Thematico-Visual Image Retrieval: How to Deal With Partially Indexed Corpora

Gérald Duffing

Thematico-Visual Image Retrieval: How to Deal With Partially Indexed Corpora

Gérald Duffing

UMR LORIA, BP 239, F-54506 Vandœuvre-les-Nancy, France.

ABSTRACT

It becomes very easy to access large amounts of images when surfing the Internet. All images, however, are not always thematically indexed. We think that partially thematically indexed corpora can be organised in a way that facilitates retrieval. We assume that, concerning visual properties, the corpus is totally indexed by means of "generic" features. Based on these indexes, a hierarchical clustering technique is used to bring together images that share some similarities: two distinct structures are built ("dendrograms").

We propose a new retrieval strategy based on a *virtual image* that captures the user's need along the retrieval session, taking into account both thematic and visual aspects. Clusters are successively selected in each dendrogram. A combined method, called *tunnels*, allows dendrograms cooperation. Images are then ranked according to the virtual image. After each retrieval step, the *virtual image* is enriched within a relevance feedback process. Theme, colour and general layout of each images can be rated and the query is updated accordingly.

In our experiments, we used two different corpora (2470 and 1100 images) to assess the performance of our thematico-visual approach within different indexing conditions. Experimentation results confirm the relevance of our approach and suggests improvement possibilities.

Keywords: Image Retrieval; Image Analysis; Clustering; Partially-Indexed Corpora

1. INTRODUCTION

It becomes very easy to access large amounts of images over the Internet. When dealing with image corpora's building, it remains however difficult to examine each picture to label it, even with a few indexing terms. In this paper, we argue that *partially indexed* corpora can be organised in a way that facilitates retrieval. For that purpose, we propose an adapted retrieval process.

Some systems allow only *keyword querying*, and will only consider the indexed part of the corpus. The indexing structure varies between classical keyword vectors (RIVAGE,[9] Cabri-n[21]), to complex data structures (VIMSYS,[8] MMIS,[7] MULTOS[18]). The difficulty of the indexing task varies accordingly. *Content-based retrieval* systems allow retrieval based on visual similarity comparison: each image has been characterised off-line with a set of useful visual features (for example, TradeMark,[25] Art-Museum,[12] VisualSEEk,[22] SurfImage[13]). Hybrid systems try to integrate both approaches: in CHABOT,[15] keywords are associated with a combination of visual predicates; in QBIC,[14] the similarity measure relies on keyword, shape, colour and texture...

Our proposal falls in the latter category. We believe that content-based systems allowing visual retrieval give little attention to the thematic (semantic) part of the query. To allow correct thematic retrieval, a complete indexing process would have to be carried out. However, even if the corpus is not totally indexed, we think that keywords remain the best mediating object between users' desires and image content. To bridge the gap between text-based and visual-based queries without loss of high semantic search capabilities, we propose a new retrieval strategy based on a prior corpus organisation that takes into account both thematic and visual aspects.

In this paper, we first examine how a corpus may be characterised and organised (§ 2). Then we propose our retrieval process approach and show how to take advantage of this organisation (§ 3). Finally, current experimentations and results are presented and discussed (§ 4).

E-mail: duffing@loria.fr

2. VISUAL AND THEMATIC CORPUS ORGANISATION

In our approach, images are characterised on a thematic *and* visual point of view. Our first goal is to select some relevant visual features, that can highlight image similarities. Then, we address the thematic content of images. In both cases, these features are used thereafter for classification purposes.

2.1. Describing image's visual content

Each image can be characterised by a set of visual features: colour, texture or shape are widely used.[17] However, some features are particularly well adapted to a given domain, whereas they achieve poor results in another one. Moreover, these features must be close to human perception, in order to take easily into account user judgements.[16]

Each corpus should be able to choose well adapted visual features to describe images. It is the retrieval system's responsibility to cope with this collection of different features. For our experiments, we have not assumed that domain-dependent knowledge could help selecting relevant features. As a consequence, we use "generic" features, such as basic colour, texture and shape. They are computed over the entire image (e.g. a colour histogram), and also on small image areas: a fixed grid is applied on the image, defining 32x32 pixels squares called *tiles*. Each tile is thus characterised by colour and texture features. This improves localisation and allows "layout" comparison.

Texture is described with a subset of Haralick's features.[10] Among fourteen features at four different orientations, we selected only four ones, namely "angular second moment", "contrast", "correlation", and "entropy", and used the average value for all orientations. According to our experiments, those indices are powerful enough to provide a rough characterisation of images.

Colour is described with a very general feature: the colour histogram.[23] Colours are represented in L*u*v* colour space, which is a device-independent and perceptually uniform colour model. A subset of 128 representative colours have been selected from the entire corpus and each image has been quantized according to this reduced colour map.

Shape description is more difficult, as extraction of the "relevant" shape is not straightforward.[1] We defined shape feature as spatial organisation of visually homogeneous areas (according to colour and texture features), which is based on the local characterisation described above.

