# 2

# Towards Intelligible Query Processing in Relevance Feedback-Based Image Retrieval Systems

Belkhatir Mohammed
*Faculty of IT, Monash University*

## 1. Introduction

We propose in this paper the specification of an image retrieval architecture based on a relevance feedback framework which operates on high-level image descriptions instead of their extracted low-level features. This framework features a conceptual model which integrates visual semantics as well as symbolic relational characterizations and operates on image objects, abstractions of visual entities within a physical image. Also, it manipulates a rich query language, consisting of both boolean and quantification operators, which therefore leads to optimized user interaction and increased retrieval performance. Let us first introduce the context of our research.

In order to cope with the storing and retrieval of ever-growing digital image collections, the first retrieval systems (cf. [Smeulders et al. 00] for a review of the state-of-the-art), known as content-based, propose fully automatic processing methods based on low-level signal features (color, texture, shape...). Although they allow the fast processing of queries, they do not make it possible to search for images based on their semantic content and consider for example red apples or Ferraris as being the same entities simply because they have the same color distribution. Failing to relate low-level features to semantic characterization (also known as the **semantic gap)** has slowed down the development of such solutions since, as shown in [Hollink 04], taking into account aspects related to the image content is of prime importance for efficient retrieval. Also, users are more skilled in defining their information needs using language-based descriptors and would therefore rather be given the possibility to differentiate between red roses and red cars.

In order to overcome the semantic gap, a class of frameworks within the framework of the European Fermi project proposed to model the image semantic and signal contents following a sharp process of human-assisted indexing [Mechkour 95] [Meghini et al. 01]. These approaches, based on elaborate knowledge-based representation models, provide satisfactory results in terms of retrieval quality but are not easily usable on large collections of images because of the necessary human intervention required for indexing.

Automated systems which attempt to deal with the semantics/signal integration (e.g. iFind [Lu et al. 00] and the prototype presented in [Zhou & Huang 02]) propose solutions based on textual annotations to characterize semantics and on a relevance feedback (RF) scheme operating on low-level features. RF techniques are based on an interaction with a user providing judgment on displayed images as to whether and to what extent they are relevant or irrelevant to his need. For each loop of the interaction, these images are learnt and the system tries to display images close in similarity to the ones targeted by the user. As any learning process, it requires an important number of training images to achieve reasonable performance. The user is therefore solicited through several tedious and time-consuming loops to provide feedback for the system in real time, which penalizes user interaction and involves costly computations over the whole set of images. Moreover, starting from a textual query on semantics, these state-of-the art systems are only able to manage **opaque** RF (i.e. a user selects relevant and/or non-relevant documents and is then proposed a revised ranking without being given the possibility to 'understand' how his initial query was transformed) since it operates on extracted low-level features. Finally, these systems do not take into account the relational spatial information between visual entities, which affects the quality of the retrieval results.

Our RF process is a specific case of state-of-the-art RF frameworks reducing the user's burden since it involves a unique loop returning the relevant images. Moreover, as opposed to the opacity of state-of-the-art RF frameworks, it holds the advantage of being **transparent** (i.e. the system displays the query generated from the selected documents) and **penetrable** (i.e. the modification of the generated query is possible before processing), which increases the quality of retrieval results. Through the use of a symbolic representation, the user is indeed able to visualize and comprehend the intelligible query being processed. We manage transparent and penetrable interactions by considering a conceptual representation of images and model their conveyed visual semantics and relational information through a high-level and expressive representation formalism. Given a user's feedback (i.e. judgment or relevance or irrelevance), our RF process, operating on both visual semantics and relational spatial characterization, is therefore able to first generate and then display a query for eventual further modifications operated by the user. It enforces computational efficiency by generating a symbolic query instead of dealing with costly learning algorithms and optimizes user interaction by displaying this 'readable' symbolic query instead of operating on hidden low-level features.

As opposed to state-of-the-art loosely-coupled solutions penalizing user interaction and retrieval performance with an opaque RF framework operating on low-level features, our architecture combines a keyword-based module with a transparent and penetrable RF process which refines the retrieval results of the first. Moreover, we offer a rich query language consisting of several Boolean operators.

At the core of our work is the notion of **image objects** (**IOs**), abstract structures representing visual entities within an image. Their specification is an attempt to operate beyond simple low-level signal features since IOs convey the semantic and relational information.

In the remainder, we first detail the processes allowing to abstract the extracted low-level features to high-level relational description in section 2. Section 3 deals with the visual semantic characterization. We specify in section 4 the image model and develop its conceptual instantiation integrating visual semantics and relational (spatial) features. Section 5 is dedicated to the presentation of the RF framework.

## 2. From low-level spatial features to high-level relational description

Taking into account spatial relations between semantically-defined visual entities is crucial in the framework of an image retrieval system since it enriches the index structures and expands the query language. Also, dealing with relational information between image components allows to enhance the quality of the results of an information retrieval system [Ounis&Pasca 98]. However, relating low-level spatial characterizations to high-level textual descriptions is not a straightforward task as it involves highligting a spatial vocabulary and specifying automatic processes for this mapping. We first study in this section methods used to represent spatial data and deal with the automatic generation of high-level spatial relations following a first process of low-level extraction.

### 2.1 Defining a spatial vocabulary through the relation-oriented approach

We consider two types of spatial characterizations: the first describes the absolute positions of visual entities and the second their relative locations.

