Integrating Perceptual Signal Features within a Multi-facetted Conceptual Model for Automatic Image Retrieval

Mohammed Belkhatir, Philippe Mulhem, Yves Chiaramella

Integrating perceptual signal features within a bi-facetted conceptual model for automatic image retrieval

Mohammed Belkhatir, Philippe Mulhem, Yves Chiaramella

Laboratoire CLIPS-IMAG,
Université Joseph Fourier, Grenoble, France
{belkhatm, mulhem, chiara}@imag.fr

Abstract. The majority of the content-based image retrieval (CBIR) systems are restricted to the representation of signal aspects, e.g. color, texture… without explicitly considering the semantic content of images. According to these approaches a sun, for example, is represented by an orange or yellow circle, but not by the term "sun". The signal-oriented solutions are fully automatic, and thus easily usable on substantial amounts of data, but they do not fill the existing gap between the extracted low-level features and semantic descriptions. This obviously penalizes qualitative and quantitative performances in terms of recall and precision, and therefore users' satisfaction. Another class of methods, which were tested within the framework of the Fermi-GC project, consisted in modeling the content of images following a sharp process of human-assisted indexing. This approach, based on an elaborate model of representation (the conceptual graph formalism) provides satisfactory results during the retrieval phase but is not easily usable on large collections of images because of the necessary human intervention required for indexing. The contribution of this paper is twofold: in order to achieve more efficiency as far as user interaction is concerned, we propose to highlight a bond between these two classes of image retrieval systems and integrate signal and semantic features within a unified conceptual framework. Then, as opposed to state-of-the-art relevance feedback systems dealing with this integration, we propose a representation formalism supporting this integration which allows us to specify a rich query language combining both semantic and signal characterizations. We will validate our approach through quantitative (recall-precision curves) evaluations.

1 Introduction

From a user's standpoint, the democratization of digital image technology has led to the need to specify new image retrieval frameworks combining expressivity, performance and computational efficiency.

The first CBIR systems (signal-based) [11,16,18,19] propose a set of still images indexing methods based on low-level features such as colors, textures... The general approach consists in computing structures representing the image distribution such as color histograms, texture features and using this data to partition the image; thus reducing the search space during the image retrieval operation. These methods are based

on the computation of discriminating features rejecting images which do not correspond to the query image and hold the advantage of being fully automatic, thus are able to quickly process queries. However, aspects related to human perception are not taken into account. Indeed, an image cannot be sufficiently described by its moments or color histograms. The problem arising from invariants or discriminating features lies on the loss of semantic information conveyed by the image. These tools are used for restricting the search space during the retrieval operation but cannot however give a sound and complete interpretation of the content. For example, can we accept that our system considers red apples or Ferraris as being the same entities simply because they present similar color histograms? Definitely not, as shown in [9], taking into account aspects related to the image content is of prime importance for efficient photograph retrieval.

In order to overcome this weakness and allow the representation of the semantic richness of an image, semantic-based models such as Vimsys [5] and EMIR^2 [12,17] rely on the specification of a set of logical representations, which are multilevel abstractions of the physical image. The originality of these models is achieved through integration of heterogeneous representations within a unique structure, collecting a maximum of information related to the image content. However these systems present many disadvantages. First, they are not fully automatic and require the user intervention during indexing, which constitutes a major drawback when dealing with reasonable corpus of images as this process is time-consuming and leads to heterogeneous and subjective interpretations of the image semantic content. Moreover, these models do not incorporate a framework for signal characterization, e.g. a user is not able to query these systems for "red roses". Therefore, these solutions do not provide a satisfying solution to bridge the gap between semantics and low-level features.

State-of-the-art systems which attempt to deal with the signal/semantics integration [6,10,21] are based on the association of a query by example framework with textual annotation. These systems mostly rely on user feedback as they do not provide a formalism supporting the specification of a full textual querying framework combining semantics and signal descriptions and therefore exhibit poor performance in relating low-level features to high-level semantic concepts. Prototypes such as ImageRover [6] or iFind [10] present loosely-coupled solutions relying on textual annotations for characterizing semantics and a relevance feedback scheme that operates on low-level features. These approaches have two major drawbacks: first, they fail to exhibit a single framework unifying low-level and semantics, which penalizes the performances of the system in terms of retrieval effectiveness and quality. Then, as far as the querying process is concerned, the user is to query both textually in order to express high-level concepts and through several and time-consuming relevance feedback loops to complement her/his initial query. This solution for integrating semantics and low-level features, relying on a cumbersome querying process does not enforce facilitated and efficient user interaction. For instance, queries involving textually both semantics and signal features such as "Retrieve images with a purple flower" or "Retrieve images with a vegetation which is mostly green" cannot be processed. We propose a unified framework coupling semantics and signal features for automatic image retrieval that enforces expressivity, performance and computational efficiency. As opposed to state-of-the-art frameworks offering a loosely-coupled solution with a textual framework for

keyword-based querying integrated in a relevance feedback framework operating on low-level features, user interaction is optimized through the specification of a unified textual querying framework that allows to query over both signal and semantics.

In the remainder of this paper, we will first present the general organization of our model and the representation formalism allowing the integration of semantics and signal features within an expressive and multi-facetted conceptual framework. We will deal in sections 3 and 4 with the descriptions of both the semantic and the signal facets, dealing thoroughly with conceptual index structures. Section 5 will specify the querying framework. We finally present the validation experiments conducted on a test collection of 2500 personal photographs.