2.2. Describing image's thematic content

The indexing process may be long and difficult. In an Internet context, it is important that users and indexers share the same vocabulary. This would make image retrieval easier, as any image corpora over the net can be searched with the same words. The vocabulary must be generic enough, so that it can be used in any application domain.

For these reasons, we use WordNet[6] as a generic organised set of words. In this lexical reference system, words are organised into synonym sets called "synsets". Each synset represents a lexical concept and is connected to other synsets with different kinds of semantic relationships. We use the following ones only: specific, generic, meronym, and holonym.

When image corpora are built, it is likely that not all images are indexed, and that indexed images are only assigned a few indexing words, say four or five. In this work, we suppose that the collection of images is *partially* and *not finely* indexed.

2.3. Organising the corpus: a classification approach

During image retrieval sessions, users try to find "similar" images, that is, images that share something in common. We make the following assumptions.

- **Similarity is twofold**. We consider both visual and thematic similarity. When dealing with images, the pictorial dimension is of crucial importance. Nonetheless, the thematical content plays also a central role, as visual similarity is often not powerful enough to allow pertinent retrieval. As far as we know, it remains very difficult to achieve an image interpretation (*i.e.* discovering its thematic content) based only on visual features.
- **Different degrees of similarity exist**. Compared to some pattern, retrieved images can be ranked according to their similarity with this model. It is of particular importance over the Internet, as, very often, only first retrieved items are considered. Moreover, the degree of similarity allows to control the amount of retrieved images: demanding strong similarity may return few images, whereas more images could be retrieved relaxing the similarity constraint.

Our goal is to organise the corpus in a way that our assumptions are verified. We have focused on clustering techniques,[4] which aim at grouping similar objects. They have been extensively used and studied in the field of Information Retrieval,[11,24] mainly for textual documents.

We applied clustering techniques also on *image* classification, based on one visual feature, namely colour, as no general conclusion has been made concerning clustering techniques suitability for information retrieval.[5,26] As we can't rely on domain-dependent knowledge, the classification will not be optimal. However, the following benefits are expected:

- Clustering speeds up retrieval process, as queries are matched against clusters' representatives, and as only images contained in the selected clusters will be further analysed. This is particularly true for visual similarity comparisons, which are often time-consuming.
- Visually similar images are clustered, and provided one image in the cluster is also thematically indexed, then every image in the same cluster gets a chance to be retrieved and displayed to the user (see our retrieval process below).

An agglomerative hierarchical clustering classifies the corpus, and two structures are obtained: a thematic **and** a visual classification are constructed based on features described in sections 2.2 and 2.1 above; the resulting structure, called "dendrogram", is reproduced on figure 1.

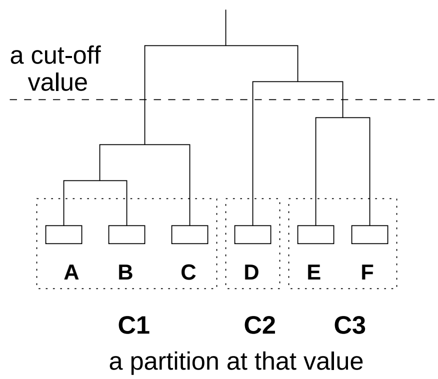


Figure 1. The dendrogram: a hierarchic structure produced by Agglomerative Hierarchic Clustering.

Dendrograms verify our assumptions: both thematic and visual similarities are represented, and different degrees of similarity can be considered by means of a cut-off value. At lower levels, clusters contain few, very similar images. The cluster size grows as we consider higher levels. This suggests a simple way to balance precision against recall*.

2.4. Associating visual knowledge to keywords

Let's consider a very simple case: the user is looking for images representing seascape. He will probably submit the keyword "sea" as a query. All images indexed with keyword "sea" or any other semantically close keyword will be retrieved, leaving aside all non-indexed images. Let's now assume that we have some hint about how the "sea" may be represented, in terms of visual features. Then we have enough material to refine the search, looking for images that feature these visual indices. Of course, "sea" may be represented very differently depending on the image: blue stormy atlantic ocean, or green pacific water around islands... This suggests that multiple visual description shall be gathered for one single word. To achieve this, we propose the concept of **realisation**, derived from the concept of *schema*[3] and defined in this paper as *a set of visual features appearing simultaneously in an image or part of an image, and associated with a keyword.*

As a realisation refers to a keyword, we can say that the set of all the realisations associated to a given keyword or "concept" provides the system with knowledge about the visual properties of this concept. We imagine that, on the Internet, a repository contains all available realisations and can be queried by various retrieval systems. Associating a keyword with its possible realisations is a kind of indexing task that has to be carried out manually.

* *Precision* is defined as the proportion of the retrieved documents which are relevant, whereas *recall* is the proportion of relevant documents that are actually retrieved.

3. OUR THEMATICO-VISUAL RETRIEVAL PROCESS

It is well accepted that visual feature are not always sufficient to retrieve relevant images. To alleviate this lack of accuracy, an adapted retrieval process is defined that will take advantage of all information given by users (user feedback). Figure 2 illustrates the main steps of our approach.

