Multimed Tools Appl (2013) 62:479–505
DOI 10.1007/s11042-011-0911-1

Improving image retrieval by using spatial relations

Carlos Arturo Hernández-Gracidas · Luis Enrique Sucar · Manuel Montes-y-Gómez

Published online: 11 November 2011

Abstract In this paper we proposed the use of spatial relations as a way of improving annotation-based image retrieval. We analyzed different types of spatial relations and selected the most adequate ones for image retrieval. We developed an image comparison and retrieval method based on conceptual graphs, which incorporates spatial relations. Additionally, we proposed an alternative term-weighting scheme and explored the use of more than one sample image for retrieval using several late fusion techniques. Our methods were evaluated with a rich and complex image dataset, based on the 39 topics developed for the ImageCLEF 2008 photo retrieval task. Results show that: (i) incorporating spatial relations produces a significant increase in performance, (ii) the label weighting scheme we proposed obtains better results than other traditional schemes, and (iii) the combination of several sample images using late fusion produces an additional improvement in retrieval according to several metrics.

Keywords Image retrieval · Spatial relations · Conceptual graphs

1 Introduction

Image retrieval consists of searching an image database in order to find those images that satisfy the needs of a user. This is a complex problem not yet completely solved,

C. A. Hernández-Gracidas (✉) · L. E. Sucar · M. Montes-y-Gómez
National Institute of Astrophysics, Optics and Electronics,
Luis Enrique Erro No. 1, Tonantzintla, Puebla, Mexico
e-mail: carloshg@ccc.inaoep.mx

L. E. Sucar
e-mail: esucar@inaoep.mx

M. Montes-y-Gómez
e-mail: mmontesg@inaoep.mx

given the difficulty in mapping a query, which may be expressed in terms of keywords or image samples, to the visual features of an image.

In this paper we explore the use of spatial relations as a way of improving image retrieval. We perform a study of the different types of relations and select those we consider as the most adequate ones for their application to image retrieval. We developed an image comparison and retrieval method based on conceptual graphs and spatial relations. High-level information is included in both processes in the form of the spatial relations among the objects detected in the image. These relations help to better represent the contents of the image and its structure.

We evaluate the relevance of spatial relations with respect to labels for retrieval. We perform improvements in label weighting by using label frequency in the database and the images individually, as well as information redundancy among sample images and textual description of the retrieval topic. Besides, we give evidence that the availability of several sample images for a topic helps to improve retrieval. We evaluated the use of spatial relations in retrieval using the topics from the ImageCLEF competition [1]. Results obtained give evidence of the usefulness of spatial relations, showing significant improvements in content-based image retrieval (CBIR).

This paper includes results and some methods developed in some of our previous research making use of spatial relations for image retrieval. Although most of the methods in this research have already been published in workshop and congress papers, we focused on compiling these methods and performing in-depth experiments to analyze the potential and feasibility of combining them.

1.1 The problem

Automatic image retrieval can be basically performed in one of two ways [9, 16, 35]: by using text related to the images or by using the image contents. Using text is referred as text-based image retrieval (TBIR), and it is currently the approach with the highest retrieval rates, but seriously limited by the need of manually added text. On the other hand, using image contents for retrieval is known as content-based image retrieval (CBIR) [32, 35]. A variation of CBIR is query by example (QBE), where the tendency is to use a set of images (which could be just one image) to obtain low-level image features. These features are expected to be useful to describe the general idea behind a search topic. A particular case of QBE, known as annotation-based image retrieval (ABIR), consists of using the same low-level features to try to identify objects in the image, and then to associate these objects to a label. In ABIR retrieval is based on the set of labels assigned to the images to be retrieved. Labels are compared and the more labels in common with the query image, the better that image is positioned in the retrieval list. As we can see, most CBIR methods are based solely on low-level visual features (color, texture, shape, etc.), however, their main drawback is that they tend to be confused at the moment of distinguishing between two visually similar, though conceptually different objects, which consequently causes them to obtain erroneous results. This is actually part of a well known problem called "semantic gap" [26].

Given that most state of the art retrieval methods using image contents are based on low-level features, textual information or direct human interaction, ABIR is still an open problem, and the use of spatial relations could be useful for adding complementary information to ABIR. Spatial information is directly linked to the objects of interest in the image, and how they interact, both, in the scene and among each other. Spatial relations are useful to know the position of an object using other objects as reference and providing with high-level information with respect to the image contents. These relations provide with a context for annotations, helping to compare images by their structure (consisting of labels and spatial relations together) and reducing the semantic gap. Experimental results give evidence that both, modeling and application of spatial relations to improve ABIR, provide better results than the use of traditional methods based on low-level information extracted from the images.

1.2 Motivation

In this paper we focus on the application of knowledge derived from spatial relations to improve ABIR. We emphasize the fact that spatial relations should not be employed as the exclusive source of information for improving retrieval; however, their use as a complementary source of information along with other sources provides relevant knowledge to the processes of image comparison and retrieval. Two main issues can be identified in relation to the use of spatial relations for ABIR:

1. It is necessary to determine which spatial relations are possible to be retrieved depending on the domain of application, as well as determining which relations are useful to be applied for solving problems in ABIR.
2. It is also fundamental to clarify how to model the set of spatial relations among the objects in a scene and how to use these relations for the important problems of image comparison and retrieval.

These issues will be solved in this work by carefully selecting the most adequate set of spatial relations for ABIR and by using a spatio-conceptual image representation, suitable for image comparison and retrieval.

1.3 Contributions

The method we propose consists of using a set of region-level annotated images for image retrieval. These regions are used to obtain spatial relations in the image itself. The combination of image annotations and spatial relations is used to compare sample images to each of the images in a database, obtaining a list of similarities. Once we have this list, we retrieve the k most similar images and are able to answer the query. The most relevant contributions of our work are the following:

– The analysis of topological relations and order relations, and the selection and validation of the set of spatial relations adequate for image retrieval.
– An image retrieval method combining concepts and spatial relations, which is based on conceptual graphs. This method is intended to use high-level information acquired by annotating images, and use this information to better answer image retrieval.

- A label weighting method, called $MTFIDF$. This is an adaption of $TFIDF$, a frequently used term-weighting method, which in its original form performed poorly. The modified method prioritizes those labels co-occurring with the smallest amount of labels in an image.
- A label weighting method called topic-specific weighting, using late fusion to take advantage of the availability of several sample images. This information is complemented by the use of the textual description of the retrieval topic and this weighting is intended to consider terms as more relevant when they are frequent along the sample images and textual description of each specific topic.
- An extensive experimental evaluation of the proposed methods for improving image retrieval by incorporating spatial relations, label weighting and combining several sample images. By evaluating several scenarios we intend to provide evidence on the advantages of each individual method, and the combination of them.

2 Related work

Several models for representing both, topological and order relations, have been proposed in the literature. Some of the most important methods for topological relations are the 4-intersection model [13], the 9-intersection model [12], the Voronoi-based 9-intersection model [8] and the model based on the Euler number [41]. Models for representing order relations have also been proposed, such as symbolic projections [7], cardinal directions represented as cones or defined by projections [15] and the direction-relation matrix [17]. These methods are not specifically intended to be used in CBIR, and their usefulness has not been extensively evaluated in such a field.

There are several previous works that incorporate spatial relations for image retrieval. In [34], they introduced a deductive system intended to extend text-based image retrieval systems. Using an initial set of relations, together with a set of rules that are applied, it allows to derive additional spatial relations. This system is shown to be complete in 3D spaces, but incomplete in 2D spaces, which represents an important limitations when working with images. Although this is an interesting and feasible idea, this work does not suggest a way to perform several related processes, which limits the appliability of their method.