In order to model the spatial data, we consider the «relation-oriented» approach which allows explicitly representing the relevant spatial relations between IOs without taking into account their basic geometrical features. Our study features the four modeling and representation spaces:

- The Euclidean space gathers the image pixels coordinates. Starting with this information, all knowledge related to the other representation spaces can be inferred.
- The Topological space is itself linked to the notions of continuity and connection. We consider five topological relations and justify this choice by the fact that these relations are exhaustive and relevant in the framework of an image indexing and retrieval system. Let io1 and io2 two IOs. These relations are ($s_1$=**P**,io1,io2) : 'io1 is a part of io2', ($s_2$=**T**,io1,io2) : 'io1 touches io2 (is externally connected)', ($s_3$=**D**,io1,io2) : 'io1 is disconnected from io2', ($s_4$=**C**,io1,io2) : 'io1 partially covers (in front of) io2' and ($s_5$=**C_B**,io1,io2) : 'io1 is covered by (behind) io2'. Let us note that these relations are mutually exclusive and characterized by the important property that each pair of IOs is linked by only one of these relations.
- The Vectorial space gathers the directional relations: Right ($s_6$=**R**), Left ($s_7$=**L**), Above ($s_8$=**A**) and Below ($s_9$=**B**). These relations are invariant to basic geometrical transformations such as translation and scaling.
- In the metric space, we consider the fuzzy distance relations Near ($s_{10}$=**N**) and Far ($s_{11}$=**F**). Discrete relations are not considered since providing a query language which allows a user to quantify the distance between two visual entities would penalize the fluidity of the interaction.

### 2.2 Automatic spatial characterization

**Topological relations.** In our spatial modeling, an IO *io* is characterized by its center of gravity io_c and by two pixel sets: its interior, noted io_i and its border io_b. We define for an image an orthonormal axis with its origin being the image left superior border and the basic measure unity, the pixel. All spatial characterizations of an object such as its border, interior and center of gravity are defined with respect to this axis.

In order to highlight topological relations between IOs, we consider the intersections of their interior and border pixel sets through a process adapted from [Egenhofer 91]. Let io1 and io2 be two IOs, the four intersections are: io1_i ∩ io2_i, io1_i ∩ io2_b, io1_b ∩ io2_i and io1_b ∩ io2_b. Each topological relation is linked to the results of these intersections as illustrated in table 1. The strength of this computation method relies on associating topological relations to a set of necessary and sufficient conditions linked to spatial attributes of IOs (i.e. their interior and border pixel sets).

| Topological Relation \ Intersections | io1_b ∩ io2_b | io1_i ∩ io2_b | io1_b ∩ io2_i | io1_i ∩ io2_i |
|---|---|---|---|---|
| (P, io1, io2) | ∅ | ≠ ∅ | ∅ | ≠ ∅ |
| (T, io1, io2) | ≠ ∅ | ∅ | ∅ | ∅ |
| (D, io1, io2) | ∅ | ∅ | ∅ | ∅ |
| (C, io1, io2) | ∅ | ∅ | ≠ ∅ | ≠ ∅ |
| (C_B, io1, io2) | ∅ | ≠ ∅ | ∅ | ≠ ∅ |

Table 1. Characterization of topological relations with the intersections of interior and border pixel sets of two IOs

**Directional relations**. The computation of directional relations between io1 and io2 is based on their centers of gravity io1_c($x1_c$, $y1_c$) and io2_c($x2_c$, $y2_c$), the minimal and maximal coordinates along x axis ($x1_{min}$, $x2_{min}$ & $x1_{max}$, $x2_{max}$) as well as the minimal and maximal coordinates along y axis ($y1_{min}$, $y2_{min}$ & $y1_{max}$, $y2_{max}$) of their four extremities.

We will say that io1 is at the left of io2, noted (L,io1,io2) iff. $(x1_c<x2_c) \wedge (x1_{min}<x2_{min}) \wedge (x1_{max}<x2_{max})$.

io1 is at the right of io2, noted (R,io1,io2) iff. $(x1_c>x2_c) \wedge (x1_{min}>x2_{min}) \wedge (x1_{max}>x2_{max})$.

We will say that io1 is above io2, noted (A,io1,io2) iff. $(y1_c>y2_c) \wedge (y1_{min}>y2_{min}) \wedge (y1_{max}>y2_{max})$.

io1 is below io2, noted (B,io1,io2) iff. $(y1_c<y2_c) \wedge (y1_{min}<y2_{min}) \wedge (y1_{max}<y2_{max})$.

We illustrate these definitions in figure 1 where the IO corresponding to huts (io1) is above the IO corresponding to the grass (io2). It is however not at the left of the latter since $x1_c<x2_c$ but $x1_{min}>x2_{min}$.

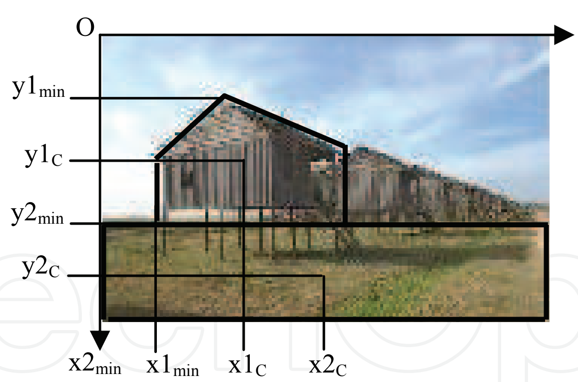



Figure 1. Characterization of directional relations

**Metric relations.** In order to distinguish between the Near and Far relations, we use the constant $D_{sp}= d(\vec{0},0.5*[\sigma_1,\sigma_2]^T)$ where d is the Euclidean distance between the null vector $\vec{0}$ and $[\sigma_1,\sigma_2]^T$ is the vector of standard deviations of the localization of centers of gravity for each IO in each dimension from the overall spatial distribution of all IOs in the corpus. $D_{sp}$ is therefore a measure of the spread of the distribution of centers of gravity of IOs. This distance agrees with results from psychophysics and can be interpreted as the bigger the spread, the larger the distances between centers of gravity are. We will say that two IOs are **near** if the Euclidean distance between their centers of gravity is inferior to $D_{sp}$, **far** otherwise.

