces, and its design plays an important role in the retrieval performance. It should allow the user to express a boolean composition query which successfully corresponds to his/her mental image. Region categories must be perceptually coherent so that the user can easily select relevant PQCs and NQCs for the mental image. The perceptual coherence of categories rely on the clustering performance and on the chosen region descriptor. The descriptor should be relevant with respect to the application. Region mean color was suitable for both photoagency and TV news scenarios but texture would be preferable for aerial images, for instance. The user is assisted in his selection of categories by the representative regions and, if necessary, the region browsing facility to check the overall content of category. The query refinement is a natural mechanism in our approach. Examination of the retrieved images helps the user decide what query categories in his/her selection are relevant given the mental image and the database content. It can also help the user adjust the range radius parameter to better match required search precision.

5.4 “Image Google”

Our framework can be viewed as an “Image Google” [2] from both the query and the matching aspects.

[2] Google : http://www.google.com

The image indexing by the category labels naturally mapped our problem to a text-retrieval one where documents are searched by their keywords. The inverted file technique was used to process boolean queries of region categories. The following analogy with the text retrieval terminology can be made :

image	→ document
region	→ word
region category	→ concept
neighbor category	→ synonym
union of neighbor categories	→ hyperonym
set of region categories	→ thesaurus
query by boolean composition	→ boolean query

However our query formulation is more general than Google. Since Google performs an exact keyword match the equivalent would be to perform an exact color match. But the categories selected by the user in our approach define *visual concepts* which are more general than an exact color.

From the user point of view the analogy is also remarkable. A Google query can be considered as a "query by mental document" rather than a "query by example document". To reach the "mental document" [3], the user expresses a boolean query from keywords. Note that Sivic and Zisserman recently proposed a "Video Google" approach [45] for object matching in videos. Their matching technique is inspired from text retrieval, but the query paradigm remains a partial query by example.

By letting the user have a direct access to the regions in the database and formulate boolean queries, the thesaurus interface provides *rich user expression*. Furthermore our approach requires *simple user interaction* since the user simply ticks the relevant categories from the thesaurus, without having to search a prior example or draw a sketch. We think the **rich user expression from simple interaction** is a challenging aspect in future work in visual information retrieval.

5.5 Positioning with respect to existing techniques and notions

In this section we review other work which employ techniques or notions related to some parts of our framework.

[3] The idea of the "mental document" may be more or less precise in the user's mind : it may be an already seen document or more generally a document related to a particular topic.

“Visual thesaurus” : some approaches introduced an idea of *Visual Thesaurus* of image blocks or regions [40,29,52]. Although different from one another, they all rely on a *supervised* learning process of visual features, either to represent user-driven visual groupings [40], or to learn domain-dependent visual similarity [29], or to learn visual description of predefined semantic classes [52]. In all these approaches, the thesaurus generation requires a significant user interaction in the supervised learning process. On contrary, the construction of our region photometric thesaurus is totally unsupervised. Besides none of these approaches supports boolean composition query.

Inverted files : inverted files have already been used in the CBIR context. They have used as an underlying indexing structure on complex visual descriptors to perform query by example (the example being an image in [49][12] or an image part in [45]). In the Viper system [49] inverted tables are built on the 80,000 visual attributes (both global and local) which are used to index the images. In [12] inverted tables are built on “codewords”, obtained after quantization of the region descriptors. In the Video Google approach [45], inverted tables are built on the “visual words”, obtained after quantization of the local descriptors. In all three cases inverted files are not exploited for mental image search or boolean search.

“Visual keywords” : several notions have been introduced in the literature to refer to the same idea of quantization/grouping of local visual descriptors : “codewords” [12], “visual words” [45], “picture words” [18], “visual keywords” [27]. But these approaches keep the “visual keywords” at the indexing and matching level and do not let the user explicitly use them to express a boolean query as we do. In our case we used the term “visual concepts” to refer to the region categories which also rely on local (segmented region, here) descriptor clustering. The specificity of our visual concepts is that they can be more or less precise depending on the user selected range radius.