An image retrieval system using spatial information as a complementary element is presented in [30]. This retrieval system works on the WWW using a web crawler, which employs textual information obtained from the retrieved web pages, as well as the image names. This information is complemented by low-level features such as color, and high-level features such as spatial relations. This system represents image contents by means of a graph and image similarity is measured in terms of graph isomorphism, object similarity, object position similarity, topological similarity and distance similarity. Human interaction is needed for the object recognition and annotation to be performed, and image search can be performed based on hand sketches or sample images. One disadvantage of this method is the need of human interaction, which limits its appliability. Image queries are simple and more related to the object recognition task, which makes it difficult to determine how well the method works and the usefulness of spatial relations for image retrieval.

In [31] they present another retrieval system which adds an extra spatial relation coding in the retrieval process. Their model includes 6 spatial relations: *left, right, up, down, touch* and *front*. Similarity is measured using knowledge about the objects in the image and their spatial relations. Automatic segmentation and annotation methods are employed for that purpose, and retrieval is based on a sample image and relevance feedback. One disadvantage is that the experiments are performed on a limited set of labels and images, and tests on more complex and realistic label sets are not provided. The set of relevant images is determined according to the image contents, which gives no evidence as whether the method will work with more elaborate topics.

More theoretical work on analyzing properties and problems regarding the use of CBIR and spatial relations is presented in [40]. There they verify the consistency of spatial relations and analyze certain information intractability issues.

From the approach we are following to image retrieval, a very relevant step is to identify concepts in the image, which is known as region-level image annotation. Some frequently used methods for automatically performing this kind of image annotation are [6, 10, 29]. Particularly, some using of spatial relations for image annotation has been tried at [39], with promising results; there they define 4 order relations as neighborhoods and divide the image into a grid, which is later matched using automatic image segmentation.

Some of the main disadvantages these methods have are: (i) Human interaction is needed, which limits their appliability to real problems; (ii) Experiments are simplified, since reduced sets of labels and images are employed; (iii) Image retrieval is evaluated based on a concept detection task; (iv) Finally, although these methods use spatial relations for image retrieval, performance is not compared with and without adding spatial relations, which makes unclear how good their approach for adding spatial relations was.

In our research we determine the adequate set of spatial relations and determine their relevance for image retrieval. We experiment with a complex image set and a considerable number of labels. Finally, we consider several sample images to perform image retrieval, through a late fusion mechanisms.

Although the experiments in this paper are based on manually segmented and annotated images, we understand the unfeasible manual processing becomes when image resources grow. Even more, we suggest in the future segmentation and annotation methods will be robust enough to allow for a completely automatic processing of the image, which will allow for the application of our methods to any image database.

An example of the use of graphs (non-conceptual) for these tasks is [4], where they use an image representation based on graphs, even using spatial relations for image retrieval. The image comparison method derives from [5] and the evaluation is performed on a set of painting images.

Several semantic representations have been explored previously for general information and image retrieval, and particularly conceptual graphs have been used several times. An example of this is [3], where a multifacetic image representation is used for indexing and retrieval. There is another example of the use of conceptual graphs for image modeling and retrieval in [25], where they even take into account spatial relations to have a better representation.

3 Determining the set of spatial relations

3.1 Spatial relations

Spatial relations are those relations that can be determined for an object with respect to another (which is known as the reference). These relations give information about the relative position of the object of interest in the scene. Frequently, spatial relations are determined in a binary fashion, however, certain spatial relations, such as: *among*, *surrounded by*, *closer to*, among others, are better understood when they are defined with respect to more than one object of reference. Three basic types of spatial relations [11] are the ones that are more frequently used, and they are:

1. Topological relations. These relations are determined for two objects and are preserved even if topological transformations, such as translation, rotation and scaling, are applied. To apply these relations to image retrieval, we must consider relations between two surfaces (bidimensional objects also called regions) in a bidimensional space (the image itself). Figure 1 left shows the 8 possible topological relations between two surfaces in a bidimensional space.
2. Order or direction relations. These are based on the definition of order and represent information regarding the position of an object with respect to another. This kind of relation is variable to rotations but is preserved to scaling and translation. In Fig. 1 right we show the possible order relations between a pair of surfaces for a bidimensional space.
3. Metric relations. These use measurements such as distance and direction. This kind of spatial relation is affected by scaling but not by rotation or translation. *Two kilometers far* is an example of a metric relation.

We focus on the analysis of the spatial relations and the determination of an adequate set, useful for ABIR. We have chosen topological and order relations, considering that both of them can provide relevant and complementary information. However, not all of these relations can be observed in an image, and even from those which can be observed, they are not equally relevant for our purpose. This analysis is developed next.

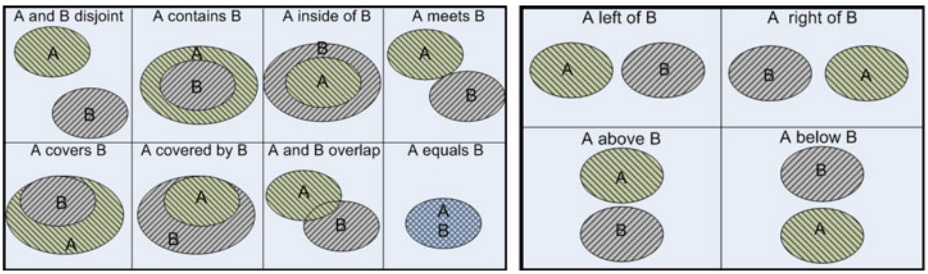


Fig. 1 *Left*: topological relations between surfaces in a bidimensional space. *Right*: order relations between surfaces in a bidimensional space

3.2 Selecting the relevant spatial relations

3.2.1 Topological relations

Given that topological relations have been modeled in several previous works [30, 31, 34] with favorable results, we consider their inclusion as an important aspect. There are, however, inconvenients which oblige us to simplify them.

Certain ambiguities are introduced when a manual segmentation is performed due to different interpretations that could be given to the same image. Some factors such as scene knowledge and domain knowledge when interpreting the expected shape of an object could take to variations in the segmentation. In contrast, an automatic segmentation is expected to provide the same results under the same conditions (except when a non-deterministic algorithm is employed). Figure 2 shows an example of how the way spatial relations are determined is affected by the interpretation when a manual segmentation is carried out. Three different interpretations for the possible relation are provided. The three of them are feasible for a manual segmentation, while the last one would be the only one provided by an automatic segmentation. According to our analysis, we decided to discard some topological relations due to the following reasons:

- The spatial relation *equal* must be discarded, since in the image plane, an equality relation means that one object covers exactly another object with the same characteristics in the image. The problem with this relation is that the object that is covered cannot be perceived and the evaluation of such relation is not feasible.
- In the case of manual segmentations, the relations *contains*, *inside of*, *meets*, *covers*, *covered by* and *overlap* show ambiguities that could lead to certain

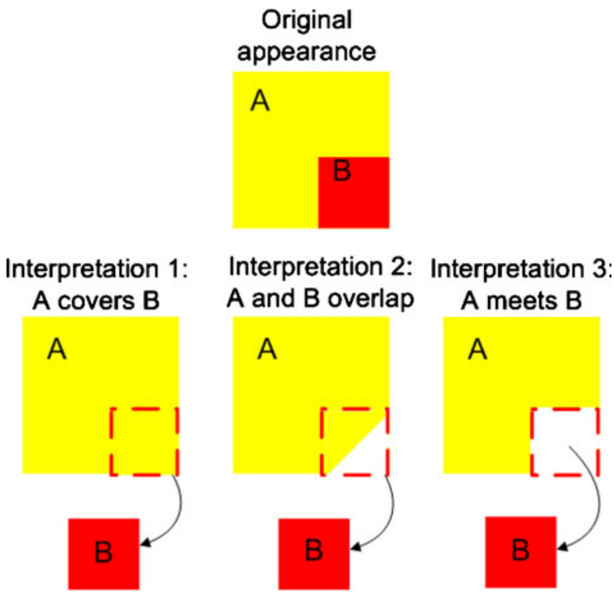


Fig. 2 Example of the ambiguities when an image is manually segmented. *Above*: the image as it originally appears. *Below*: three possible interpretations for this interaction (**A** *covers* **B**, **A** and **B** *overlapped* and **A** *meets* **B**). These three interpretations are feasible with a manual segmentation, while the last one would be the only one provided by an automatic segmentation

confusions when the segmentation is being carried out. Additionally, when the segmentation is automatic, one pixel cannot be assigned to more than one region, so the intersection between any pair of regions will always be empty (∅) and evaluating the relations *contains*, *inside of*, *covers*, *covered by* and *overlap* is not possible.