### 2.3 From low-level features to symbolic spatial relations

So as to deduct knowledge from partial spatial information and to enforce computational efficiency, composition rules are used to infer relations between two IOs io1 and io2 from the relations generated between io1, io2 and a third IO io3. For example, if io1 is at the left of io3 and io3 at the left of io2 then io1 is at the left of io2.

Composition rules on spatial relations are dynamically processed when constructing index spatial representations. Let us note moreover that there are existing implications between spatial relations characterized in different modeling spaces. We identified the following implications related to the topological relations only:

- (P,io1,io2)→ ¬ (T, io1, io2) ∧ ¬ (D, io1, io2) ∧ ¬ (C, io1, io2) ∧ ¬ (C_B, io1, io2)
- (T,io1,io2)→ ¬ (P, io1, io2) ∧ ¬ (D, io1, io2) ∧ ¬ (C, io1, io2) ∧ ¬ (C_B, io1, io2)
- (D,io1,io2)→ ¬ (P, io1, io2) ∧ ¬ (T, io1, io2) ∧ ¬ (C, io1, io2) ∧ ¬ (C_B, io1, io2)
- (C,io1,io2)→ ¬ (P, io1, io2) ∧ ¬ (T, io1, io2) ∧ ¬ (D, io1, io2) ∧ ¬ (C_B, io1, io2)
- (C_B,io1,io2)→ ¬ (P, io1, io2) ∧ ¬ (T, io1, io2) ∧ ¬ (D, io1, io2) ∧ ¬ (C, io1, io2)

These implications illustrate the fact that there exists a unique topological relation between two IOs.

We identified the following implications related to the directional relations:

- (L,io1,io2) → ¬ (R,io1,io2); (R,io1,io2) → ¬ (L,io1,io2)
- (A,io1,io2) → ¬ (B,io1,io2); (B,io1,io2) → ¬ (A,io1,io2)

These implications illustrate the fact that an IO io1 is either at the left or at the right of a second IO io2. Also, it is either above, either below io2.

We identified the following implications between metric relations only:

- (N,io1,io2) → ¬ (F,io1,io2); (F,io1,io2) → ¬ (N,io1,io2)

These implications illustrate the fact that an IO io1 is either near, either far from a second IO io2.

Finally, we identified the following implications between spatial relations of distinct natures:

- (P, io1, io2) → N, io1, io2), if io1 is **part of** io2, then it is **near** io2.
- (T, io1, io2) → (N, io1, io2), if io1 **touches** io2, then it is **near** io2.

We propose in the next section to highlight the image visual semantics, i.e. semantic concepts linked to IOs.

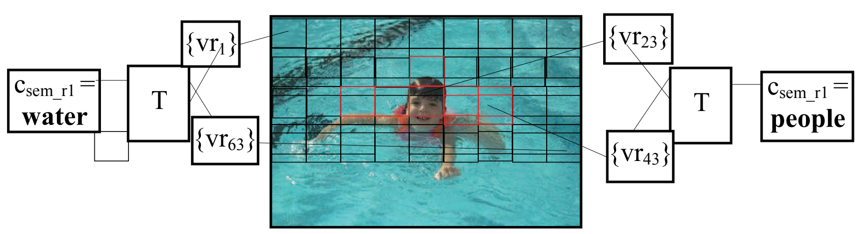
## 3. Characterizing the visual semantics

Semantic concepts are learned and then automatically extracted given a visual ontology. Its specification is strongly constrained by the application domain. Indeed, the development of cross-domain multimedia ontologies is currently limited by the difficulty to automatically map low-level signal features to semantic concepts [Naphade et al. 06]. Our efforts have been focused towards developing an ontology for general-purpose photography.

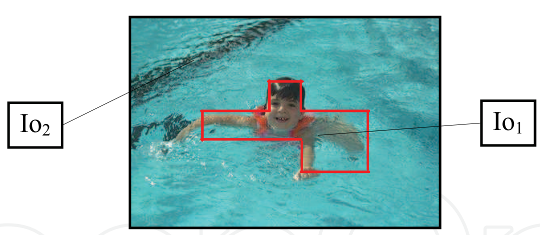
Several experimental studies presented in [Mojsilovic&Rogowitz 01] have led to the specification of twenty categories or picture scenes describing the image content at a global level. Web-based image search engines (google, altavista) are queried by textual keywords corresponding to these picture scenes and 100 images are gathered for each query. These images are used to establish a list of semantic concepts characterizing objects that can be encountered in these scenes. A total of 72 semantic concepts to be learnt and automatically extracted are specified.

Figure 2. Image patches corresponding to semantic concepts: ground, sky, vegetation, water, people, mountain, building

A three-layer feed-forward neural network with dynamic node creation capabilities is used to learn these semantic concepts. Labeled image patches cropped from home photographs constitute the training corpus T (example images are provided in figure 3). Low-level color and texture features are computed for each of the training images as an input vector for the neural network.




a)Learning framework linking each grid-based region with a semantic-concept and its recognition result




b)Recognition results are reconciled across all regions to highlight IOs

Figure 3. Architecture for the highlighting of IOs and the characterization of their corresponding semantic concept

Once the neural network has learned the visual vocabulary, the approach subjects an image to be indexed to a multi-scale, grid-based recognition against these semantic concepts. An image to be processed is scanned with grids of several scales. Each one features visual regions $\{vr_i\}$ characterized by a feature vector of low-level color and texture features. The latter is compared against feature vectors of labeled image patches corresponding to semantic concepts in the training corpus T (figure 3.a)). Recognition results for all semantic concepts are computed and then reconciled across all grid regions which are aggregated according to configurable spatial tessellation (figure 3.b)) in order to highlight IOs. Each IO is linked to a semantic concept with maximum recognition value.