2 The proposed method: Signal/Semantic integration within an expressive conceptual framework

In state-of-the-art CBIR systems, images cannot be easily or efficiently retrieved due to the lack of a comprehensive image model that captures the structured abstractions, the signal information conveyed and the semantic richness of images. To remedy such shortcomings, visual semantics and signal features are integrated within an image model which consists of a physical image level and a conceptual level. The latter is itself a multi-facetted framework supported by an expressive knowledge representation formalism: conceptual graphs.

2.1 An image model integrating signal and semantic features

The first layer of the image model (fig.1) is the physical image level representing an image as a matrix of pixels.
The second layer is the conceptual level and is itself a tri-facetted structure:

- The first facet called *object facet* describes an image as a set of **image objects**, abstract structures representing visual entities within an image. Their specification is an attempt to operate image indexing and retrieval operations beyond simple low-level processes [18] or region-based techniques [2] since image objects convey the visual semantics and the signal information at the conceptual level. Formally, this facet is described by the set I_{IO} of image object identifiers.
- The second facet called ***visual semantic facet*** describes the image semantic content and is based on labeling image objects with a semantic concept. In the example image of fig. 1, the first image object (Io1) is tagged by the semantic concept *Hut.* Its formal description will be dealt with in section 3.
- It is itself partitioned in two sub-facets. The ***color facet*** describes the image color content in terms of symbolic perceptual features and consists in characterizing image objects with color concepts. In the example image of fig. 1, the second image object (Io2) is associated with symbolic colors *Cyan* and *White* and symbolic textures . The signal facet will be described in detail and formalized in section 4. It features

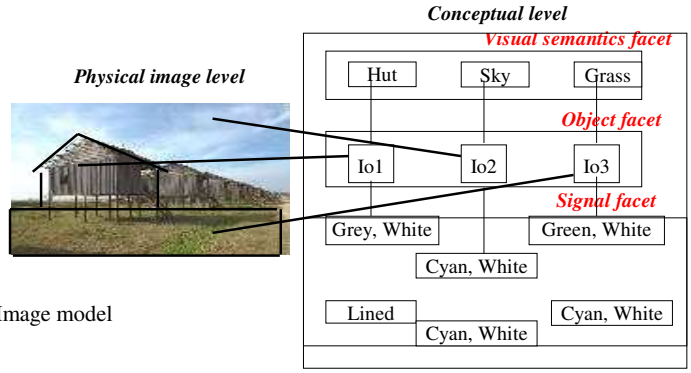


Fig. 1. Image model

2.2 Representation formalism

In order to instantiate this model within a framework for image retrieval, we need a representation formalism capable to represent image objects as well as the visual semantics and signal information they convey. Moreover, this representation formalism should make it easy to visualize the information related to an image. It should therefore combine expressivity and a user-friendly representation. As a matter of fact, a graph-based representation and particularly conceptual graphs (CGs) [22] are an efficient solution to describe an image and characterize its components. The asset of this knowledge representation formalism is its flexible adaptation to the symbolic approach of image retrieval [12,18,19]. It allows indeed to represent components of our CBIR architecture and to develop an expressive and efficient framework as far as indexing and querying operations.

Formally, a conceptual graph is a finite, bipartite, connex and oriented graph. It features two types of nodes: the first one graphically represented by a rectangle (fig. 2) is tagged by a concept however the second represented by a circle is tagged by a conceptual relation. The graph of fig. 2 represents a man eating in a restaurant. Concepts and relations are identified by their type, itself corresponding to a semantic class. Concept and conceptual relation types are organized within a lattice structure partially ordered by '$\leq$' which expresses the relation 'is a specialization of'. For example, Person $\leq$ Man denotes that the type *Man* is a specialization of the type *Person*, and will therefore appear in the offspring of the latter within the lattice organizing these concept types. Within the scope of the model, conceptual graphs are used to represent the image content at the conceptual level. Each image (respectively user query) is represented by a conceptual graph called document index graph (respectively query graph) and evaluation of similarity between an image and a query is achieved through a correspondence function: the conceptual graph projection operator.

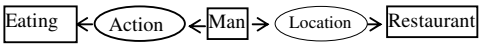


Fig. 2. An example of conceptual graph

3 A descriptive model for semantics: the visual semantics facet

3.1 Conceptual structures for the visual semantics facet

Semantic concept types are learned and then automatically extracted given a visual thesaurus. The construction of a visual thesaurus is strongly constrained by the application domain, indeed dealing with corpus of medical images would entail the elaboration of a visual thesaurus that would be different from a thesaurus considering computer-generated images. In this paper, our experiments presented in section 6 are based on a collection of personal photographs. Let us detail the elaboration of our visual thesaurus.

Several experimental studies presented in [14] have led to the specification of twenty categories or picture scenes describing the image content at the global level. Web-based image search engines (google, altavista) are queried by textual keywords corresponding to these picture scenes and 100 images are gathered for each query. These images are used to establish a list of concept types characterizing objects that can be encountered in these scenes. This process highlights seven major semantic concept types: people, ground, sky, water, foliage, mountain/rocks and building. We then use WordNet to produce a list of hyponyms linked with these concept types and discard terms which are not relevant as far as indexing and retrieving images from personal corpus are concerned. Therefore, we obtain a set of concept types which are specializations of the seven major semantic concept types. We repeat the process of finding hyponyms for all the specialized concept types. The last step consists in organizing all these concept types within a multi-layered lattice ordered by a specific/generic partial order. In fig. 3, the second layer of the lattice consists of concepts types which are specifications of the major semantic concept types, e.g. *face* and *crowd* are specifications of *people*. The third layer is the basic layer and presents the most specific concept types, e.g. *man*, *woman*, *child* are specifications of *individual*.