We consider that the best choice is to simplify the spatial relations to evaluate if there is any contact between a pair of regions or not, evaluating only two topological relations: (1) *Disjoint*. The same as in the original definition, it means that the intersection between the two regions equals ∅; (2) *Adjacent*. It means that there is some intersection between the two regions. We use the word *adjacent* to differentiate it from *meets*, given that although it is the same as *meets* for automatic segmentations, in the case of a manual segmentation, *adjacent* includes the concepts *contains*, *inside of*, *meets*, *covers*, *covered by* and *overlap*. With this simplification we only evaluate if the intersection between two regions is empty (*disjoint*) or not (*adjacent*), reducing the computational cost, and consequently the processing time. Also, there will always be a topological relation between a pair of regions, since any pair of regions is either *adjacent* or *disjoint*.

3.2.2 Order relations

Order relations are important for an adequate representation of image contents. Finding labels where there are contradictions is an advantage of the inclusion of these relations (for example, an image where a region labeled *sky* is below another region labeled *grass*, which is very unlikely). There is an inconvenient with the original set of order relations when we look at previous work, as well as the kind of images and segmentations. Next we provide a number of considerations about the needs for order relations in image retrieval. The set of order relations we obtain are summarized in Table 1.

1. Given the irregular shape of the segmented regions, a strict evaluation where every pixel of one region must be *above, below, left* or *right* of the other region, seems inadequate. For that reason, we decided to evaluate the position of the regions with respect to their center of mass.
2. In terms of order, we can see that two objects will always be related in two ways. For example, an **airplane** could be, at the same time, *above* and *left of* a **house**. For such reason we decided to divide order relations into horizontal and vertical

Table 1 The set of spatial relations used in this work and their classification as directed or undirected

		Relation	Type
Topological relations	1	Adjacent	Undirected
	2	Disjoint	Undirected
Order relations			
Horizontal relations	3	Beside (either left or right)	Undirected
	4	Horizontally aligned	Undirected
Vertical relations	5	Above	Directed
	6	Below	Directed
	7	Vertically aligned	Undirected

We divide them into three groups: topological, horizontal and vertical relations

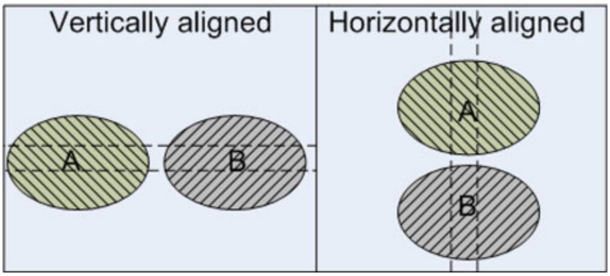


Fig. 3 Horizontal and vertical alignment relations. These relations are determined by drawing a *vertical* and a *horizontal stripe*, respectively, proportional to the image size. When the *center* of mass of two regions falls into the same stripe, they are said to be aligned

relations. This way we guarantee there will always be one and only one relation between a pair of regions in each group.

3. To minimize the consequences of the use of the center of mass for evaluating these relations, we introduce horizontal and vertical alignment relations, which help in cases when the distance between the center of mass of two regions is just a few pixels. In order to evaluate the alignment we draw a vertical stripe for horizontal alignment and a horizontal stripe for vertical alignment. The width of

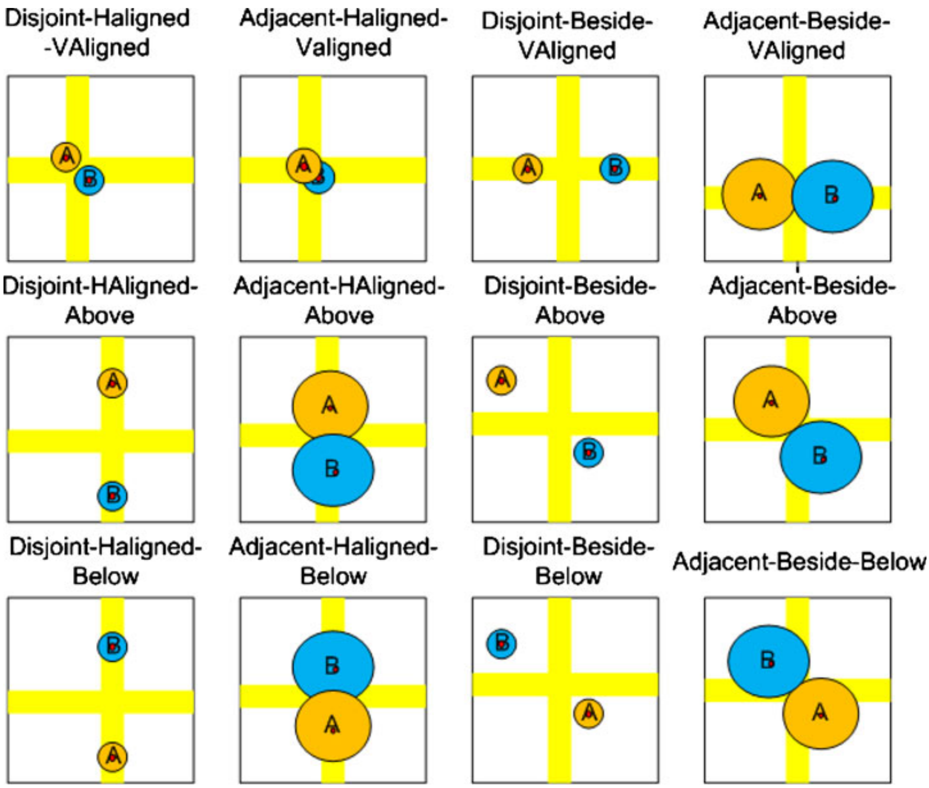


Fig. 4 Verification of the feasibility of the 12 combinations for the 7 spatial relations we selected. The circular regions (A, B) represent two segments in each image, and the *yellow* (*light gray*) *stripes* represent the alignment threshold for the regions

the stripe is proportional to the image size, and whenever two centers of mass fall into the same stripe, we say they are aligned. This provides the evaluation of order relations with more flexibility. Alignment relations are illustrated in Fig. 3.
4. Although in some domains *left of* and *right of* should be differentiated, in general we assume that this difference is not significant, so we group these two relations as *beside*.

3.3 Validation of the spatial relations

After our analysis we defined 7 spatial relations, divided into three independent groups, so there is one and only one relation in each group for each pair of regions, giving a total of three relations at a time between each pair of regions. An important step now is to verify if all of the combinations are feasible in a 2D image. Keeping in mind that we have 2 topological relations, 2 horizontal relations and 3 vertical relations, we have 12 combinations that must be evaluated in order to determine if there are any restrictions. Figure 4 shows examples where these combinations are present, so there was no case that turned out to be impossible.

4 Image retrieval incorporating spatial relations

For the objective we pursue, which is incorporating spatial relations to improve ABIR, we propose a three-step process:

1. Image representation. We use a simple and effective image representation, which captures relevant image features, while at the same time allows for a fast image comparison. This representation is based on the information provided by image annotations and spatial relations between pairs of annotated regions in each image.
2. Image comparison. Once we had the image representation, we developed a measure to compare how similar an image is to another, the similarity measure we developed was based on the work presented in [27]. This measure is based on evaluating two different kinds of similarity and observing the impact of each of them with respect to the other in the retrieval results.
3. Image retrieval. Image retrieval is achieved by using the afore mentioned image comparison measure. Based on this measure we rank the list of images and retrieve a sub-list which will be considered as the most similar images to a sample provided.