## 4. A model for semantic/relational integration

We propose an image model combining visual semantics and relational characterization through a bi-facetted representation (cf. figure 4). The image model consists of both a physical image level representing an image as a matrix of pixels and a conceptual level. IOs convey the visual semantics and the relational information at the conceptual level. The latter is itself a bi-facetted framework:

- The **visual semantics facet** describes the image semantic content and is based on labeling IOs with a semantic concept. E.g., in figure 4, the second IO (Io2) is tagged by the semantic concept *Water*. Its conceptual specification is dealt with in section 4.1.
- The **relational facet** features the image relational content in terms of symbolic spatial relations. E.g., in figure 4, Io1 is **inside** Io2. Its conceptual specification is dealt with in section 4.2.

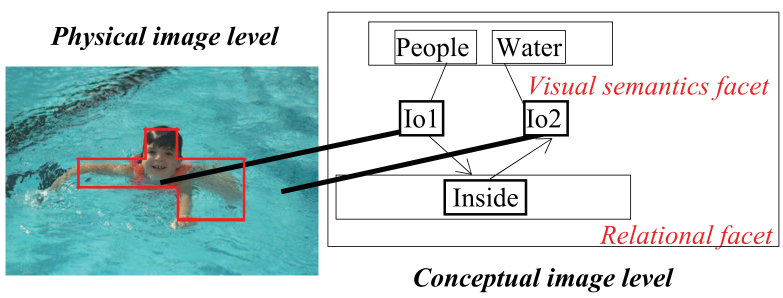



Figure 4. Image Model

To instantiate this model within an image retrieval framework, we use a representation formalism capable to model IOs as well as the conveyed visual semantics and relational information. This formalism should moreover make it easy to visualize the image information, especially as far as the interaction with the user within a RF framework is concerned. A graph-based representation and particularly conceptual graphs (CGs) [Sowa 84] is an efficient solution to describe an image and characterize its components. CGs have indeed proven to adapt to the symbolic approach of image retrieval [Mechkour 96] [Belkhatir et al. 04] [Belkhatir 05a] [Belkhatir et al. 05b]. CGs allow to represent components of our image retrieval architecture and to specify expressive index and query frameworks. Formally, a CG is a finite, bipartite and directed graph. It features two types of nodes: concept and relation nodes. In the graph [Tools with Artificial Intelligence]←(Entitled) ←[Book]→(Published_by)→[I-Tech], concepts are between brackets and relations between parentheses. This graph is equivalent to a first-order logical expression where concepts and relations are connected by the conjunction operator (boolean **AND**):

$\exists$ x,y,z s.t. (Book=x) $\wedge$ (Tools with Artificial Intelligence=y) $\wedge$ (I-Tech=z) $\wedge$ Entitled(x,y) $\wedge$ Published_by(x,z).

It is semantically interpreted as: the book entitled Tools with Artificial Intelligence is published by I-Tech. Concepts and conceptual relations are organized within a lattice structure partially ordered by the IS-A ($\leq$) relation. Person $\leq$ Man, e.g., denotes that the concept *Man* is a specialization of the concept *Person*, and will therefore appear in the offspring of the latter within the lattice organizing these concepts. In our model, CGs are used to represent the image content at the conceptual level.

### 4.1 Representation of the visual semantics facet

An instance of the visual semantics facet is represented by a set of CGs, each one containing an *Io* concept linked through the conceptual relation *is_a* to a semantic concept: [Io]→(is_a)→[$c_{sem}$[i]]. E.g., graphs [Io1]→(is_a)→[People] and [Io2]→(is_a)→[Water] are the representation of the visual semantics facet in figure 4 and can be translated as: the first IO (Io1) is associated with the semantic concept *people* and the second IO (Io2) with the semantic concept *water*. We use WordNet to elaborate a visual ontology that reflects the is_a relation among the semantic concepts. They are organized within a multi-layered lattice ordered by a specific/generic partial order (a part of the lattice is given in figure 5).

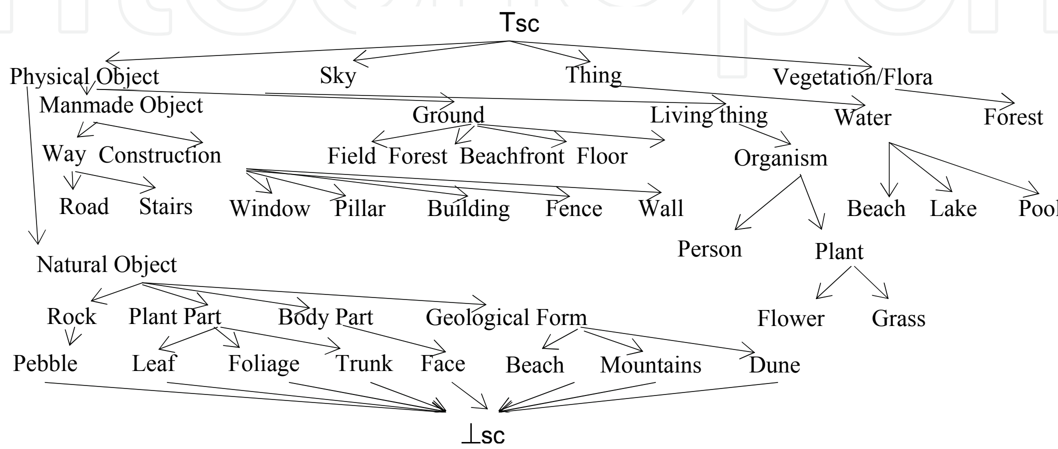



Figure 5. Lattice organizing semantic concepts

We now focus on the relational facet by first proposing structures for the integration of relational information within our strongly-integrated framework and then specifying their representation in terms of CGs.