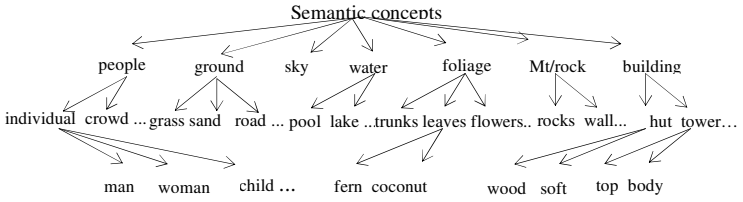


Fig. 3. Lattice of semantic concept types

A feed-forward neural network is used to learn these semantic concept types from a set of training and testing images. Low-level features [8] are computed for each training object and organized within a vector used as the input of the neural network. The learning being completed, an image is then processed by the network and the recognition results are aggregated to highlight image objects. An image object is thus characterized by a vector of semantic concept types, each one being linked to a value of recognition certainty. For instance, in the image of fig. 1, the second image object labeled as Io2 is characterized by a vector which has a significant value for the seman-

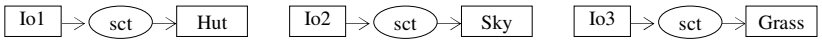
tic concept type *hut* and small values related to other semantic concepts. At the CG representation level, the semantic concept type with the highest recognition certainty is kept. As a matter of fact, Io2 will be represented by the semantic concept type *hut*. We will now specify the model organizing the visual semantics facet and deal with its representation in terms of CGs.

3.2 Model of the visual semantics facet

The model of visual semantics facet gathers semantic concept types and their lattice induced by a partial order relation: Msy = (SC, sct)

- SC is the set of visual semantics concept types.
- sct: I_{IO} → SC associates to each image object its semantic concept type.

Image objects are represented by *Io* concepts and the set SC is represented by a lattice of semantic concept types partially ordered by the relation $\leq_{vs}$. An instance of the visual semantics facet is represented by a set of CGs, each one containing an *Io* type linked through the conceptual relation *sct* to a semantic concept type. The basic graph controlling the generation of all visual semantic facet graphs is: [Io]→(sct)→[SC]. For instance, the following graphs are the representation of the visual semantics facet in fig. 1 and can be translated as: the first image object (Io1) is associated with the semantic concept type *hut*, the second image object (Io2) with the semantic concept type *sky* and the third image object (Io3) with the semantic concept type *grass*.



The integration of signal information within the conceptual level is crucial since it enriches the indexing framework and expands the query language with the possibility to query over both semantics and visual information. After presenting our formalism, we will now focus on the signal facet and deal with theoretical implications of integrating signal features within our multi-facetted conceptual model. This integration is not straightforward as we need to characterize low-level signal features at the conceptual level, and therefore specify a rich framework for conceptual signal indexing and querying. We first propose conceptual structures for the signal facet and then thoroughly specify the conceptual model for the signal facet and its representation in terms of CGs.

4 The signal facet: from low-level signal data to symbolic characterization

Our architecture and its supported operational model make it possible for a user to combine low-level features with visual semantics in a fully textual conceptual framework for querying. However, querying textually on low-level features requires specifying a correspondence process between color names and color stimuli.

Our symbolic representation of color information is guided by the research carried out in color naming and categorization [1] stressing a step of correspondence between color names and their stimuli. We will consider the existence of a formal system S_{nc} of color categorization and naming [7] which specifies a set of color categories Cat with a cardinal C_{cat}. These color categories are the C_i where variable i belongs to [1, C_{cat}]. Each image object is then indexed by two types of conceptual structures featuring its color distribution: boolean and quantified signal concepts.

When providing examples of the specified conceptual structures, we will consider that the color naming and categorization system highlights four color categories: cyan(c), green(gn), grey(g) and white(w) (Cat={cyan, green, grey, white}).

4.1 Index structures

Boolean signal index concept types (*BSICs*), gathered within the set B_{SI} are supported by a vector structure v_B with a number of elements equal to the number C_{cat} of color categories highlighted by the naming and categorization system. Values $v_B[i]$, $i \in [1,C_{cat}]$ are booleans stressing that the color category C_i is present in non-zero proportion within the considered image object. The semantics conveyed by BSICs is the 'And' semantics. As a matter of fact, these concept types feature the signal distribution of image objects by a conjunction of color categories. For instance, the first image object (Io1) corresponding to the semantic concept type *hut* in fig.1 is characterized by the BSIC <c:0,gn:0,g:1,w:1>, which is translated by Io1 having a signal distribution including grey **and** white.

We wish to extend the definition of BSICs and quantify by a variable known as **color category value** the integer percentage of pixels corresponding to a color category. The color category value corresponds to the standardization of the pixel percentages of each color category. Quantified signal index concept types (*QSICs*), gathered within the set Q_{SI} are supported by a vector structure v_Q with a number of elements equal to the number C_{cat} of color categories highlighted by the naming and categorization system. Values $v_Q[i]$, $i \in [1,C_{cat}]$ are the color category values. These concept types feature the signal distribution of image objects by a conjunction of color categories and their associated color category values. Let us note that the sum of category values is always 100, the color distribution being fully distributed between all color categories. The second image object (Io2) corresponding to the semantic concept type *sky* in fig.1 is characterized by the QSIC <c:59,gn:0,g:0,w:41>, which is translated by Io2 having a signal distribution including 59% of cyan **and** 41% of white.