Query by Sketch : as mentioned in section 1, query by sketch is another interaction mode which allows the user to express his/her mental image by drawing a sketch of it. It is particularly suitable to search for specific shapes [7] or specific spatial layouts such as in Geographic Information Systems [11]. In a general CBIR context, query by sketch also allows the user to specify the photometric appearance of the sketched regions : color selection is usually performed through a synthetic colorpicker such as in VisualSeek [48], QBic [38] or MARS [22] and texture from a texture example in [22]. The query interaction in our approach and query by sketch are complimentary : our thesaurus enables intuitive selection of region photometric features which are relevant for the mental image, while the latter allows specification of shape and spatial layout in the query. The addition of a query by sketch mode in our framework will be proposed in section 6.

6 Perspectives and concluding remarks

The original approach we presented offers rich perspectives for future work. They include :

- **Association with text ontologies :** in addition to visual descriptors, if regions are annotated with keywords, we can form semantic categories. Since our visual indexing scheme is very similar to keyword annotation, combining both visual and semantic categories with a text ontology would be straightforward. Boolean composition queries could then be performed on both visual and semantic content.
- **Other visual descriptors :** instead of mean color, any other region photometric descriptor with adequate similarity metric can be used to form region categories, such as color distribution [15] or texture and also geometric ones (such as position and area). The choice for other descriptors can be motivated by the domain of application (e.g. dedicated texture or shape descriptors for medical applications). The requirement for the descriptor is that the corresponding visual thesaurus is meaningful to the user. The approach proposed in [29] maybe investigated to integrate texture in our thesaurus.
- **Hierarchical categorization :** hierarchical categorization for the category generation process may become necessary for two problems : 1) to deal with very large image databases or 2) to integrate multiple descriptors (visual ones as well as keywords). The derived thesaurus and indexing scheme would both become hierarchical.
- **Query interface :** to make the query interface more intuitive, the following aspects should be investigated : 1) a more perceptual way to select γ radius instead of the box, 2) other strategies to define the category region representatives which are shown to the user.
- **Spatial relations :** so far we have only allowed the user to perform query by specifiying the presence and the absence of region types within images. Spatial relations between regions is another kind of information which can be relevant in some case. As discussed before, the combination of the region photometric thesaurus with a sketch functionality in the query interface can be easily achieved. On the indexing side, the challenge will be to incorporate spatial relations in the inverted tables.
- **Advanced information retrieval mechanisms :** the text-retrieval analogy mentioned in section 5.4 motivates the use of proven text-retrieval techniques to improve the user's satisfaction, such as relevance feedback on image composition, weighting scheme (for example in a similar way as Wang applied *tf.idf* to region-based image retrieval [53]) and result ranking to provide an order of relevance in the retrieved images.

We presented a new approach to implement the mental image search paradigm : the boolean composition of region categories. Since no starting image example is required to start the search, it provides a solution to the page zero problem. It can be used as a standalone approximate search engine as well as a prefiltering process to provide relevant starting images for existing query-by-example or relevance feedback techniques. The user formulates boolean queries through the region photometric thesaurus to specify the types of regions which are present and absent in his/her mental image. The very simple user interaction enables sophisticated boolean composition queries combined with the range query mechanism to adjust the precision of the visual search. Boolean composition matching is performed by boolean search with inverted files. This approach maps the CBIR problem into the information retrieval context on both query and indexing aspects and is viewed as an "Image Google".

We showed the viability of this framework with two applications. We first presented a search scenario in a photostock database using generic composition query. In the second scenario we showed that the domain specific knowledge, in a TV news context, could be taken into account to formulate the composition query. We observed in both cases that the user composition helps the implicit selection of relevant regions, in other words, that visual semantics emerges from the boolean visual composition expressed by the user.
Various future work directions were proposed for this effective, new and simple framework; we think these directions are challenging in the multimedia information retrieval context.

Acknowledgment

We would like to thank Aicha El Golli for generating the Kohonen maps of the visual thesaurus. We would also like to thank TF1 Channel for providing TV news videos.

References

[1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval.* Addison-Wesley, 1999.

[2] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Functions.* Plenum, New York NY, 1981.