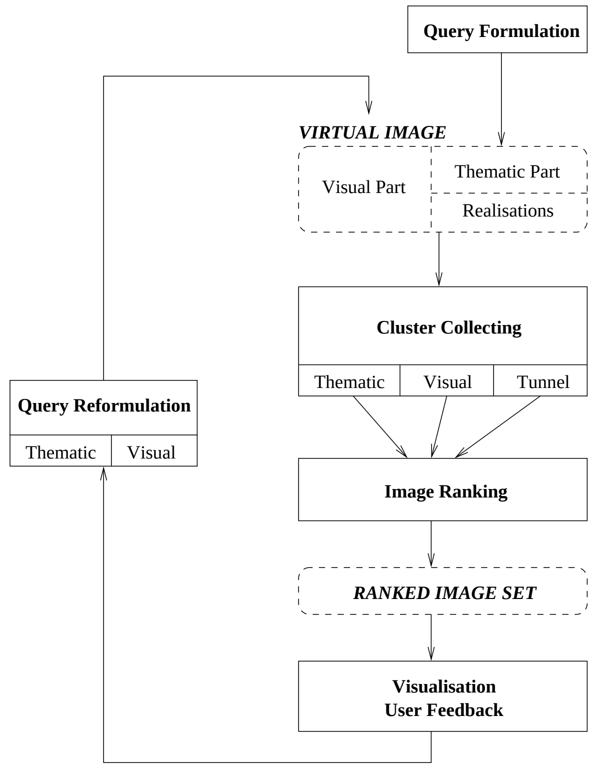


Figure 2. The Retrieval Process

3.1. A virtual image as a query

The system has to derive a representation of its understanding of user's needs. Each iteration will bring more information about the particular themes that images should represent, about the visual aspect they should have, thanks to user feedback mechanism.

To model this information, we propose the **virtual image** concept. Initial query, user feedback and features associated with displayed and judged images will permit *virtual image* construction. Four different parts are maintained during the retrieval session, corresponding to wanted and unwanted image features, for both thematic and visual aspects.

At the beginning of any session, the virtual image is empty. The user can construct a first query by adding to the virtual image some keywords. The vocabulary that can be used as query items is limited to nouns from WordNet. Each selected item can be constrained: "absolutely", "rather" or "possibly" present or absent in the image. This first query is mostly thematic, because no indications have been given about what the images to retrieve should look like. However, as realisations may be associated to keywords, we allow the user to browse through these visual examples. If he selects one or more realisations, visual conditions are added to the query. The visual part of the virtual image will be automatically updated during user feedback processing.

3.2. Cluster collecting and image ranking

The retrieval process consists in two main steps: clusters collecting and image ranking. Clusters collecting searches the thematic or visual dendrogram using an ascendant method (adapted from a method proposed by Van Rijsbergen[19]). We first localise interesting low-level clusters, by comparing their centroids with the corresponding part of the virtual image. Then, we perform a constrained generalisation of these low-level clusters, using constraints that control the minimum/maximum size of the cluster, its dispersion rate, and the maximum distance between centroid and query. This method is applied to both thematic and visual dendrograms. Images from selected clusters are ranked according to their distance to the virtual image. To refine cluster collecting, thematically indexed images are further used within our *tunnel* mechanism, in order to collect more clusters from the visual dendrogram. Three steps can be identified:

–**1**– **Thematic clusters collecting**. — The thematic hierarchy is searched using the thematic part of the virtual image (cosine measure[20] is evaluated between "positive" part of the virtual image and the cluster centroid). Depending on cluster size, though, not all images may be equally relevant to the virtual image: this motivates additional thematic ranking process that will be used for the *tunnel* mechanism (see below). Let us suppose that cluster $C1$ is selected, in which image A is the best ranked.

–**2**– **Visual clusters collecting – direct search**. — The visual hierarchy is searched using the visual (colours) part of the virtual image. The "positive" part of the virtual image and the cluster centroid are matched using histogram intersection method.[23] This results in some likely interesting clusters, in which images are ranked according to the visual part (and the thematic part, whenever the image is thematically indexed) of the virtual image. This pure visual search may discover new interesting images. Recall is favoured against precision, though.

–**3**– **Visual clusters collecting – tunnel search**. — We believe that links can be established between the thematic dendrogram and the visual one, yielding to a thematico-visual synergy. This is based on the following hypothesis: "Thematic similarity and visual similarity are not independent". In step 1, we assumed that image A from thematic cluster $C1$ is particularly relevant. The tunnel hypothesis tells us to consider visually similar images, that is to locate image A in the visual dendrogram, and to consider that some "neighbourhood" around A as relevant ("neighbourhood" is controled by the constrained generalisation process). Figure 3 illustrates this thematico-visual cooperation: two thematically relevant images are chosen to explore the visual dendrogram, in which two clusters are chosen. The advantage of tunnel search is that **non-indexed images have a chance to be retrieved**, based on indirect thematic criteria.