The three retrieval steps are explained with more detail in this section. Additionally, other techniques are considered to improve retrieval results. Two different approaches for label weighting are an important part of this paper. On the one hand, we compare different ways of considering label frequency; on the other hand, we give labels a specific weight for each topic. Taking advantage of the availability of several sample images for the set of topics developed for the ImageCLEF 2008, we were able to compare and evaluate a number of late fusion schemes, which is also a contribution of this work. We describe these weighting and fusion methods in this section as well.

4.1 Step 1: image representation

In [19] we introduced a method for the representation of images based on spatial relations and conceptual graphs (CGs) [36]. Spatial relations were divided into three groups, namely: topological relations, horizontal relations, and vertical relations, whose selection was performed following the methodology we explained in Section 3. In [19] we also made use of CGs to express the spatial relations among labels (objects) from an image. CGs are finite, connected, and bipartite graphs formed of two types of nodes: concepts (in our case labels) and relations (in our case spatial relations). Figure 5 shows an example of how the images in the database [14] are segmented and annotated, and how the spatial relations in the images can be represented by means of CGs. In accordance with Table 1, all of the relations were defined as undirected except for the vertical relations *above* and *below*.

4.2 Step 2: image comparison

The similarity between a pair of images is measured using two different similarity measures: conceptual similarity (S_c) and relational similarity (S_r). S_c measures how similar two graphs are by counting how many concepts (labels) they have in common, while S_r measures how similar the relations among the concepts in common are. For so doing, we used spatio-conceptual graphs [20]. Images are compared in two different fashions using the following equations:

1. Concepts, represented by the image labels are compared using the equation:

$$S_c = \frac{2n(G_c)}{n(G_1) + n(G_2)} \tag{1}$$

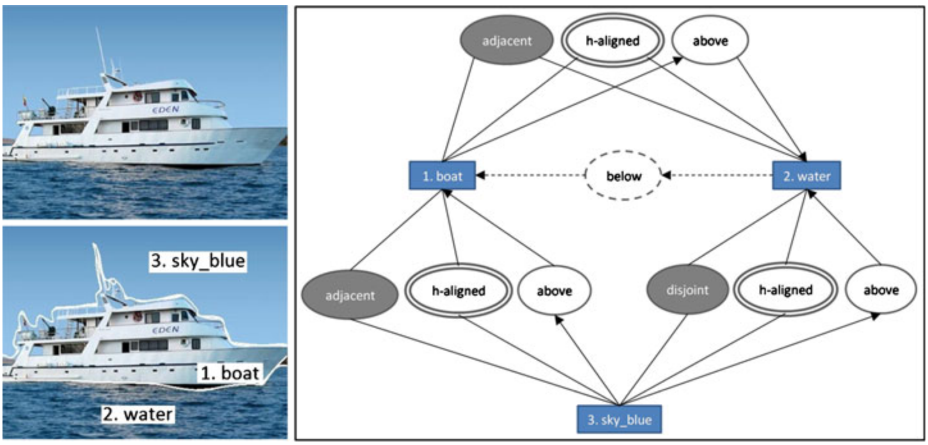


Fig. 5 *Top-left*: one of the images in the IAPR-TC12. *Bottom-left*: the same image, segmented and annotated. *Right*: conceptual graph indicating the spatial relations in the image. Topological relations are shown with *filled nodes*, horizontal relations appear with *double-lined border*, and vertical relations appear with *single-lined border*

where $n(G_c)$ is the number of concept nodes the two graphs have in common, and $n(G_1)$ and $n(G_2)$ are the number of concept nodes in graphs G_1 and G_2, respectively.

2. Spatial relations among the labels in common are also compared using the equation:

$$S_r = \frac{2m(G_{Tc}) + 2m(G_{Hc}) + 2m(G_{Vc})}{3m_{G_c}(G_1) + 3m_{G_c}(G_2)} \tag{2}$$

where S_r considers the three relational graphs, thus $m(G_{Tc})$, $m(G_{Hc})$ and $m(G_{Vc})$ represent the number of arcs (relations) in common between the two compared images, for topological, horizontal and vertical relations, respectively.

Conceptual similarity (S_c) measures the proportion of labels in common between the two images, with respect to the total number of labels in both images. Relational similarity (S_r) compares the number or edges in common between the two images, for topological, horizontal and vertical relations.

The similarity between two images is measured by S, which considers both S_c and S_r, giving each a weight depending on a constant α

$$S = \alpha S_c + (1 - \alpha) S_r \tag{3}$$

For two images to be compared, they have to be preprocessed by segmenting and annotating them. After this process is done, spatial relations are computed in order to build their CGs. Once we have the CG for both images, they can be compared based on (3).

Even though some of the methods we mention for image retrieval use conceptual graphs for their representation, the use of spatial relations does not seem to be completely taken advantage of, and relegated to be simple extra information mixed with several other (low-level) features to be used in retrieval. We, on the other hand, incorporate high-level image contents in the form of spatial relations to the image comparison process and using this information we expect to obtain better results for image retrieval.

4.3 Step 3: image retrieval

With the previous representation we have explored the use of spatial relations as high-level support information for representing and comparing images. To retrieve a set of images corresponding to a topic, the sample image is compared against all of the images in the database, obtaining a list of ranked images according to their similarity with the sample image, from which the top k are kept. Figure 6 presents a block diagram for this process.

4.4 Label weighting

In traditional image retrieval, term weighting is a commonly used tool to determine the most relevant elements, which is usually done by assigning such relevance with respect to the frequency with which a term appears. In ABIR a "term" could be represented by a label in the annotation vocabulary. $TFIDF$ (*term frequency-inverse document frequency*) [22] is one such term weighting method, which is directly related to our experiments. In the first stage of our experiments, we determined

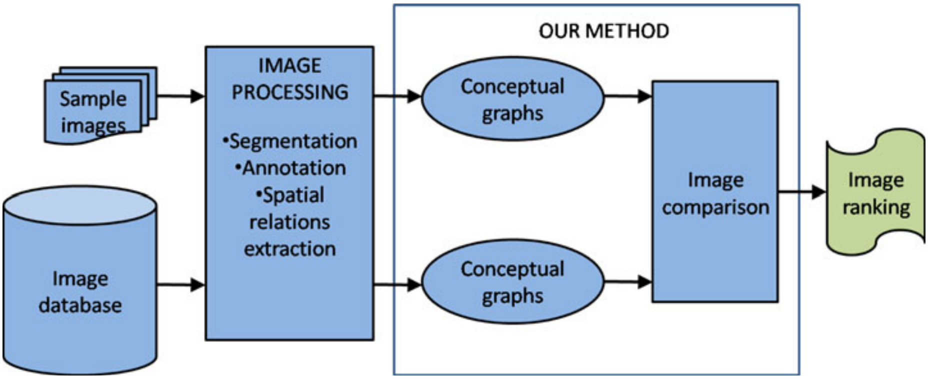


Fig. 6 Block diagram for the image retrieval method using spatio-conceptual graphs. For two images to be compared they must be segmented and annotated. Spatial relations are computed and the corresponding spatio-conceptual graphs are obtained. These graphs are compared using the similarity formulas we described and, by repeatedly applying this process between the sample image and each image in the database, a list of images, ranked according to their similarity, is obtained

which weighting approach is more adequate to consider the relevance of labels in retrieval. The weighting schemes are directly applied by first adding the weight of each of the k labels in the image (w_i), and then multiplying the accumulated weight by the similarity measure S, so we obtain $S_f = WS$ as the final similarity value.

$$W = \sum_{i=1}^{k} w_i \tag{4}$$

The three weighting schemes we consider are:

– **Uniform weights.** $w_i = 1$. Giving an equal and constant weight to each label. Results with this schema are provided as our baseline.
– **Inverse global frequency label weighting.** $w_i = \frac{1}{|\{I:l_i \in I\}|}$. Where $|\{I : l_i \in I\}|$ is the total number of images where label l_i appears. A simple scheme, similar to the IDF part in $TFIDF$, where each term is given a weight inverse to the number of times it appears along the image collection. The basic idea is that less frequent labels are considered to give more information than those common to a big amount of images.
– $TFIDF_{ij} = TF_{ij} \times IDF_i$ [22]. It is the traditional $TFIDF$ measure, where $TF_{ij} = \frac{n_{ij}}{N_j}$ (the occurrence of label l_i in image I_j is divided by the number of labels in image I_j), and $IDF_i = log\frac{|D|}{|\{I:l_i \in I\}|}$ (the number of images in the collection is divided by the number of images containing label l_i).