[3] 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), Rocquencourt, France, pages 25–28, 2001.

[4] C. Carson and al. Blobworld: A system for region-based image indexing and retrieval. *Proc. of International Conference on Visual Information System, LNCS vol. 1614*, pages 509–517, 1999.

[5] M. La Cascia, S. Sethi, and S. Sclaroff. Combining textual and visual cues for content-based image retrieval on the world wide web. *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL)*, june 1998.

[6] I. J. Cox, M. L. Miller, and T. P. Minka. The bayesian image retrieval system, pichunter: Theory, implementation and psychological experiments. *IEEE Transactions on Image Processing*, 9(1):20–37, 2000.

[7] A. DelBimbo and P. Pala. Visual image retrieval by elastic matching of user sketches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2), february 1997.

[8] A. DelBimbo and E. Vicario. Using weighted spatial relationships in retrieval by visual contents. *IEEE workshop on Image and Video Libraries*, June 1998.

[9] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society*, 39:1–38, 1977.

[10] Y. Deng and B. S. Manjunath. An efficient low-dimensional color indexing scheme for region based image retrieval. *Proc. IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP), Phoenix, Arizona*, March 1999.

[11] M. J. Egenhofer. Query processing in spatial query by sketch. *Journal of Visual Languages and Computing (JVLC)*, 8(4):403–424, 1997.

[12] H.J. Zhang F. Jing, M. Li and B. Zhang. An effective region-based image retrieval framework. *Proceeding of ACM Multimedia*, pages 456–465, 2002.

[13] J. Fauqueur. Contributions to image retrieval by their visual components. *PhD Thesis, UVSQ - INRIA, (in french)*, 2003. http://www-rocq.inria.fr/˜fauqueur/recherche/.

[14] J. Fauqueur and N. Boujemaa. Logical query composition from local visual feature thesaurus. *International Workshop on Content-Based Multimedia Indexing (CBMI), Rennes, France*, 2003.

[15] J. Fauqueur and N. Boujemaa. Region-based image retrieval: Fast coarse segmentation and fine color description. *Journal of Visual Languages and Computing (JVLC), special issue on Visual Information Systems*, 15(1):69–95, 2004.

[16] M. Flickner and al. Query by image and video content: the qbic system. *IEEE Computer*, 28(9):23–32, 1995.

[17] H. Frigui and R. Krishnapuram. Clustering by competitive agglomeration. *Pattern Recognition*, 30(7):1109–1119, 1997.

[18] C. Y. Fung and K. J. Loe. Learning primitive and scene semantics for image for classification and retrieval. *ACM Multimedia*, 1999.

[19] V. Gouet and N. Boujemaa. Object-based queries using color points of interest. *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL)*, 2001.

[20] A. Gupta and al. The virage image search engine: an open framework for image management. *SPIE Storage and Retrieval for Image and Video Databases*, 2670, 1996.

[21] A. Hiroike, Y. Musha, A. Sugimoto, and Y. Mori. Visualization of information spaces to retrieve and browse image data. *International Conference on Visual Information System (VIS)*, 1999.

[22] T. Huang, S. Mehrotra, and K. Ramchandran. Multimedia analysis and retrieval system (mars) project. *Proceedings of the 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval*, 1996.

[23] D.J. Harper J.M. Jose, J. Furner. Spatial querying for image retrieval: a user-oriented evaluation. *international ACM SIGIR conference*, pages 232 – 240, 1998.

[24] T. Kohonen. *Self-Organizing Maps*. Springer Verlag, New York, 1997.

[25] J. Laaksonen, E. Oja, M. Koskela, and S. Brandt. Analyzing low-level visual features using content-based image retrieval. *International Conference on Neural Information Processing (ICONIP). Taejon, Korea*, 2000.

[26] B. LeSaux and N. Boujemaa. Unsupervised robust clustering for image database categorization. *IAPR International Conference on Pattern Recognition (ICPR)*, 2002.

[27] J.H. Lim. Learnable visual keywords for image classification. *ACM conference on Digital libraries*, pages 139–145, 1999.

[28] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, COM-28:84–95, 1980.