4.2 Index structures

As far as querying is concerned, our conceptual architecture is powerful enough to handle an expressive and computationally efficient language consisting of boolean and quantified queries:

– A user shall be able to associate visual semantics with a boolean conjunction of color categories through an *And* query, such as Q1: "Find images with a grey and

white hut", a boolean disjunction of color categories through an *Or* query, such as Q2: "Find images with either cyan or grey sky" and a negation of color categories through a *No* query, such as Q3: "Find images with non-white flowers".

– As far as quantified queries, *At Most* queries (such as Q4: "Find images with a cloudy sky (at most 25% of cyan)") and *At Least* queries (such as Q5: "Find images with lake water (at least 25% of grey)") associate visual semantics with a set of color categories and a percentage of pixels belonging to each one of these categories. We specify also literally quantified queries (*Mostly*, *Few*) which can prove easier to handle by an average user, less interested in precision-oriented querying.

In the following sections we will present the conceptual structures supporting the previously defined query types.

4.2.1 Boolean signal query concept types. There are three categories of concepts types supporting boolean querying: *And* signal concept types (*ASCs*), gathered within the set B_{And} represent the color distribution of an image object by a conjunction of color categories; *Or* signal concept types (*OSCs*), gathered within the set B_{Or}, by a disjunction of color categories and *No* signal concept types (*NSCs*), gathered within the set B_{No}, by a negation of color categories. These concepts are respectively supported by vector structures v_{And}, v_{Or} and v_{No} with a number of elements equal to the number C_{cat} of color categories. Values $v_{And}[i]$, $v_{Or}[i]$ and $v_{No}[i]$, $i \in [1,C_{cat}]$ are non-null if the color category C_i is mentioned respectively in the conjunction, disjunction or negation of color categories within the query. The ASC corresponding to the color distribution expressed in query Q1 is $<$c:0, gn:0, g:1, w:1$>_{And}$. The color distribution expressed in query Q2 is translated in the OSC $<$c:1, gn:0, g:1, w:0$>_{Or}$. Finally, the color distribution expressed in query Q3 is translated in the NSC $<$c:0, gn:0, g:0, w:1$>_{No}$.

4.2.2 Signal quantified query concept types. There are two types of numerically quantified signal concept types : *At Most* signal concept types (*AMSCs*) gathered within the set Q_{AM} and *At Least* signal concept types (*ALSCs*) gathered within the set Q_{AL} that are respectively supported by vector structures v_{AM} and v_{AL} with a number of elements equal to C_{cat}.

If the color category C_i is specified in a query, values $v_{AM}[i]$ and $v_{AL}[i]$ ($i \in [1,C_{cat}]$) are non-null and correspond respectively to the maximum pixel percentage associated with C_i (translating the keyword 'At Most') and the minimum pixel percentage associated with C_i (translating the keyword 'At Least'). For instance, the color distribution expressed in query Q4 is translated in the AMSC $\mathbf{<}$c:25, gn:0, g:0, w:0$\mathbf{>_{AMSC}}$ whereas the color distribution expressed in query Q5 is translated in the ALSC $\mathbf{<}$c:0, gn:0, g:25, w:0$\mathbf{>_{ALSC}}$.

Expressing a query with numerical quantification is precision-oriented and an average user might find it cumbersome. Therefore we introduce literally quantified queries such as *Mostly* queries (e.g. "Find images with a bright sky (*Mostly* cyan)") and *Few* queries (e.g. "Find images with a cloudy sky (*Few* cyan)"). These queries are supported by sets Q_{Mostly} and Q_{Few} of *Mostly* and *Few* signal concept types. We set up a correspondence between the quantifier *Mostly* and the numeral quantification *At Least*

50%. Also, the quantifier *Few* will correspond to the numeral quantification *At Most 10%*.

After introducing structures for conceptual signal characterization, we propose a formal model organizing the signal facet. This model is then instantiated in the CG representation formalism within our image retrieval framework.

4.3 A conceptual model for the signal facet

The model of the signal facet M_{SI} is given by the model MI_{SI} of the signal index facet and the model MQ_{SI} of the signal query facet where $MI_{SI} = (I_{SI}, RI_{SI})$ and $MQ_{SI} = (Qu_{SI}, RQ_{SI})$:

- I_{SI} is the set of signal index structures: $I_{SI} = \{B_{SI} , Q_{SI}\}$
- RI_{SI} is the set of signal index conceptual relations: $RI_{SI} = \{b\_si, q\_si\}$

  b_si : $I_{IO} \rightarrow B_{SI}$ and q_si : $I_{IO} \rightarrow Q_{SI}$ associate image object identifiers with boolean and quantified signal index concept types.
- Qu_{SI} is the set of signal query structures: $Qu_{SI} = \{B_{And}, B_{Or}, B_{No}, Q_{AM}, Q_{AL}, Q_{Mostly}, Q_{Few}\}$
- RQ_{SI} is the set of signal query conceptual relations: $RQ_{SI} = \{$and_si , or_si, no_si, am_si, al_si, mostly_si, few_si$\}$

  and_si : $I_{IO} \rightarrow B_{And}$; or_si : $I_{IO} \rightarrow B_{Or}$ and no_si : $I_{IO} \rightarrow B_{No}$ associate image object identifiers with ASCs, OSCs and NSCs.

  am_si : $I_{Si} \rightarrow Q_{AM}$ and al_si : $I_{IO} \rightarrow Q_{AL}$ associate image object identifiers with AMSCs and ALSCs.