### 4.2 Conceptual representation of the relational facet

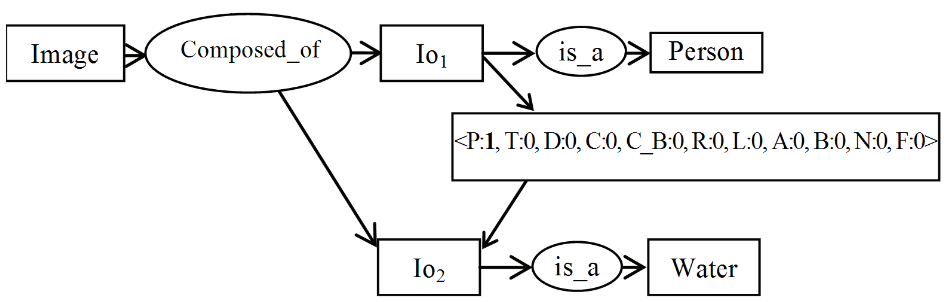
Each pair of IOs are related through an index spatial meta-relation (ISR), compact structure summarizing spatial relationships between these IOs. ISRs are supported by a vector structure **Sp** with eleven elements corresponding to the previously explicited spatial relations. Values Sp[i], $i \in [1,11]$ are booleans stressing that the spatial relation $s_i$ links the two considered IOs. E.g., the first and second IOs (Io2) respectively corresponding to semantic concepts *person* and *water* in figure 4 are related by the ISR <P:**1**, T:0, D:0, C:0, C_B:0, R:0, L:0, A:0, B:0, N:0, F:0>, which is translated by Io1 being **inside** (**part of**) Io2.

Our framework proposes an expressive query language which integrates visual semantics and symbolic spatial characterization through boolean operators. A query which associates visual semantics with a boolean disjunction of spatial relations such as **Q**: "Find images with people **at the left** OR **at the right** of buildings" can therefore be processed (user-formulated queries are studied in [Belkhatir 05b]). *Or* spatial concepts (OSCs) are conceptual structures semantically linked to the disjunction boolean operator and specified for the processing of such a query. They are supported by the vector structure $\mathbf{Sp_{or}}$ such that $Sp_{or}(i)$, $i \in [1,11]$, is a non-null boolean value if the spatial relation $s_i$ is mentioned in the disjunction of spatial relations within the query. The OSR <P:0, T:0, D:0, C:0, C_B:0, R:**1**, L:**1**, A:0, B:0, N:0, F:0>$_{OR}$ corresponds to the spatial characterization expressed in Q.

In our conceptual representation of the spatial facet, spatial meta-relations are elements of partially-ordered lattices organized with respect to the type of the query processed. There are two types of basic graphs controlling the generation of all the relational facet graphs. **Index spatial graphs** link two IOs through an ISR: **[Io1]→(ISR)→[Io2]**. **Query spatial graphs** link two IOs through *And*, *Or* or *Not* spatial meta-relations **[Io1]→(ASR)→[Io2]**; **[Io1]→(OSR)→[Io2]** and **[Io1]→(NSR)→[Io2]**. Eg, the index spatial graph [Io1]→(<P:**1**, T:0, D:0, C:0, C_B:0, R:0, L:0, A:0, B:0, N:0, F:0>)→[Io2] is the index representation of the spatial facet in figure 4 and is interpreted as the first IO (Io1) is related to the second IO (Io2) through the ISR <P:**1**, T:0, D:0, C:0, C_B:0, R:0, L:0, A:0, B:0, N:0, F:0>. The query spatial graph [Io1]→(<P:0, T:0, D:0, C:0, C_B:0, R:**1**, L:**1**, A:0, B:0, N:0, F:0>$_{OR}$)→[Io2] is the representation of query Q.

### 4.3 Image index and query representations

Image index and query representations are obtained through the combination (**join** operation [Sowa 84]) of CGs over the visual semantics and relational facets. We propose the graph unifying all visual semantics and spatial CG representations of the image proposed in figure 4:




## 5. A relevance feedback framework strongly integrating visual semantics and relational descriptions

We present a RF framework enhancing the state-of-the-art techniques as far as two major issues are concerned. First, while most image RF architectures are designed to deal with global image features, our framework operates at the IO level and the user is therefore able to select visual entities of interest to refine his search. Moreover, the user has a total control of the query process since the system displays the query generated from the images he selects and allows its modification before processing.

### 5.1 Use case scenario

Our RF framework operates on the whole corpus or on a subset of images displayed after an initial query image was proposed. The user refines his search by selecting IOs of interest. In case the user wants to refine the spatial characterization between a pair of visual entities (e.g. the user is interested in retrieving people either inside, in front of or at the right of a water area), he first queries with the semantic concepts corresponding to these entities (here '*water* and *people*') and then enrich his characterization through RF. The system translates the phrase query 'water and people' in a visual semantics graph:

[Image]→(composed_of)→[$Io_1$]→(is_a)→[water]
→ [$Io_2$]→(is_a)→[people]

The latter is processed and the results are given in figure 6.


Figure 6. First retrieval for the query "water and people"

When the RF mode is chosen, the system displays all IOs within images relevant to the query 'water and people'. The user chooses to highlight 3 pairs of IOs (figure 7) within displayed images which are relevant to his need (i.e. present the specific visual semantic and spatial characterizations he is interested in).