Image ranking uses both positive and negative parts of the virtual image *i.e.,* what is wanted and unwanted by the user. Negative feedback is carefully handled for thematic and visual aspects, but we give more importance to positive feedback. The following formula is used to rank images:

$$S = w * S_t + (1 - w) * S_v.$$

where the visual score (S_v) and thematic score (S_t) may be weighted by the user (w factor) to give more or less importance to visual search against thematic search. Scores are computed as follows:

- **Thematic score** mainly consists in a cosine measure introduced in the vector model.[20] Moreover, we use *realisations* to alter this thematic score. Whenever realisations have been added to the query, each retrieved image is further analysed with the visual features they contain. We try to determine in what extent the realisations' visual elements are present in the image. If the image presents some similarity with a *realisation*, then we slightly raise the thematic score. Image analysis by means of *realisations* is very incertain and lead us to use them as "bonus" that may, however, alter image ranking.
- **Visual score** consists in global and local similarity evaluation. Global evaluation compares images' global colour histogram to the corresponding part of the virtual image. Furthermore, if a typical layout has been chosen by the user (see feedback issues, below), retrieval precision can be considerably improved by introducing a *local* visual matching. Images (that have been characterised locally, splitting image into 32x32 tiles) are then considered tile-by-tile, so that their overall composition are compared.

3.3. User feedback issues and reformulation

It is important to gather information as precise as possible from the user. Actually, we may have retrieved images for which one point of view only (theme or visual aspect) is relevant: the system needs to know in what extent a retrieved image is relevant or not to the user. This is the only way to extract accurate information that will allow proper query reformulation. To achieve this, we propose a threefold judgement on each image.

Theme is crucial, because our system is oriented towards thematic retrieval. This kind of judgement is used to determine what themes are to be searched or avoided. A relevant image may feature new interesting themes that should be included in the next virtual image, by means of reformulation. **Colour** has been chosen as the most representative among visual indices. Colour is a fundamental visual property of images, as users are very sensitive to it; it has also a great discriminant power.[23] **Layout** seems to be a good criterion candidate, as it captures the whole composition of the image, and then some "shape" properties. In this work, "layout" refers to spatial localisation of blobs, each blob featuring some homogeneous visual properties (in terms of colour and texture).

The user can choose, among the proposed images, the most representative one (if any). This "typical" layout will be used for image ranking purposes at the next retrieval step. This limitation to one image only is intended to facilitate the visual reformulation problem as well as the local visual similarity evaluation. For each judgement type, user can "accept", "reject" or have "no idea" about each selected image. This vote can be moderated with a weight ranging from 1 to 10.

Reformulating a virtual image involves complex tasks, especially because queries consist in both thematic and visual conditions. **Thematic reformulation** takes into account the previous virtual image, updating the keywords' weights according to user feedback (using the vector model). **Visual reformulation** is far more complex. Each type of feature has its own strategy to alter its component depending on user feedback. Colour feature is the simplest case to handle: wanted and unwanted colours are modelled as a vector admitting a positive or negative normalised value. This vector can be easily updated according to user feedback.

4. EXPERIMENTS

4.1. Corpus characteristics and indexing

Experiments aims at evaluating the impact of partial indexing on retrieval in our approach. Therefore, we can "hide" some part of the index. Indexing choices are carried out randomly, or by hand. *Random indexing* consists in choosing randomly a certain amount of images (say, 20% of the total amount of images), and to make indexing keywords "visible" to the system. *Ad hoc indexing* relies on manual selection of images for which indexing should be visible. It is worth noting that *ad hoc* indexing concerns far less images than random indexing: this is on purpose, since we want to prove that our approach can achieve good performance, provided few images — chosen by hand — are indexed.

We use two different corpora. Corpus #1 contains 2470 heterogeneous images (pictures, paintings, ...). Three cases have been tested: random indexing of 20% and 40% of images, and hand-indexing of 5% of images. Corpus #2 contains 1100 images (Buildings pictures and Fine Arts paintings), and 5% of them have been indexed by hand. Each image, of course, has been characterised off-line with visual features (colour, texture and "shape").

4.2. Experiment description

In this work, we want the assess performance of the thematico-visual approach. To do so, we let the user conduct an entire session on corpus #1 (i.e. consisting in as many iterations as he wants). Our goal is to determine whether the system allowed retrieval of interesting images, *i.e.* thematically and visually relevant images. As a retrieval session may return numerous images, and as each image is ranked according to its expected visual and thematic relevance, we say that a retrieval session is "successful" whenever a certain amount of best ranked images are thematically and visually relevant. We give particular importance to final ranking.

We assume that images are presented to the user by groups of n images. We consider only the k first groups for evaluation. Each of these $k \times n$ images is evaluated by the user from a visual and thematic point of view. Therefore, a symbolic grade is assigned to each image: A (very relevant), B (relevant), C (rather relevant), D (irrelevant), and E (totally irrelevant), according to its visual and thematic user-relevance. Our decision rules are as follows: an image visually and thematically relevant gets grade A; an image visually or thematically relevant only gets grade C; finally, an image that is not thematically nor visually relevant gets grade E (of course, other decision rules may be adopted, depending on a particular search context).