Additionally, and as a product of our observation of the behavior of these measures for term weighting, we use the label weighting method we proposed in [21]. This method consists of a modified version of the $TFIDF$ scheme. We call this modification $MTFIDF$, which is defined as:

$$MTFIDF_{ij} = TF_{ij} \times MIDF_i \tag{5}$$

In this case, TF_{ij} is the same as in $TFIDF_{ij}$, and $MIDF_i$ is defined as:

$$MIDF_i = log \frac{K}{\sum_{\{I:l_i \in I\}} N_i} \tag{6}$$

This is a modified version of IDF_i, where a constant (K) is divided by the sum of the number of labels in each image (N_i) where label l_i ($\{I : l_i \in I\}$) appears. In our experiments, the value of K is equal to the number of images in the database. For example, if label *tree* appears in 3 images and there were 4, 3 and 6 labels in each of them, respectively, the denominator in $MIDF_i$ would be $\sum_{\{I:l_i \in I\}} N_i = 4 + 3 + 6 = 13$. The idea behind this weighting scheme is that labels co-occurring with many other labels should be considered less relevant than those co-occurring with just a few labels.

4.5 Late fusion

In CBIR frequently just one sample image is used to represent a retrieval topic. However, an advantage frequently overlooked of having more than one sample image for the same topic is that it provides with more information that could be taken advantage of for retrieval. Late fusion, in the form of list fusion, could be performed by fusing lists obtained by executing different retrieval systems in order to combine them and obtain better results. Some possibilities for this kind of fusion are: *round robin*, *combSUM*, *combMNZ*, *Borda count*, *Condorcet*, *raw score value* (RSV) and *fuzzy Borda count* [24, 33, 37, 38].

Regarding the fusion methods we use in our experiments, we resorted to list addition [20], also known as linear combination of scores or SUM [28], and the maximum of the lists [20], or simply MAX [28], two basic methods with interesting results; *combMNZ* [24], which is considered as the baseline in different works; and *fuzzy Borda count* [2], given that it has provided better results than *combMNZ*.

5 Experiments

Our experiments are divided into three groups:

1. Experiments with a single sample image. We experimented with image retrieval using a single sample image. An important aspect of this work lays on determining if the spatio-conceptual representation is adequate for image retrieval. For so doing, we varied α in the evaluation of conceptual and relational similarity (3). In general, if results are better with α values smaller than 1, then we can infer that spatial relations are useful for image retrieval. At the same time, we evaluated the impact of the different term weighting schemes. We compared three existing term weighting methods (uniform weighting, used as the baseline; frequency-based weighting; and $TFIDF$), against a modified version of $TFIDF$ (called $MTFIDF$).
2. Experiments with multiple sample images. Given the fact that each retrieval topic is visually described by three sample images in the set of 39 topics designed for the ImageCLEF 2008, we also performed experiments using these three images and a late fusion scheme. We experimented with four different fusion methods: SUM, MAX, *combMNZ* and *fuzzy Borda count*. These methods are

compared to the average of the three individual retrievals, used as the baseline. For each of the sample images, an individual retrieval is performed and after this, the three retrieved lists are combined by using one of the mentioned fusion methods.

3. Topic-specific weighting (TSW). These experiments are related to fusing redundant information coming from the images and the textual description corresponding to a topic. The idea behind this weighting is that the information contained in the annotations given to the sample images and/or the textual description of the topic, could be useful to define a topic-specific label weighting. Although we include the textual description for this weighting, this method cannot be considered as text-based image retrieval, since text is only used to identify labels in the sample images but not to match these labels along the image collection. For example, if all of the sample images contain the labels *church* and *sky*, and just one of them contains the label *lamp* it is likely that *church* and *sky* are more relevant than *lamp* for that particular topic, without regard to their frequency along the whole image set. Figure 7 illustrates this. For TSW we perform the fusion of the labels in the sample images by using *combMNZ* and *fuzzy Borda*. Additionally, we can include in this fusion the labels contained in the textual description. To combine TSW with the best weighting scheme obtained in our experiments ($MTFIDF$) we try two simple alternatives, which are adding ($W = TSW + (MTFIDF)$) or multiplying

	2125		4947		8062
1	'church'	1	'sky-night'	1	'sky-night'
2	'tower'	2	'sky-night'	2	'street'
3	'tower'	3	'tree'	3	'street'
4	'sky-night'	4	'tower'	4	'couple-of-persons'
		5	'tower'	5	'church'
		6	'church'	6	'tower'
				7	'tower'
				8	'lamp'

Topic 15. Relevant images will show churches or cathedrals at night: the building is illuminated and/or the background is black. Images of cathedrals or churches during the day are not relevant. Other night shots without cathedrals are not relevant.

Fig. 7 Example of the use of TSW. *Top*: the three sample images for a topic, along with their labels. *Bottom*: the textual description for the same topic. The most frequent labels are highlighted with *green*, less frequent labels are highlighted with *yellow*, labels appearing just once are not highlighted at all

$(W = TSW \times (MTFIDF))$ both weights. These considerations provide us with eight variations in total.

5.1 Experimental setup

5.1.1 The database

The image database used in our experiments is the IAPR-TC12 [18], which was used for the ImageCLEF 2008 photo retrieval task [1]. This image set consists of 20,000 images of sports events, people, animals, cities and landscapes. Particularly, the manually segmented and annotated version of it, the SAIAPR-TC12 [14] was chosen for the experiments, given that it provides a reliable dataset that allows focusing more on image retrieval, than on the effects of automatic segmentation and/or annotation. However, it must be highlighted that our method is appliable to automatic segmentation and/or annotation as well. Some images from the SAIAPR-TC12 are shown in Fig. 8.

5.1.2 The topics

For an objective evaluation of our method, we resorted to the 39 topics developed for the ImageCLEF 2008 photo retrieval task [1]. The purpose of this task is to retrieve a set of relevant images from the whole image set, by using textual or visual information. Topics are expressed, for this reason, in both forms. In terms of text, a topic is expressed with a sentence in natural language. On the image side, three sample images are provided per topic. We base our work on the use of the visual part, discarding the textual information provided for the topics. The availability of three sample images per topic makes it a good alternative for testing data fusion,

Fig. 8 Sample images from the IAPR-TC12 collection

since we can perform a retrieval using each sample image and then fuse the results of the three retrievals to better characterize the topic.

For the ImageCLEF 2008, 39 retrieval topics are provided. In order to evaluate how accurate a retrieval is, the list of relevant images for each topic is provided. This, combined with a set of accuracy measures, gives a reliable parameter for comparing with other methods for image retrieval. Besides, given that these topics (along with the IAPR-TC12 image set) have already been used in other research, a comparison is possible.

5.1.3 Evaluation measures

In this paper we use the following evaluation terms:

- Precision (P) measures the fraction of the retrieved images that are considered as relevant, so P-20 measures the precision after the first 20 documents are retrieved.
- Recall (R) on the other hand is the fraction of the relevant images that were successfully retrieved. In our experiments, recall is computed over the total number of images retrieved, i.e., 1,000 images.
- Finally, the average precision (AP) combines P and R, to emphasize how early, relevant images are retrieved. MAP, in turn, describes the mean of the AP over the 39 topics.