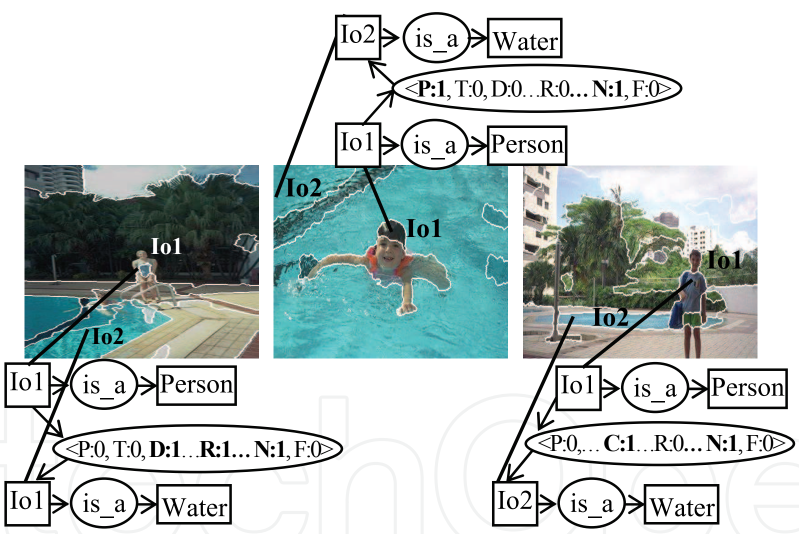
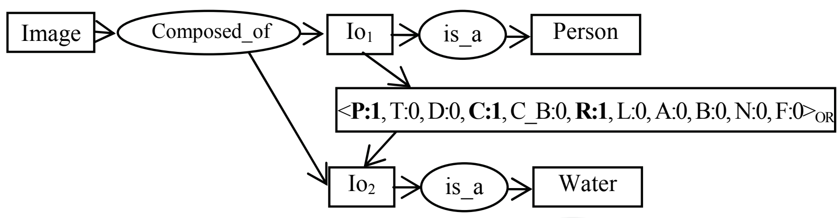



Figure 7. Selected IOs and their conceptual representation

The system is then expected to generate a generalized and accurate representation of the user's need from the conceptual information conveyed by the selected IOs.
According to the user's selection, the system should find out that the user focuses on images containing a person either being inside, in front of or at the right of water. Our RF framework therefore processes the ISRs of the selected pairs of IOs so as to construct the OSR <**P:1**, T:0, D:0, **C:1**, C_B:0, **R:1**, L:0, A:0, B:0, N:0, F:0>$_{OR}$. The spatial query graph [Io1]→[<**P:1**, T:0, D:0, **C:1**, C_B:0, **R:1**, L:0, A:0, B:0, N:0, F:0>$_{OR}$]→[Io2] is then generated. Finally, visual semantics and spatial query graphs are aggregated to build the full query graph:




### 5.2 Relevance feedback algorithm

The algorithm summarizing the RF mode is as follows:

Given a query with semantic concepts $SC_i$, **generate** a visual semantics graph $G_{sem}$.
**Process** the graph and **display** relevant images.
If the user selects the RF mode, **highlight** IOs then **take into account** the n pairs of IOs selected by the user.
<u>Regarding the spatial subfacet</u>
The n selected pairs of IOs are characterized by n ISRs supported by vector structures $[Sp]_k$ ($k \in [1,n]$) such that values $[Sp(i)]_k$, $i \in [1,11]$ are booleans stressing that in the $k^{th}$ ISR the spatial relation $s_i$ links the considered pair of IOs.
**Generate four *Or* spatial relations** respectively corresponding to the topological relations, the right/left and above/below directional relations and finally the metric relations considering the n ISRs (let us note that we generate an OSR for each group of relations which are said incompatible, i.e. one IO cannot be both at the left and at the right of an other IO, also one IO cannot be both near and far from an other IO etc…). These OSRs are supported by vector structures $[Sp_{OR}]_j(i)$, $j \in [1,4]$ , $i \in [1,11]$ such that:

- $[Sp_{OR}]_1(i)$ is a boolean value equal to 1 if a topological relation $s_i$ ($i \in [1,5]$) relates the IOs in one of the n pairs selected by the user and all other boolean values are null ($[Sp_{OR}]_1(i)=0 \ \forall \ i \in [6,11]$).
- $[Sp_{OR}]_2(i)$ is a boolean value equal to 1 if a directional relation right/left $s_i$ (i=6 or i=7) relates the IOs in one of the pairs selected by the user and all other boolean values are null.
- $[Sp_{OR}]_3(i)$ is a boolean value equal to 1 if a directional relation above/below $s_i$ (i=8 or i=9) relates the IOs in one of the pairs selected by the user and all other boolean values are null.
- $[Sp_{OR}]_4(i)$ is a boolean value equal to 1 if a metric relation $s_i$ (i=10 or i=11) relates the IOs in one of the pairs selected by the user and all other boolean values are null.

**Generate the respective *Or* query graphs** $G_{spa\_k}$: $[IO] \rightarrow (<[Sp_{OR}]_j(i)>) \rightarrow [IO]$, $j \in [1,4]$, $i \in [1,11]$
**Aggregate** (join operation [Sowa 84]) CGs $G_{spa\_1}$, $G_{spa\_2}$, $G_{spa\_3}$ and $G_{spa\_4}$ to generate the spatial query graph $G_{spa}$.
**Aggregate** (join operation) visual semantics and spatial query graphs $G_{sem}$ and $G_{spa}$. Each query (like document index representations) is indeed represented by a global CG resulting from the aggregation of CGs over the visual semantics and relational facets called **image query graph**.