We define a quality measure, noted Q, to evaluate the retrieval session performance, based on relevance judgements of the user (as described above), and on ranking:

Query Number	Query Formulation	Total nbr of relevant images
1	People at the airport	6
2	Military aircrafts	116
3	Sea birds	33
4	Boats	55
5	Soldiers in the desert	11
6	A cliff and a river	13
7	Cities views	70
8	Mountain sports	21
9	Trains or locomotives	43
10	Old cars	41
11	Fruits sales at the market	11

Table 1. A set of queries for evaluation.

$$Q = N \sum_{b=1}^{k} \left((k - b + 1) \sum_{p=1}^{n} v_{b,p} \right) \tag{1}$$

With:

b	group number,
p	image position within a group,
$v_{b,p}$	relevance value associated with image at position p in group b,
N	normalisation factor.

The numerical value $v_{b,p}$ is derived from symbolic grades. In this experiment, we chose the following values: $A = +5$, $B = +3$, $C = 0$, $D = -1$ and $E = -2$. Constant N may be set to 1, so that Q is a "general quality measure", or to $1/nk$, to obtain an "average quality measure".

A set of 11 queries has been defined, in order to evaluate system performance on corpus 1. Table 1 lists these queries, along with a manual evaluation of the amount of potentially relevant images in the corpus. Given the set of queries described above, retrieval sessions have been conducted in different indexing conditions, namely 20%, 40% and *ad hoc* indexing. Only the 10 first images † are taken into account, yielding a global performance measure Q. As the user may tune visual against thematic weights to optimise the results, we also indicate the visual weight (in a range from 0 to 100). Table 2 shows the results corresponding to our 11 queries under different indexing conditions. The information provided in these tables corresponds to whole retrieval sessions (*i.e.* after possible feedback). It is organised as follows: for each query (col. 1), and under various indexing conditions (col. 2), relevance judgements are reported for the 10 best ranked images (col. 3 to 12): "T" stands for "thematically relevant", "V" for "Visually relevant". A session quality score Q is computed according to equation 1, with $k = 2$, $n = 5$, $N = 1$ (col. 13). Finally, the visual weight used to produce the reported results is given (col. 14). Table 3 provides a visual representation of our retrieval results for some queries.

4.3. Discussion

4.3.1. The impact of initial indexing

Initial thematic indexing is of particular importance. We obtained the following average scores according to the three indexing conditions we considered: 35 at random 20%, 42 at random 40%, and 51 at *ad hoc* indexing. Not surprisingly, *ad hoc* indexing achieved the best results, even if far less images are actually indexed in this case. Moreover, indexing more images does not systematically permits better results.

† We consider 2 groups of 5 images, that is $k = 2$ and $n = 5$ in equation 1.

Query Number	Index Cond.	Image Ranks										Quality Score	Visual Weight
		1	2	3	4	5	6	7	8	9	10		
1	20	TV	TV	TV	V	TV			TV	TV		44	70
1	40	TV	TV	TV	TV	TV						40	100
1	AH	TV	TV	TV	TV	TV		TV				47	70
2	20	TV	TV	T	TV	T	T			T		24	60
2	40	TV	TV	TV	TV	TV	TV	T				49	70
2	AH	TV	TV	TV	TV	TV	T	T	T		T	48	90
3	20	TV	TV	TV	V	V	TV	T			V	31	90
3	20	TV	TV	TV	TV	V	TV	V	V			41	70
3	40	TV	TV	TV	TV	V	V		T			34	100
3	40	TV	TV	TV	TV	V	V	V		T	T	38	30
3	AH	TV	TV	T	TV	TV	T	T	T			36	40
4	20	TV	V	T	V	TV	V		V	TV		16	60
4	40	TV	T	T	TV	TV	T				T	24	50
4	AH	TV	TV	TV	TV	T		T	T	T	T	38	80
5	20	TV	TV	TV	TV	TV	V		TV			49	100
5	40	TV	TV	TV	TV					T		32	50
5	AH	TV	TV	TV	TV	TV	TV		TV	T	TV	63	70
6	20	TV	TV	TV	TV				TV			37	100
6	40	V	TV	TV	TV	V	V	TV	TV	V	V	40	100
6	40	TV	TV	TV	TV	TV	TV	V	V	V	TV	60	100
6	AH	TV	TV	TV	TV	TV	TV	T	V	V	V	55	60
7	20	TV	TV	V	TV	TV	V	T				34	90
7	40	TV	TV	TV	TV	TV					TV	47	100
7	40	TV	TV	TV	TV	T		TV		T	TV	46	60
7	AH	TV	TV	TV	TV	TV	TV	T	V	TV	V	60	60
8	20	TV	TV	TV	TV							26	60
8	40	TV	TV	TV		TV						26	70
8	AH	TV	TV	TV	TV	TV	TV	TV			TV	61	100
9	20	TV	TV	TV								12	60
9	40	TV	TV	TV	TV	T	T				T	34	75
9	40	TV	TV	TV	TV	TV		T			T	44	70
9	AH	TV	TV	TV	TV	T	TV		TV	T		46	20
10	20	TV	TV		TV	TV	TV			TV	TV	47	95
10	40	TV	TV	TV	TV	TV	TV	TV	TV			61	60
10	AH	TV	TV	TV	TV	TV	T	T	T		T	48	70
11	20	TV	TV	TV	TV		TV	V	V		V	39	15
11	20	TV	TV	TV	TV	TV	TV	V	T	V	V	55	90
11	40	TV	TV	TV	TV	TV		TV	TV			54	50
11	AH	TV	TV	TV	TV	TV	TV	TV		T	V	58	100