[29] 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.

[30] W. Y. Ma and B. S. Manjunath. Netra: A toolbox for navigating large image databases. *Multimedia Systems*, 7(3):184–198, 1999.

[31] S. MacDonald and John Tait. Search strategies in content-based image retrieval. *international ACM SIGIR conference*, 2003.

[32] J. MacQueen. Some methods for classification and analysis of multivariate observations. *Proc. of the Fifth Berkeley Symp. on Math. Stat. and Prob.*, 1:281–296, 1967.

[33] J. Malki, N. Boujemaa, C. Nastar, and A. Winter. Region queries without segmentation for image retrieval by content. In *Proc. of International Conference on Visual Information System (VIS)*, pages 115–122, 1999.

[34] B.S. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG-7: Multimedia Content Description Interface*. Wiley, ISBN: 0-471-48678-7, 2002.

[35] C. Meilhac and C. Nastar. Relevance feedback and category search in image databases. *IEEE International Conference on Multimedia Computing and Systems*, 1999.

[36] B. Moghaddam, H. Biermann, and D. Margaritis. Defining image content with multiple regions of interest. *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL)*, 1999.

[37] C. Nastar, M. Mitschke, C. Meilhac, and N. Boujemaa. Surfimage: A flexible content-based image retrieval system. *ACM Multimedia Conference Proceedings, Bristol, UK*, 1998.

[38] W. Niblack, R. Barber, W. Equitz, M. Flickner, and al. The qbic project: querying images by content using color, texture, and shape. *Proc. SPIE (Storage and Retrieval for Image and Video Databases)*, 1908:173–187, 1993.

[39] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *SPIE Storage and Retrieval for Image and Video Databases*, II(2185), Feb. 1994.

[40] R. W. Picard. Toward a visual thesaurus. *MIT Technical Report TR358*, 1995.

[41] J. Rissanen. Modeling by shortest data description. *Automatica*, 1978.

[42] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Does organisation by similarity assist image browsing? *international ACM SIGCHI conference*, pages 190–197, 2001.

[43] Y. Rubner. Perceptual metrics for image database navigation. *PhD Thesis, Stanford University*, 1999.

[44] S. Sclaroff, L. Taycher, and M. La Cascia. Imagerover: A content-based image browser for the world wide web. *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL)*, june 1997.

[45] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. *Proceedings International Conference on Computer Vision (ICCV)*, pages 1470–1477, 2003.

[46] A Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(12):1349–1380, 2000.

[47] J. R. Smith and S. F. Chang. Tools and techniques for color image retrieval. *IST/SPIE Proceedings*, pages 426–437, 1996.

[48] J. R. Smith and S. F. Chang. Visualseek: A fully automated content-based image query system. *ACM Multimedia Conference, Boston, MA, USA*, pages 87–98, 1996.

[49] D. Squire, W. Muller, H. Muller, and J. Raki. Content-based query of image databases, inspirations from text retrieval: inverted files, frequency-based weights and relevance feedback. *11th Scandinavian Conference on Image Analysis (SCIA) Kangerlussuaq, Greenland*, 1999.

[50] M. Swain and D. Ballard. Color indexing. *International Journal of Computer Vision (IJCV)*, 7(1):11–32, 1991.

[51] I. Keller T. Meiers, T. Sikora. Hierarchical image database browsing environment with embedded relevance feedback. *IEEE International Conference on Image Processing (ICIP)*, 2002.

[52] C. Town and D. Sinclair. Content based image retrieval using semantic visual categories. *ATT Technical Report*, 2001.

[53] J. Z. Wang and Y. Du. Rf*ipf: A weighting scheme for multimedia information retrieval. *IEEE International Conference on Image Analysis and Processing (ICIAP)*, 2001.

[54] I. H. Witten, A. Moffat, and T. C. Bell. *Managing gigabytes: compressing and indexing documents and images*. Van Nostrand Reinhold, 115 Fifth Avenue, New York, NY 10003, USA, 1994.

[55] T. Huang Y. Rui and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. *IEEE International Conference on Image Processing (ICIP)*, 1997.