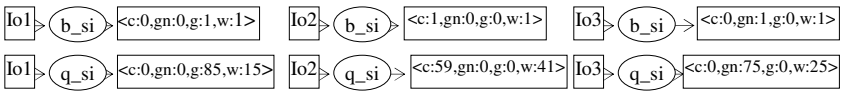
  mostly_si : $I_{IO} \rightarrow Q_{Mostly}$ and few_si : $I_{IO} \rightarrow Q_{Few}$ associate image object identifiers with *Mostly* and *Few* signal concept types.

Let us note that and_si and or_si are specialized relations of b_si. Also, am_si, al_si, mostly_si, few_si are specialized relations of q_si.

Image objects are represented by *Io* concepts and signal index and query structures are organized within a lattice of concept types. An instance of the signal facet is represented by a set of canonical CGs, each one containing an *Io* type possibly linked through signal conceptual relations to signal concept types.

There are two types of basic graphs controlling the generation of all signal facet graphs. The firsts are **index signal graphs**: [Io]→(b_si)→[B_{SI}]; [Io]→(q_si)→[Q_{SI}]. The seconds are **query signal graphs**: [Io]→(and_si)→[B_{And}]; [Io]→(or_si)→[B_{Or}] ; [Io]→(no_si)→[B_{No}] ; [Io]→(am_si)→[Q_{AM}]; [Io]→(al_si)→[Q_{AL}]; [Io]→(mostly_si)→[Q_{Mostly}] and [Io]→(few_si)→ [Q_{Few}].

The following index graphs are the representation of the signal facet in fig. 1:



5. The querying module

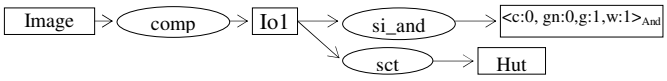
In image retrieval systems, the typical mode of user interaction relies on the query by example process [18]: the user provides a set of example images as an input, the system generates a query and then outputs images that are the most similar. This mode of interaction suffers from the fact that the user's need remains implicit, i.e. given the input images chosen by the user, the system has thus to use its knowledge of the image content to extract implicit information and construct a query. This process can be very complex and lead to ambiguities and poor retrieval performances when dealing with high-level characterizations of an image. Our conceptual architecture is based on a unified textual-based framework allowing a user to query over both the visual semantics and the signal facets. This obviously enhances user interaction since contrarily to query by example systems, the user becomes in 'charge' of the query process by making his needs explicit to the system through full textual querying. We will present in the following some queries involving boolean and quantified signal concepts, study their transcription within our conceptual framework and then deal with operations related to query processing. We will thus specify the organization of concept type lattices.

5.1 Query expression

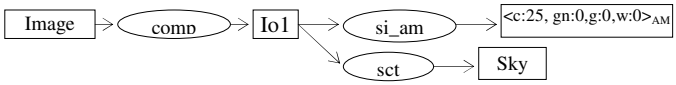
A general query is defined through a combination of selection criteria over the visual semantics and the signal facets. A query image q is represented by a 3-tuple (I_{IO}, q_{vs}, q_{si}) where I_{IO} is the set of image objects, q_{vs} and q_{si} are instances of the visual semantics and query signal facet models.

We propose to study the transcription in our model and then the processing of two types of queries for obvious space restrictions: the first one associates visual semantics with a boolean signal concept type (*And* signal concept type), the second associates visual semantics with a quantified signal concept type (*At Most* signal concept type).

5.1.1 ***Find images with a grey and white hut.*** In our formalism, it is translated as: $q=(I_{IO},q_{vs},q_{si})$ with $I_{IO}=\{Io1\}$; $q_{vs}=(\{Hut\},\{(Io1,Hut)\})$; $q_{si}=(\{<c:0,gn:0,g:1,w:1>_{And}\}, \{(Io1,<c:0,gn:0,g:1,w:1>_{And})\})$. In the CG representation formalism, we have:



5.1.2 ***Find images with a cloudy sky (At Most 25% of cyan).*** The transcription of this query in our conceptual framework is: $q=(I_{IO}, q_{vs}, q_{si})$: $I_{IO}=\{Io1\}$; $q_{vs}=(\{Sky\}, \{(Io1,Sky)\})$; $q_{si}=(\{<c:25,gn:0,g:0,w:0>_{AM}\}, \{(Io1,<c:25,gn:0,g:0,w:0>_{AM})\})$.
The transcription of this query in the CG representation formalism gives:



5.2 The projection operator

An operational model of image retrieval based on the CG formalism uses the graph projection operation for the comparison of a query graph and a document graph. This operator allows to identify within a graph g_1 sub-graphs with the same structure as a given graph g_2, with nodes being possibly restricted, i.e. their types are specialization of g_2 node types. If it exists a projection of a query graph Q within a document graph D then the document indexed by D is relevant for the query Q.

Formally, the projection operation $\wp : q \rightarrow d$ exists if there is a sub-graph of d verifying the two following properties:

- There is a unique document concept which is a specific of a query concept, this being valid for any query concept. This property ensures that all elements describing the query are present within the image document, and their image is unique.
- For any relation linking concepts c_{q1} and c_{q2} of q, there is the same relation between the two concepts c_{d1} and c_{d2} of d, such as $\wp(c_{q1}) = c_{d1}$ and $\wp(c_{q2}) = c_{d2}$.

At the implementation level, brute-force coding of the projection operation would result in exponential execution times. Based on the work in [19], we enforce the scalability of our framework using an adaptation of the inverted file approach for image retrieval. This technique consists in associating indexed keywords to the set of documents whose index contain it. Treatments that are part of the projection operation are performed during indexing following a specific organization of CGs which does not affect the expressiveness of the formalism.