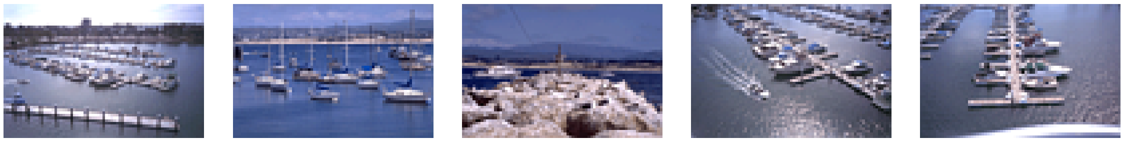
Table 2. Query results.

It is clear that retrieval techniques based on visual similarity evaluation only are not sufficient, since visual indices are not powerful enough to characterise accurately *all* kind of images. In this situation, part of the corpus may be "invisible" to the system. Similar situation occurs when images are not indexed, and therefore not accessible from a thematic point of view. The thematico-visual approach tries to avoid both of these situations.

4.3.2. Similarity evaluation reliability

In the thematico-visual approach, the *tunnel* mechanism has been devised to allow collaboration of thematic and visual retrieval techniques. Experiments show that, for each query, an interesting set of images has been retrieved, by means of thematic and/or visual similarity measures. However, these measures cannot always distinguish relevant

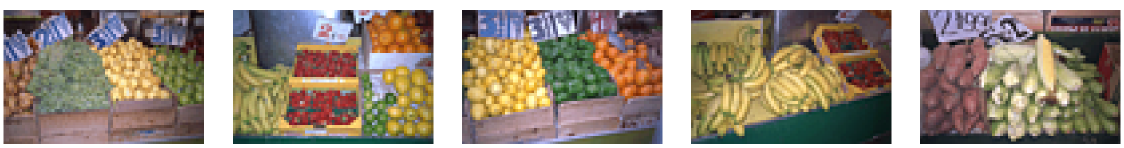
(a) Query #1 (*ad hoc* indexing)



(b) Query #4 (40% indexing)

(c) Query #9 (40% indexing)

(d) Query # 11 (20% indexing) without layout feedback



(e) Query # 11 (20% indexing) after layout feedback

Table 3. Results on corpus 1 for queries 1, 4, 9, and 11.

from irrelevant images: *noise* seems to be the price to pay to obtain satisfactory recall.

Figure 4 shows how thematic (T), visual (V) and tunnel (U) techniques performed during this experiment on corpus 1.

Results are first reported for all images (col. 2), then for relevant images only (col. 3, thematically *and* visually relevant). In both cases, the value reported is the average number of images retrieved using each technique. The last column is the ratio between the former results, therefore indicating the method reliability.

We observe that the *tunnel* mechanism is fairly reliable: 62.5% of tunnel-retrieved images turned out to be relevant. Only 30% of images retrieved from visual techniques only proved to be relevant.

The hypothesis associated to the *tunnel* mechanism turned out to be verified: we observe that most of the images retrieved by means of *tunnel* have been judged as relevant by the user. In this way, the *tunnel* can be used to *confirm*

Retrieval Technique	Average nbr of retrieved images	Avg. nbr of relevant retrieved images	Relevance ratio
(T)hematic	6.7	2.76	41.2%
(V)isual	13.0	3.97	30.5%
T(U)nnel	1.6	1.00	62.5%

Table 4. Retrieval technique performance.

image relevance. As a heuristic, we can say that an image retrieved by means of visual techniques *and* by means of *tunnel* is more likely to be relevant than another.

4.3.3. Query reformulation performance

We shall focus here on visual reformulation, that permitted retrieval results improvement. Global colour-based reformulation allows the system to focus on a certain kind of images, namely those which are closer to the user's desires. As a complement, layout-based reformulation allows the system to refine image ranking (see results on figs. 3(d) and 3(e)).