### 5.3 Matching query and index structures

**The Projection Operator**. An operational model of image retrieval based on the CG formalism uses the graph projection operation for the comparison of an image query graph and an image 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 a projection of an image query graph $I_Q$ within an image document graph $I_D$ exists then the image document indexed by $I_D$ is relevant for the image query $I_Q$.

Formally, the projection operation $\wp : I_Q \rightarrow I_D$ exists if there is a sub-graph of $I_D$ 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 $I_Q$, there is the same relation between the two concepts $c_{D1}$ and $c_{D2}$ of $I_D$, such as $\wp(c_{Q1}) = c_{D1}$ and $\wp(c_{Q2}) = c_{D2}$.

However, brute-force implementations of the projection would result in exponential execution times. Based on the work in [Ounis&Pasca 98], we use an adaptation of the inverted file approach for image retrieval. We specify lookup tables associating visual semantics concepts to the set of image documents whose index contain it. Treatments that are part of the projection are performed during indexing following a specific organization of CGs which does not affect the expressiveness of the formalism. Moreover, lattices organizing spatial relations are defined by mathematical partial orders and not hard-coded, which allows fast query processing. We discuss in the next section the organization of the lattice for processing queries with OSMs.

**Processing queries with OSMs.** ISRs are organized within an *Or* lattice to process a query conveying a boolean disjunction of spatial relations such as “Find images with people at the left **or** at the right of buildings”. This query is first translated in its graph representation (cf. section 4.2). Semantic concepts *huts* and *grass* are processed by the lattice of semantic concepts. The link between the generated OSR <P:0, T:0, D:0, C:0, C_B:0, **R:1**, **L:1**, T:0, B:0, N:0, F:0>$_{OR}$ and its equivalent ISR is not straightforward. A new category of meta-relations eliciting this link by taking into account **dominant** spatial relations (i.e. spatial relations mentioned in a query as they have a higher importance in the ordering process of ISRs within the lattice, other spatial relations are called **secondary**) shall be introduced. These concepts are index spatial meta-relations **with dominant** $\mathbf{d_{OR}}$, where $d_{OR}$ is the set of dominant spatial relations. They are supported by a vector structure $s_d$ with eleven elements corresponding to spatial relations $s_i$. Values $s_d[i]_{i\in[11]}$ such that $s_i \in d_{OR}$ characterize the presence of dominant spatial relations and values $s_d[j]_{j\in[1,11]}$ such that $j \neq i$, the presence of secondary spatial relations within the spatial characterization of the considered IOs. Index spatial meta-relations **with dominant** $\mathbf{d_{OR}}$ are specializations of OSRs and generalizations of ISRs as far as the lattice organization is concerned. The OSR <B:0…D:1,I:0…U:1…>$_{OR}$ is related to its equivalent ISR with dominant {left, right}: <P:0, T:0, D:0, C:0, C_B:0, **R:1**, **L:1**, T:0, B:0, N:0, F:0> as highlighted in the lattice of figure 8. As a matter of fact, the most relevant images provided by the system present people at the left or at the right of buildings, i.e. people and buildings related through only dominant spatial relations. This symbolic spatial characterization is represented by the highlighted ISR (*sr*) in figure 8. Other images are composed of people either at the right or at the left of buildings with at least one additional spatial relation not mentioned in the query linking the two semantic concepts. In the lattice, ISRs representing such characterizations are descendants of *sr*. Formally, sub-lattices of index spatial meta-relations with dominant $d_{OR}$ are partially ordered by $\leq_{OR}$:

$$\forall a,b \text{ index spatial meta-relations with dominant } d_{OR},\ a \leq_{OR} b \Leftrightarrow (a=\perp_{OR}) \vee (b=\top_{OR}) \vee [(\forall i\in[1,11], s_i\in d_{OR}, (a_{[i]} = 0 \wedge b_{[i]} = 1) \vee (a_{[i]} = 1 \wedge b_{[i]} = 1) ) \wedge (\forall j\in[1,11], s_j \notin d_{OR}, (b_{[i]} = 0 \wedge a_{[i]} = 1) \vee (b_{[i]} = 1 \wedge a_{[i]} = 1))]$$

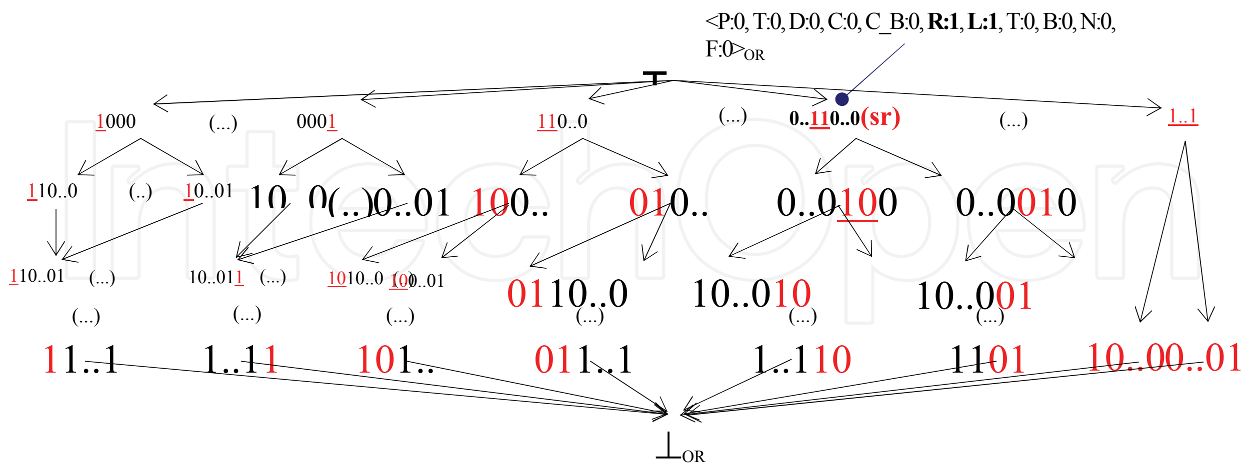



Figure 8. Lattice Processing *Or* Spatial Meta-relations

## 7. Conclusion

We have specified within the scope of this paper a framework combining semantics and relational (spatial) characterizations within a coupled architecture in order to address the semantic gap.