5.3 Organizing concept type lattices for effective and computationally efficient retrieval

In the following concept type lattices (fig. 4,5), the graphical arrow corresponds to a specialization operation and we consider that Cat={cyan, green, grey, white}.

5.3.1 Processing an *And* query. BSICs are organized within the *And* lattice (fig. 4) to process an *And* query. When a query such as "Find images with a grey and white hut" is formulated, it is first translated in a query CG with the semantic concept type *hut* processed by the lattice of semantic concept types (fig. 3) and the ASC $<c:0,gn:0,g:1,w:1>_{And}$. This ASC is then related to its equivalent BSIC as highlighted in fig. 4. The most relevant images provided by the system have a hut with grey and white only, this symbolic color distribution is represented by the highlighted BSIC (b_1) in fig. 4. Other images are composed of a hut with a color distribution including grey and white and at least one secondary color category. In the lattice, BSICs representing such color distributions are sons of b_1.

The general organization of this lattice is such that BSICs with a unique non-zero component are sons of the maximum virtual element T_{And}. They represent the perception of a unique color category in an image object. The BSIC with all non-zero components is at the bottom of the hierarchy, it is the minimum element noted $\perp_{And}$. This

concept is a specialized concept of all BSICs presenting at least a non-zero component. Formally, we define a partial order in the *And* lattice of BSICs noted $\leq_{And}$ by:

$$\forall\ a,b \in B_{SI}\ \ a \leq_{And} b \Leftrightarrow [a = \bot_{And} \vee b = \top_{And}] \vee [\neg\exists k \in [1,C_{cat}] / b_{[k]} = 1 \wedge a_{[k]} = 0] \quad (\mathbf{1})$$

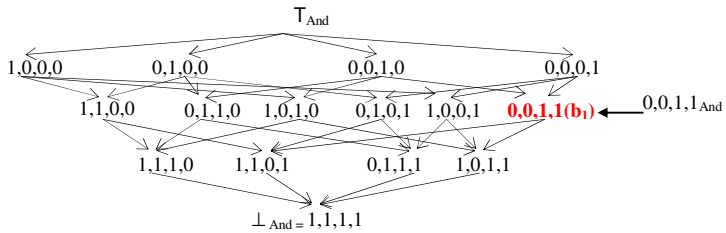


Fig. 4. Lattice processing *And* queries

5.3.2 Processing an *At Most* query. When a query such as "Find images with a cloudy sky (i.e. with a color distribution that includes at most 25% of cyan)" is formulated, it is translated in a query CG with the semantic concept type *sky* processed by the lattice of semantic concept types (fig. 3) and the AMSC <c:25,g:0,gn:0,w:0>$_{AM}$. However, the link between this AMSC and its equivalent QSIC is not straightforward. Therefore we introduce a new category of concepts types bridging the gap between AMSCs and QSICs by taking into account dominant color categories (i.e. categories mentioned in a query as they have a higher importance in the ordering process of signal concepts within the lattice, other color categories are called secondary). These concept types are QSICs with dominant d_{AM}, where d_{AM} is the set of dominant color categories. They are supported by a vector structure $v_{AMd}[i]$ with a number of elements equal to C_{cat}+1. The $v_{AMd}[i]_{i\in[1,Ccat+1]}$ values such that $C_i \in d_{AM}$ are the maximum pixel percentages of dominant color categories and the $v_{AMd}[j]_{j\in[1,Ccat+1]}$ such that $j \neq i$ correspond to the pixel percentages of secondary color categories ranked in ascending order. A component summing pixel percentages of secondary color categories noted $\sum$ is introduced. By construction, this element is the maximum value among the $v_{AMd}[j]_{j\in[1,Ccat+1]}$. QSICs with dominant d_{AM} are therefore specializations of AMSCs and generalizations of QSICs and link AMSCs to QSICs. The AMSC <c:25,g:0,gn:0,w:0>$_{AM}$ is related to its equivalent QSIC with dominant {cyan}: <25,25,25,25,75> as highlighted in fig. 5a. As a matter of fact, the most relevant images provided by the system have a sky with 25% of cyan and a remaining proportion uniformly distributed between the 3 secondary color categories (25% each in our example). Others are images with a sky having a color distribution that includes less than 25% of cyan, the remaining proportion *p* being in the best cases uniformly distributed between the 3 secondary color categories.

Formally, sub-lattices of AMSCs with dominant d_{AM} (framed structure in fig. 5a) are partially ordered by $\leq_{AM}$:

$$\forall\ a,b \text{ QSICs with dominant } d_{AM},\ a \leq_{AM} b \Leftrightarrow [a=\bot_{AM} \vee b=\top_{AM}] \vee [\forall j \in [1,C_{Cat}] / Cat_j \in d_{AM},\ 1 \leq a_{[j]} \leq b_{[j]}] \quad (\mathbf{2})$$

Sub-lattices of concept types with components corresponding to dominant color categories being equal (framed structure in fig. 5b) are partially ordered by $\leq_{AM\_eq}$:

$\forall$ a,b QSICs with dominant d_{AM} having components that correspond to dominant color categories being equal, $a \leq_{AM\_eq} b \Leftrightarrow (\forall j, k \in [1, C_{Cat} + 1] / Cat_j \notin d_{AM} \wedge Cat_k \notin d_{AM}, \sum_{j,k} | b_{[j]} - b_{[k]} | \leq \sum_{j,k} | a_{[j]} - a_{[k]} |)$ **(3)**

Let us note than the *At Least* lattice has a symmetric organization and will not be dealt with for space restriction.