Visual retrieval favours recall against precision. As a consequence, "noise" is the main drawback of this approach. We think, however, that this is necessary, as the corpus is not totally indexed. Actually, relevant images may not be indexed, and visual retrieval — combined with the *tunnel* mechanism — is the only chance we have to retrieve those images. Figure 3 illustrates our results: whereas thematic descriptions allows the retrieval of two interesting images only (row 1), the *tunnel* mechanisms acts as a relay and allows retrieval of 6 more images, that are not actually indexed. The virtual image and the relevance feedback process are responsible for improving image ranking, and therefore to allow those non-indexed images to be presented.

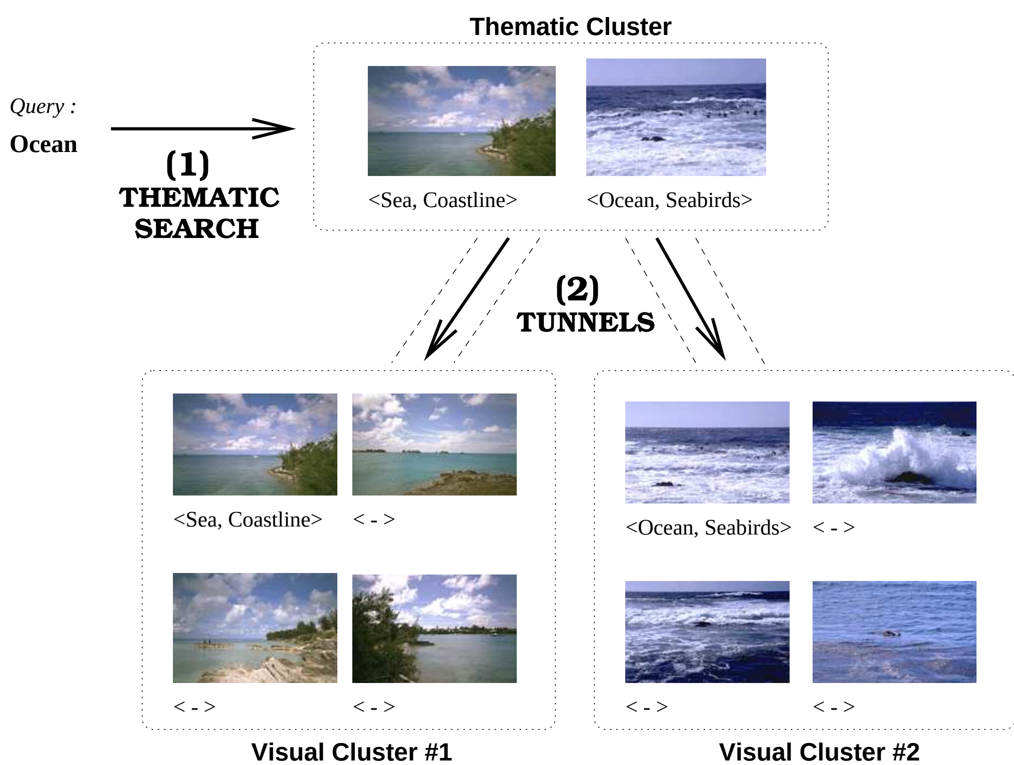


Figure 3. An example of thematico-visual retrieval results.

5. CONCLUSION AND FUTURE WORK

We have proposed a method that allows thematic and visual image retrieval in partially-indexed corpora. Our strategy relies on a preliminary corpus organisation into two hierarchical structures by a clustering process. User-system interactions are given a particular importance: the retrieval process relies on user judgments that provide precise feedback data. This information is gathered and processed within a *virtual image*.

This vision of image retrieval is well adapted to the Internet, where large partially thematically-indexed corpora can be found. We think that different corpora over the Internet could be accessed and searched using our system, provided some general knowledge is available: a collection of *realisations* that stores numerous visual examples used to refine retrieval, and a collection of *visual indices*, giving each retrieval engine the opportunity to exploit various image descriptions (downloading, for example, all necessary chunks of code to access visual indexes, or compare features). Figure 4 illustrates our proposition.

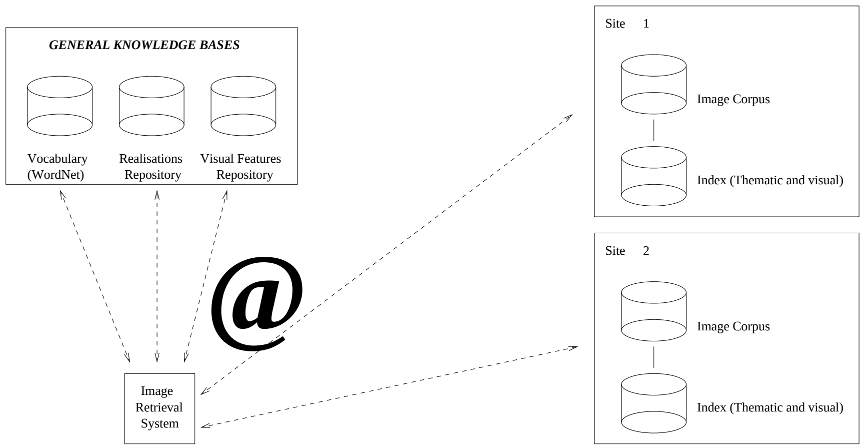


Figure 4. Distributed Image Retrieval.