We use MAP, P-20 and R computed over the 39 topics, to measure the performance of the methods we evaluated.

5.2 Experiments with a single sample image

In our experiments with a single sample image, we verified the usefulness of spatial relations by varying α in S, we also obtained evidence that $MTFIDF$ was the best label-weighting schema, being the one with the best results in most of the experiments. Figure 9 shows results for MAP, P-20 and R, respectively. Table 2 summarizes the same results. Assigning non-uniform weights to labels seems to be the most adequate solution, given that $MTFIDF$ is the scheme with the best performance, with a relative improvement of 58%, 32% and 20% with respect to the baseline for MAP, P-20 and R, respectively. We have noticed that relevant images in several topics are simple images, i.e., containing just a few objects. Given that $MTFIDF$ gives more weight to a label when there are just a few objects in the image, this kind of weighting favors the retrieval of such topics. On the other hand, the inverse global frequency label weighting is the one that takes more advantage of the use of spatial relations. While $TFIDF$ and $MTFIDF$ take into account the number of labels in the image, the inverse global frequency label weighting ignores the number of labels in the images, avoiding to reduce relevance to images where more objects, and consequently, more spatial relations, appear.

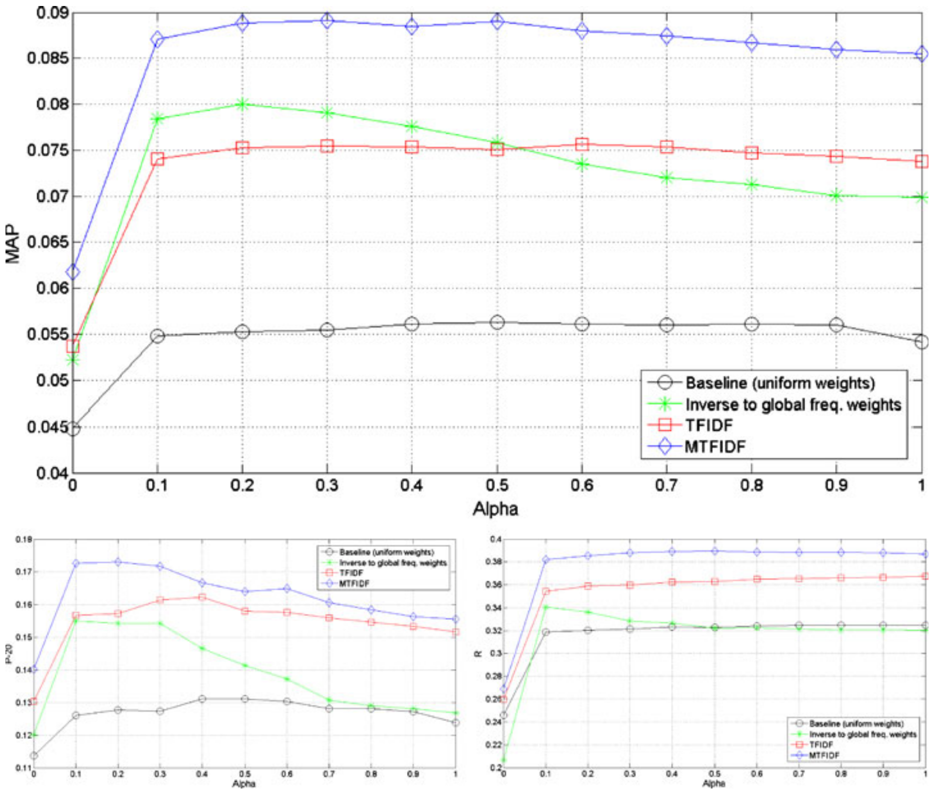


Fig. 9 Performance of retrieval with four different label weighting schemes, varying the weight of spatial relations and labels (α). *Top*: mean average precision (MAP). *Bottom-left*: precision at the first 20 images retrieved (P-20). *Bottom-right*: recall (R)

5.3 Experiments with multiple sample images

Our experiments show that combining several images through late fusion provides a rather significant increase in performance. Figure 10 shows these results for MAP, P-20 an R, respectively. Table 3 summarizes the same results. In general terms, SUM is the method that shows the best performance with a relative improvement with respect to the baseline of 49%, 39% and 34%, measured by MAP, P-20 and R, respectively. Low redundancy seems to affect SUM less than the other fusion methods given that the three fused lists contain the same kind of information and consequently, the direct addition takes advantage of the smallest existing redundancy, which is not done by the other methods. On the other hand, MAX is the method that takes the most advantage of the use of spatial relations, given that its basic idea is to always choose the highest similarity among the fused lists (remember that similarity is measured in conceptual and spatial terms).

Table 2 Summary of the results comparing different label weighting schemes (MAP, P-20 and R)

Method	$\alpha = 1$	Best $\alpha < 1$	Relative improvement (%)
MAP			
Uniform weights	0.0542	0.0563	3.87
Inverse global freq.	0.0699	0.0800	**14.45**
TFIDF	0.0738	0.0756	2.44
MTFIDF	0.0855	**0.0891**	4.21
Improvement with respect to the baseline			**58.26**
P-20			
Uniform weights	0.1239	0.1312	5.89
Inverse global freq.	0.1269	0.1551	**22.22**
TFIDF	0.1517	0.1624	7.05
MTFIDF	0.1556	**0.1731**	11.25
Improvement with respect to the baseline			**31.93**
R			
Uniform weights	0.3247	0.3249	0.0
Inverse global freq.	0.3201	0.3406	**2.05**
TFIDF	0.3675	0.3665	−0.27
MTFIDF	0.3869	**0.3894**	0.64
Improvement with respect to the baseline			**19.85**

The second column shows the result obtained when no spatial relations are used ($\alpha = 1$), the third column shows the best result when spatial relations are used ($\alpha < 1$) and the fourth column shows the relative improvement by using spatial relations. For each measure, we show in bold the best results. We also show in bold the highest relative improvement using spatial relations with respect to the baseline (uniform weights)

5.4 Topic-specific weighting

By considering label weighting, depending on each specific topic we obtained our highest retrieval rates. Results for these experiments are shown in Table 4. These results show that this kind of fusion is better for several cases than just considering a general term weighting method. By performing a label weighting depending on the topic contents we were able to improve retrieval, being the product of *combMNZ* using the textual description and $MTFIDF$ the most adequate combination. The relative improvement of using the textual description compared to using just the image is about 5%; of using *combMNZ* compared to *fuzzy Borda* is also about 5%; and of multiplying weights compared to adding them is about 18%. Redundancy seems to be increased from the use of the textual description, and *combMNZ* is a fusion method that takes more advantage of redundancy. Finally, given that the weights given by TSW and $MTFIDF$ are probably in different scales, fusing them by a multiplication turns out to be more adequate.