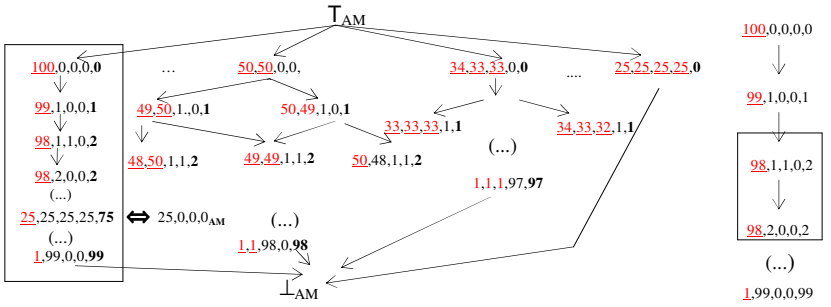


Fig. 5. (a) Sub-lattice of At Most signal concepts (b) with dominant {C1=cyan}

5.3.3 Processing a query with a literal quantifier. *Mostly* and *Few* queries involving literal quantifiers, e.g. "Find images with a bright sky (mostly cyan)" or "Find images with a cloudy sky (few cyan)", are processed accordingly to *At Most* and *At Least* queries. Indeed, the quantifier Mostly corresponds to the numeral quantification 'At Least 50%' and the quantifier Few is linked to the numeral quantification 'At most 10%'. As a matter of fact, processing these queries will not affect the computational efficiency of our model as it is based on AMSCs and ALSCs concept type lattices.

6. Experimental results

We have presented a conceptual architecture in which semantic and signal features are integrated to achieve higher expressivity as far as querying is concerned and increased retrieval accuracy. We will describe here the SIAIR image retrieval system that is an implementation of the theoretical framework presented and present several experimentation results.

The SIAIR image retrieval system implements the formal framework presented in this paper, the supported mode of interaction relying on keyword-based search. When a user enters a query, it is translated in a CG query graph as developed in section 5. It is then processed and images given by the system are ranked and displayed according to their relevance with respect to the query.

Validation experiments are carried out on a corpus of 2500 personal color photographs collected over a period of five years and used as a validation corpus in world-class publications [9,15] (fig. 1 displays a typical photograph which belongs to this

collection). Dealing with personal photographs instead of professional collections (e.g. Corel images) is guided by our research problematic which is the specification of an expressing framework enhancing techniques that allow a user to index and query over a collection of home photographs. Moreover, the quality of home photographs is not as good as the quality of professional images which leads to retrieval results being generally poorer than those for the Corel images.

Image objects within the 2500 photographs are automatically assigned a semantic concept as presented in section 3 and are characterized with conceptual signal structures presented in section 4. Eleven color categories (red, green, blue, yellow, cyan, purple, black, skin, white, grey, orange) empirically spotlighted in [4] are described in the HVC perceptually uniform space by a union of brightness, tonality and saturation intervals [13].

Given an image corpus, we wish to retrieve photographs that represent elaborate scenes involving signal characterization. We specify 22 image scenes (e.g. night, swimming-pool water…) and select within the corpus for each scene all images which are relevant. The evaluation of our formalism is based on the notion of **image relevance** which consists in quantifying the correspondence between index and query images.

We compare our approach with both state-of-the-art signal and semantics-based approaches, namely "HSV local" and "Visual keywords". The HSV local method is based on the specification of ten key colors (red, green, blue, black, grey, white, orange, yellow, brown, pink) in the HSV color space adopted by the original PicHunter system [3]. The similarity matching between two images is computed as the weighted average of the similarities between corresponding blocks of the images. As a matter of fact, this method is equivalent to locally weighted color histograms.

Visual keywords [8,9,15] are intuitive and flexible visual prototypes extracted or learned from a visual content domain with relevant semantic labels. A set of 26 specified visual keywords are learned using a neural network, with low-level features computed for each training region as an input for this network. An image is indexed to multi-scale, view-based recognition against these 26 visual keywords, recognition results across multiple resolutions are aggregated according to spatial tessellation. It is then represented by a set of local visual keyword histograms with each bin corresponding to the aggregation of recognition results. The similarity matching between two images is defined as the weighted average of the similarities between their corresponding local visual keywords histograms. The HSV local and Visual Keywords methods are presented here to compare the results of usual signal-based and semantic-based approaches to our framework combining both of these approaches.

For each of the 22 image scene descriptions (e.g. swimming-pool water), we construct relevant textual query terms using corresponding semantic and signal concepts as input to the SIAIR system (e.g. "Find images with mostly cyan" for swimming-pool water). Also each image scene description is translated in textual signal data as input to the HSV local approach ("Find images with cyan" for swimming-pool water) and in relevant visual keywords to be processed by the Visual keywords system ("Find images with a sky" for swimming-pool water). Curves associated with the *Q_SymbColor*, *Q_Symb* and *Q_Color* legends (fig. 6) illustrate respectively the results in recall and precision obtained by SIAIR, the Visual Keywords and the HSV local systems.

The average precision of SIAIR (0.5854) is approximately five times higher than the average precision of the Visual Keywords method (0.1115) and approximately 3,5 times higher than the value of average precision of the HSV local method (0.168). We notice that improvements of the precision values are significant at all recall values. This shows that when dealing with elaborate queries which combine multiple sources of information (here visual semantics and signal features) and thus require a higher level of abstraction, the use of an "intelligent" and expressive representation formalism (here the CG formalism within our framework) is crucial. As a matter of fact, SIAIR complements both state-of-the-art signal-based approaches by proposing a framework for semantic characterization and state-of-the-art semantic-based methods through signal conceptual integration, which enriches indexing languages and expands usual querying frameworks restricted to a reduced set of extracted or learned keywords (in this case the visual keywords).