This framework is instantiated by an operational model based on a sound logic-based formalism, allowing to define a representation for image documents and a matching function to compare index and query structures.

We have specified a query framework coupling keyword-based querying with a relevance feedback module managing transparent and penetrable interactions by considering conceptual characterizations of images.

The choice of conceptual graphs as an operational model is the most natural in the sense that it holds several advantages in our application context. It indeed allows the symbolic representation of all components of a multimedia indexing and retrieval architecture: queries, index documents and matching function. Moreover its simple representation is particularly well-suited for user interaction in the framework of relevance feedback.

To stress the relevance of our approach, the theoretical contributions of this paper in the domain of image indexing and retrieval are summarized below:

- We have first proposed a neural-network based architecture for the highlighting of image objects, structures abstracting the image visual entites, and the characterization of their associated semantics.
- In the perspective of unifying the semantic and relational characterizations, we have proposed an integrated model featuring a bi-facetted organization. The visual semantics facet describes the image semantic content and is based on labeling IOs with a semantic concept. The relational facet is itself based on the relational (spatial) characterizations between pairs of image objects obtained after highlighting a correspondence process between extracted low-level information and symbolic relations.

- To overcome the limitations of the keyword-based approach to query on the image content, we have proposed a high-level relevance feedback framework, allowing in particular the relational characterization of the image objects.
- We have finally proposed a correspondence model based on the conceptual graph projection operator. Its instantiation is optimized through the use of specific data structures to boost retrieval. In particular, semantic and spatial index structures are organized in lattices defined by mathematical partial orders.

## 8. References


Belkhatir, M. et al. (2004). Integrating perceptual signal features within a multi-facetted conceptual model for automatic image retrieval, *ECIR*, pp. 267-282

Belkhatir, M. (2005). Combining semantics and texture characterizations for precision-oriented automatic image retrieval, *ECIR*, pp. 457-474

Belkhatir, M. et al.: A full-text framework for the image retrieval signal/semantic integration, *DEXA* (2005), pp. 113-123

Egenhofer, M. et al. (1991). Reasoning about binary topological relations, *SSD*, 143-160

Hollink, L. et al. (2004). Classification of user image descriptions. *Int. J. Hum.-Comput. Stud.* 61(5), pp. 601-626

Lu, Y. et al. (2000). A unified framework for semantics and feature based relevance feedback in image retrieval systems. *ACM MM*, pp. 31-37

Mechkour, M. (1995). EMIR$^2$: An Extended Model for Image Representation and Retrieval, *DEXA*, pp. 395-404

Meghini, C. et al. (2001). A model of multimedia information retrieval, *J. ACM* 48(5), pp. 909-970

Mojsilovic, A. & Rogowitz, B. (2001). Capturing image semantics with low-level descriptors, *ICIP*, pp. 18-21

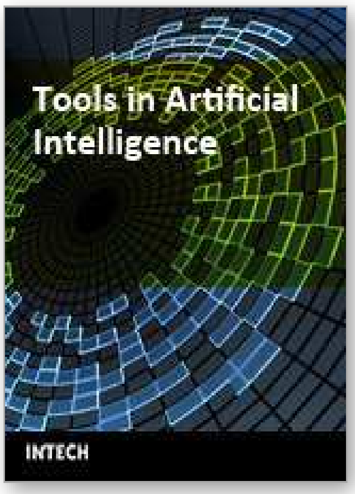
Naphade, M. et al. (2006). A Large-Scale Concept Ontology for Multimedia, IEEE MultiMedia 13(3), pp. 86-91

Ounis, I. & Pasca, M. (1998): RELIEF: Combining expressiveness and rapidity into a single system. *ACM SIGIR*, pp. 266-274

Smeulders, A.W.M. et al. (2000). Content-based image retrieval at the end of the early years. *IEEE PAMI* 22(12), pp. 1349-1380

Sowa, J.F. (1984). Conceptual structures: information processing in mind and machine, Addison-Wesley publishing company

Zhou, X.S. & Huang, T.S. (2002). Unifying Keywords and Visual Contents in Image Retrieval. *IEEE Multimedia* 9(2), pp. 23-33




**Tools in Artificial Intelligence**
Edited by Paula Fritzsche

ISBN 978-953-7619-03-9
Hard cover, 488 pages
**Publisher** InTech
**Published online** 01, August, 2008
**Published in print edition** August, 2008

This book offers in 27 chapters a collection of all the technical aspects of specifying, developing, and evaluating the theoretical underpinnings and applied mechanisms of AI tools. Topics covered include neural networks, fuzzy controls, decision trees, rule-based systems, data mining, genetic algorithm and agent systems, among many others. The goal of this book is to show some potential applications and give a partial picture of the current state-of-the-art of AI. Also, it is useful to inspire some future research ideas by identifying potential research directions. It is dedicated to students, researchers and practitioners in this area or in related fields.



INTECH
open science | open minds

**InTech Europe**
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

**InTech China**
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821