At the end of some retrieval session, a virtual image contains all information (thematic and visual) corresponding to the user's need. This virtual image has been enriched by successive relevance feedback processing. Thus, we think that it can be used to perform further off-line searching over the Internet, without any user's help.

First experimentation results confirm the relevance of our approach. We believe that it could be improved by efficient integration of other visual characteristics which will, on one hand, make possible to refine not only the user query but also the similarity measure by selecting the best characteristics according to the user and his preferences.[16,2] On the other hand, enriching the visual feature collection could increase interoperability between different corpora over the Internet.

REFERENCES

1. P. Aigrain, H. Zhang, and D. Petkovic. Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review. *Multimedia Tools and Applications*, 3:179–202, 1996.
2. G. Ciocca and R. Schettini. A relevance feedback mechanism for content-based image retrieval. *Information Processing and Management*, 35:605–632, 1999.
3. B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman. The Schema System. *International Journal of Computer Vision*, 2:209–250, 1989.
4. R. Duda and P. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, 1973.
5. A. El-Hamdouchi and P. Willett. Techniques for the measurement of clustering tendency in document retrieval systems. *Journal of Information Science*, 13:361–365, 1987.
6. C. Fellbaum, editor. *WORDNET: An Electronic Lexical Database.* MIT Press, 1998.
7. C. Goble, M. O'Docherty, P. Crowther, M. Ireton, J. Oakley, and C. Xydeas. The Manchester Multimedia Information System. In *Lecture Notes in Computer Science, vol. 580*, pages 39–55. Springer, 1992.

8. A. Gupta, T. Weymouth, and R. Jain. Semantic queries with pictures: the VIMSYS model. In *Proceedings of the 17th int. conf. on Very Large Data Bases*, pages 69–79, Barcelona, septembre 1991.
9. G. Halin. Machine Learning and Vectorial Matching for an Image Retrieval Model: EXPRIM and the System RIVAGE. In J.-L. Vidick, editor, *ACM 13th Int. Conf. on Research and Development in Information Retrieval*, pages 99–114, Brussels (Belgium), septembre 1990. Presses Universitaires de Bruxelles.
10. R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans. on Systems, Man, and Cybernetics*, SMC–3(6):610–621, 1973.
11. N. Jardine and C.J. van Rijsbergen. The use of hierarchical clustering in information retrieval. *Information Storage and Retrieval*, 7:217–240, 1971.
12. T. Kato. Database architecture for content-based image retrieval. In *Image Storage and Retrieval Systems*, volume 1662, pages 112–123, San Jose, CA, 1992. SPIE.
13. C. Nastar, M. Mischke, C. Meilhac, N. Boudjemaa, H. Bernard, and M. Mautref. Retrieving images by content: the surfimage system. In *Multimedia Information Systems*, Istanbul, 1998.
14. W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The QBIC project: querying images by content using color, texture and shape. In Wayne Niblack, editor, *Storage and Retrieval for Image and Video Databases*, pages 173–181, San Jose, CA, 1993. SPIE.
15. V. E. Ogle and M. Stonebraker. CHABOT: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48, 1995.
16. R.W. Picard and T.P. Minka. Vision Texture for Annotation. *Multimedia Systems*, 3:3–14, 1995.
17. W.K. Pratt. *Digital Image Processing*. John Wiley & Sons, New York, second edition, 1991.
18. F. Rabitti and P. Savino. Querying semantic image database. In *Image Storage and Retrieval Systems*, volume 1662, pages 69–78, San Jose, CA, 1992. SPIE.
19. C.J. van Rijsbergen and W.B. Croft. Document clustering: an evaluation of some experiments with the Cranfield 1400 collection. *Information processing and management*, 11:171–182, 1974.
20. G. Salton, E. Fox, and H. Wu. Extended boolean information retrieval. *Communications of the ACM*, 26(12):1022–1036, 1983.
21. M. Smaïl. Case-Base Reasoning Meets Information Retrieval. In *RIAO 94: Intelligent Multimedia Information Retrieval Systems and Management*, page 133, 1994.
22. J. R. Smith and S.-F. Chang. Querying by color regions using the VisualSEEk content-based visual query system. In Mark T. Maybury, editor, *Intelligent Multimedia Information Retrieval*, pages 23–41. AAAI Press, Menlo Park, 1997.
23. M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.
24. E. Voorhees. *The Effectiveness and Efficiency of Agglomerative Hierarchic Clustering in Document Retrieval*. PhD thesis, Cornell University, Ithaca, NY, Etats-Unis, 1985. Rapport Technique TR 85-705.
25. T. Whalen, E.S. Lee, and F. Safayeni. The Retrieval of Images from Image Databases. *Behaviour & Information Technology*, 14(1):3–13, 1995.
26. P. Willett. Recent trends in Hierarchic Document Clustering. *Information processing and management*, 24(5):577–597, 1988.