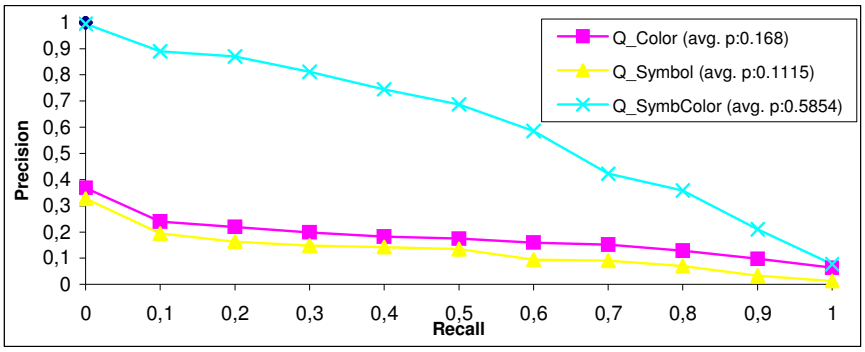


Fig. 6. Recall/Precision curves

7. Conclusion

We have proposed within the scope of this paper the formal specification of a framework combining the two existing approaches in image retrieval, i.e. signal and symbolic within a strongly-coupled architecture to achieve greater retrieval accuracy. It is instantiated by an operational model based on the CG formalism, which allows to define an image representation and a correspondence function to compare index document and query graphs. Our work has contributed both theoretically and at the experimental level to the image retrieval research topic. We have specified image objects, abstract structures representing visual entities within an image in order to operate image indexing and retrieval operations at a higher level of abstraction than state-of-the-art frameworks. We have formally described the visual semantics and the signal facets that define the conceptual information conveyed by image objects and have finally proposed a unified and rich framework for querying over both visual semantics and signal data. At the experimental level, we have implemented and evalu-

ated our framework. The results obtained allowed us to validate our approach and stress the relevance of the signal/semantics integration.

References

1. Belkhatir, M. & Mulhem, P. & Chiramella, Y. Integrating perceptual signal features within a multi-facetted conceptual model. ECIR, - , 2004.
1. Berlin, B. & Kay, P.: Basic Color Terms: Their universality and Evolution. UC Press (1991)
2. Carson, C. & al.: Blobworld: A System for Region-Based Image Indexing and Retrieval. VISUAL (1999) 509-516
3. Cox, I. & al.: The Bayesian Image Retrieval System, PicHunter: Theory, Implementation and Psychophysical Experiments. IEEE Trans. Image Processing, vol.9, no.1 (2000) 20-37
4. Gong, Y. & Chuan, H. & Xiaoyi, G.: Image Indexing and Retrieval Based on Color Histograms. Multimedia Tools and Applications II (1996) 133-156
5. Gupta, A. & Weymouth, T. & and Jain, R: Semantic queries with pictures: The VIMSYS model. VLDB (1991) 69-79
6. La Cascia & al.: Combining Textual and Visual Cues for Content-Based Image Retrieval on the World Wide Web. IEEE Workshop on CB Access of Im. and Vid. Lib. (1998) 24-28
7. Lammens, J.M.: A computational model of color perception and color naming. PhD, State Univ. of New York, Buffalo (1994)
8. Lim, J.H.: Explicit query formulation with visual keywords. ACM MM (2000) 407-412
9. Lim, J.H. & al.: Home Photo Content Modeling for Personalized Event-Based Retrieval. IEEE Multimedia, vol.10, no.4 (2003)
10. Lu, Y. & al.: A unified framework for semantics and feature based relevance feedback in image retrieval systems. ACM MM (2000) 31-37
11. Ma, W.Y. & Manjunath, B.S. 1997.: NeTra: A toolbox for navigating large image databases. ICIC (1997) 568-571
12. Mechkour, M.: EMIR^2: An Extended Model for Image Representation and Retrieval. DEXA (1995) 395-404
13. Miyahara, M. & Yasuhiro Yoshida, Y.: Mathematical Transform of (R,G,B) Color Data to Munsell (H,V,C) Color Data. SPIE Vol. 1001 (1988) 650-657
14. Mojsilovic, A. & Rogowitz, B.: Capturing image semantics with low-level descriptors. ICIP (2001) 18-21
15. Mulhem, P. & Lim, J.H.: Symbolic photograph content-based retrieval. ACM CIKM (2002) 94-101
16. Niblack, W. et al.: The QBIC project : Querying images by content using color, texture and shape. SPIE, Storage and Retrieval for Image and Video Databases (1993) 40-48
17. Ounis, I. & Pasca, M.: RELIEF: Combining expressiveness and rapidity into a single system. ACM SIGIR (1998) 266-274
18. Smeulders, A.W.M. & al.: Content-based image retrieval at the end of the early years. IEEE PAMI, 22(12) (2000) 1349-1380
19. Smith, J.R. & Chang, S.F.: VisualSEEk: A fully automated content-based image query system. ACM MM (1996) 87-98
20. Sowa, J.F. "Conceptual structures : information processing in mind and machine". Addison-Wesley publishing company (1984)
21. Zhou, X.S. & Huang, T.S.: Unifying Keywords and Visual Contents in Image Retrieval. IEEE Multimedia 9(2) (2002) 23-33