5.5 Analysis

To corroborate the usefulness of using spatial relations for image retrieval, we evaluate statistical significance by using paired *Student's t test* [23]. This is performed on

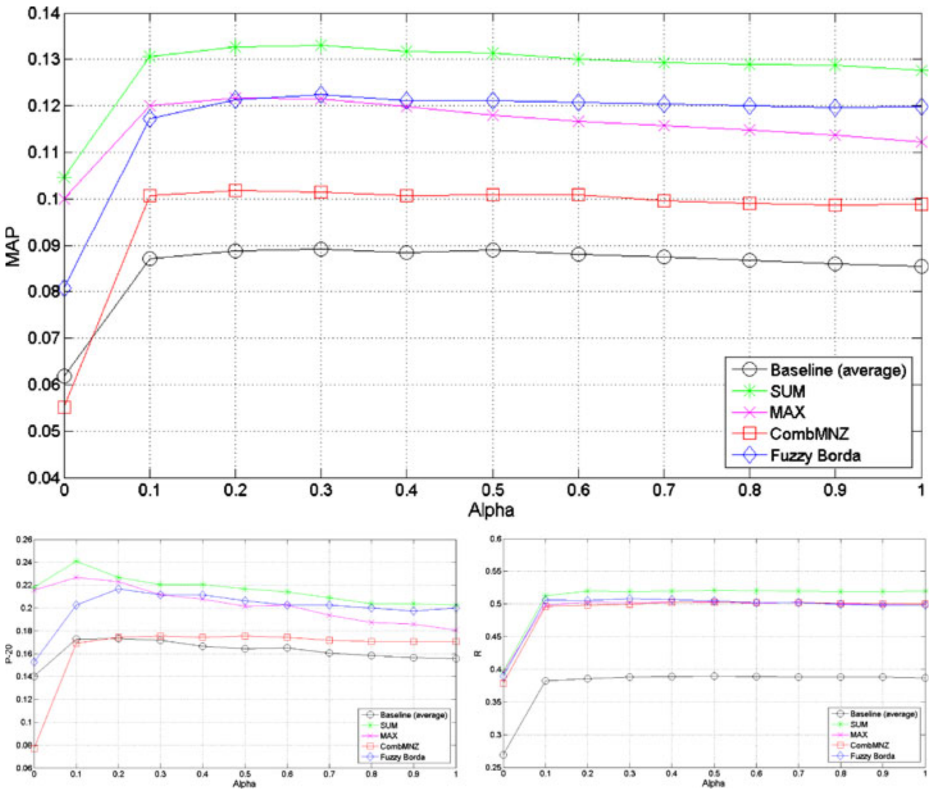


Fig. 10 Results of the evaluation of the late fusion methods. *Top*: mean average precision (MAP). *Bottom-left*: precision at the first 20 images retrieved (P-20). *Bottom-right*: recall (R)

the combination of $(MTFIDF)$+SUM+TSW comparing the use of spatial relations at any level ($\alpha < 1$) with using no spatial relations ($\alpha = 1$). From these results we highlight the fact that statistically significant differences favoring the use of spatial relations over not using them were obtained in all of the measures (except for recall). Results for this evaluation are shown in Table 5. Graphs in Figs. 9 and 10 also corroborate the relevance of spatial relations, since when $\alpha = 0$ (not using spatial relations) results are in all of the cases the lowest for every measure.

Finally, we compare the application of the best improvements we proposed (SUM, TSW and $MTFIDF$) with different combinations, and performed a hypothetical comparison to the 33 methods based on visual information provided for ImageCLEF 2008. Table 6 summarizes these results. By combining the fusion and weighting schemes we were able to improve even more our results, compared to the individual use of these methods. Particularly, $(MTFIDF)$+SUM+TSW is the combination that provides us with the best results, followed by $(MTFIDF)$+SUM. The relative improvement of $(MTFIDF)$+SUM+TSW with respect to the baseline is about 157% for MAP, while for $(MTFIDF)$+SUM is about 60% evaluated by R. It is important to highlight that just in 2 of these tests the best result was obtained by using no spatial relations, while in almost half of these tests the best results were obtained when $\alpha \leq 0.5$.

Table 3 Summary of the results comparing different late fusion schemes (MAP, P-20 and R)

Method	$\alpha = 1$	Best $\alpha < 1$	Relative improvement (%)
MAP			
Average	0.0855	0.0891	4.21
SUM	0.1277	**0.1330**	4.15
MAX	0.1122	0.1217	**8.47**
CombMNZ	0.0989	0.1018	2.93
*Fuzzy Bord*a	0.1198	0.1225	2.25
Improvement with respect to the baseline			**49.27**
P-20			
Average	0.1556	0.1731	11.25
SUM	0.2026	**0.2410**	18.95
MAX	0.1808	0.2269	**25.50**
CombMNZ	0.1705	0.1756	2.99
Fuzzy Borda	0.2000	0.2167	8.35
Improvement with respect to the baseline			**39.22**
R			
Average	0.3869	0.3894	0.64
SUM	0.5198	**0.5210**	0.23
MAX	0.5002	0.5035	**0.66**
CombMNZ	0.5002	0.5027	0.50
Fuzzy Borda	0.4981	0.5081	2.01
Improvement with respect to the baseline			**33.79**

The second column shows the result obtained when no spatial relations are used ($\alpha = 1$), the third column shows the best result when spatial relations are used ($\alpha < 1$) and the fourth column shows the relative improvement by using spatial relations. For each measure, we show in bold the best results. We also show in bold the highest relative improvement using spatial relations with respect to the baseline (average)

Regarding the computing time, we ran our image retrieval method fixing $alpha = 0.1$ with the $MTFIDF$ weighting. This process is performed on a computer with a 2 GHz Centrino 2 processor with 4 GB of RAM using Matlab. For this configuration, we obtained that for comparing one of the sample images against the 20,000 images in the database the minimum processing time is 9.61 seconds, the maximum time

Table 4 Comparison of different variations of topic-specific weighting (TSW)

Variation	MAP
CMNZ+i	0.1070
CMNZxi	0.1288
CMNZ+in	0.1144
CMNZxin	**0.1347**
FBorda+i	0.1016
FBordaxi	0.1092
FBorda+in	0.1075
FBordaxin	0.1286

We compare fusions of initial lists by mean of *combMNZ* and *fuzzy Borda*, adding (+) or multiplying (**x**) to $MTFIDF$, and with the use of just the labels in the sample images (**i**) or the labels in the sample images and their textual description (**in**). Results are obtained fixing $\alpha = 0.1$. For each measure, we show in bold the best results. We also show in bold the highest relative improvement using spatial relations with respect to the baseline

Table 5 Statistical significance tests for retrieval comparing using spatial relations against not using them

Measure	Significance	Reliability (%)
MAP	+	97.5
P-5	+	99.5
P-20	+	99.5
R	−	

The combination ($MTFIDF$)+SUM+TSW is measured by MAP, P-5, P-20 and R

Table 6 Hypothetical comparison with respect to the 33 methods based on visual information participating in ImageCLEF 2008 competition, measured by MAP and R

Method	MAP	Relat. imp. to baseline	Pos.	α	R	Relat. imp. to baseline	Pos.	α
Best results ImageCLEF 2008	0.2103	N/A	1	N/A	0.4993	N/A	1	N/A
BASELINE	0.0563	–	20	0.5	0.3247	–	17	0.7
MTFIDF	0.0891	58.41%	15	0.3	0.3894	19.92%	15	0.5
SUM	0.0961	70.79%	14	0.6	0.4227	30.18%	13	0.5
TSW	0.1050	86.67%	14	1.0	0.3714	14.37%	16	0.9
(MTFIDF)+SUM	**0.1330**	**136.37%**	**4**	**0.3**	**0.5210**	**60.45%**	**1**	**0.5**
(MTFIDF)+TSW	0.1051	86.73%	14	0.6	0.3766	15.99%	15	1.0
SUM+TSW	0.1278	127.13%	9	0.9	0.4390	35.19%	13	0.9
(MTFIDF)+SUM+TSW	**0.1447**	**157.17%**	**4**	**0.5**	**0.4990**	**53.66%**	**2**	**0.8**

The hypothetical position with each variation and their combinations is shown. The best results obtained by the competitors in ImageCLEF 2008 are shown in the first row, for each measure. For each measure, we show in bold the best results. We also show in bold the highest relative improvement using spatial relations with respect to the baseline

is 27.37 seconds, and the average time is 18.24 seconds. These values are computed over the 39 topics and each of the three sample images for each topic. Processing time could be significantly reduced if we re-implemented our algorithms in a more efficient compiler and also using more powerful computers. Image segmentation and annotation as well as spatial relation computing need to be performed only once for an image, so they can be considered as offline processing. Particularly, although in these experiments image segmentation and annotation are manually performed, in the future we expect to be able to automate both processes.

6 Conclusions and future work

6.1 Conclusions

In this paper we have explored a series of methods for improving CBIR by including spatial relations present in the image. The interest in adding spatial relations to the retrieval process lays on the fact that they are high-level information closer to the needs of the user of an image retrieval system. The method we employ consists of using a set of segmented and annotated images for image retrieval. The image segments are used to obtain spatial relations in the image itself. The combination of image annotations and spatial relations is used to compare sample images to each of the images in a database, obtaining a list of similarities. Once we have this list, we retrieve the k most similar images and are able to answer the query. The conclusions we derive from our results are: (i) The use of spatial relations helps to improve retrieval results, compared to just using image annotations; (ii) Label weighting, based on label frequency or topic-specific features, seems to be an adequate aid for retrieval; (iii) Late fusion, using several sample images produces better performance in retrieval than using a single one.

The most relevant contributions of our work are the following:

- The analysis of topological relations and order relations, and the selection and validation of the set of spatial relations adequate for image retrieval.
- An image retrieval method combining concepts and spatial relations, which is based on conceptual graphs. This method is intended to use high-level information acquired by annotating images, and use this information to better answer image retrieval.
- A label weighting method, called $MTFIDF$. This is an adaption of $TFIDF$, a frequently used term-weighting method, which in its original form performed poorly. The modified method prioritizes those labels co-occurring with the smallest amount of labels in an image.
- A label weighting method called topic-specific weighting, using late fusion to take advantage of the availability of several sample images. This information is complemented by the use of the textual description of the retrieval topic and this weighting is intended to consider terms as more relevant when they are frequent along the sample images and textual description of each specific topic.
- An extensive experimental evaluation of the proposed methods for improving image retrieval by incorporating spatial relations, label weighting and combining several sample images. By evaluating several scenarios we intend to provide evidence on the advantages of each individual method, and the combination of them.

6.2 Future work

Some possibilities we could explore in the future work are the following:

- Take advantage of the label hierarchy available for the SAIAPR TC-12 image database, in order to better generalize the concepts and obtain a more flexible retrieval.
- Perform additional experiments with an automatic annotation method.
- Analyze the spatial relations we simplified, as well as other spatial relations that could be added. An important advantage of our methodology is that the set of spatial relations used can be easily replaced in a way that it is possible to evaluate different representations of the spatial relations.

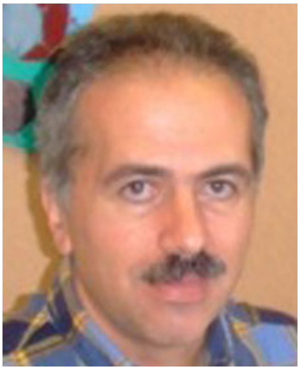
Acknowledgements This research was partially supported by CONACyT under project grant CB-2008-01-103878 and postdoctoral fellowship 10216. Thanks to Dr. Antonio Juárez who kindly processed our lists with his fusion methods to obtain results for *fuzzy Borda count* and *combMNZ*.

References

1. Arni T, Sanderson M, Clough P, Grubinger M (2008) Overview of the imageclef 2008 photographic retrieval task. Working Notes of the CLEF
2. Aslam JA, Montague M (2001) Models for metasearch. In: Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval, pp 276–284
3. Bekhatir M, Chiaramella Y, Mulhem P (2005) A signal/semantic framework for image retrieval. In: JCDL 005: proceedings of the 5th ACM/IEEE-CS joint conference on digital libraries. ACM, pp 368–368
4. Berretti S, Bimbo AD, Vicario E (2003) Weighted walkthroughs between extended entities for retrieval by spatial arrangement. IEEE Trans Multimedia 5(1):52–70
5. Berretti S, Del Bimbo A, Vicario E (2001) Efficient matching and indexing of graph models in content-based retrieval. IEEE Trans Pattern Anal Mach Intell 23(10):1089–1105. doi:10.1109/34.954600
6. Carneiro G, Chan AB, Moreno PJ, Vasconcelos N (2007) Supervised learning of semantic classes for image annotation and retrieval. IEEE Trans Pattern Anal Mach Intell 29:2007
7. Chang SK, Shi QY, Yan CW (1987) Iconic indexing by 2-d strings. IEEE Trans Pattern Anal Mach Intell 9(3):413–428
8. Chen J, Li Z, Li C, Gold CM (1998) Describing topological relations with voronoi-based 9-intersection model. Int Arch Photogramm Remote Sens 32(4):99–104
9. Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age. ACM Comput Surv 39:1–60
10. Duygulu P, Barnard K, de Freitas JFG, Forsyth DA (2002) Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In: Proceedings of the seventh European conference on computer vision, vol 4, pp 97–112. citeseer.ist.psu.edu/duygulu02object.html
11. Egenhofer MJ (1989) A formal definition of binary topological relationships. In: Proceedings of the 3rd international conference, FODO 1989 on foundations of data organization and algorithms, pp 457–472
12. Egenhofer MJ (1993) Definitions of line-line relations for geographic databases. IEEE Data Eng Bull 16(3):40–45
13. Egenhofer MJ, Franzosa RD (1991) Point-set topological spatial relations. Int J Geogr Inf Syst 5(2):161–174
14. Escalante H, Hernández-Gracidas C, González J, López A, Montes M, Morales E, Sucar L, Nor LV, Grubinger M (2010) The segmented and annotated iapr-tc12 benchmark. Comput Vis Image Underst 114(4):419–428
15. Frank AU (1992) Qualitative spatial reasoning about distances and directions in geographic space. J Visual Lang Comput 3(4):343–371
16. Goodrum A (2000) Image information retrieval: an overview of current research. Inf Sci 3:2000
17. Goyal RK, Egenhofer MJ (2000) Cardinal directions between extended spatial objects. IEEE Trans Knowl Data Eng (in press)
18. Grubinger M (2007) Analysis and evaluation of visual information systems performance. PhD thesis, School of Computer Science and Mathematics, Faculty of Health, Engineering and Science, Victoria University, Melbourne, Australia
19. Hernández-Gracidas C, Sucar LE (2007) Markov random fields and spatial information to improve automatic image annotation. In: Proceedings of the 2nd Pacific Rim conference on advances in image and video technology, PSIVT'07. Springer, Berlin, Heidelberg, pp 879–892
20. Hernández-Gracidas C, Sucar LE, Montes-y Gómez M (2009) Modeling spatial relations for image retrieval by conceptual graphs. In: Proceedings of the first Chilean workshop on pattern recognition
21. Hernández-Gracidas C, Juárez A, Sucar LE, Montes-y Gómez M, Villaseñor L (2010) Data fusion and label weighting for image retrieval based on spatio-conceptual information. In: Proceedings of the 9th international conference on adaptivity, personalization and fusion of heterogeneous information (RIAO)
22. Jones KS (1972) A statistical interpretation of term specificity and its application in retrieval. J Doc 28:11–21
23. Kanji GK (1993) 100 statistical tests / Gopal K. Kanji. Sage Publications, London, Newbury Park, California

24. Lee JH (1997) Analyses of multiple evidence combination. In: SIGIR '97: proceedings of the 20th annual international ACM SIGIR conference on research and development in information retrieval. ACM, New York, pp 267–276
25. Mechkour M, Berrut C, Chiaramella Y (1995) Using a conceptual graph framework for image retrieval. In: Proc. of the MMM95 (multimedia modeling) conference, pp 127–142
26. Millard DE, Gibbins NM, Michaelides DT, Weal MJ (2005) Mind the semantic gap. In: HYPERTEXT '05: proceedings of the sixteenth ACM conference on hypertext and hypermedia. ACM, New York, pp 54–62
27. Montes M, Gelbukh A, López-López A, Baeza-Yates R (2000) Comparison of conceptual graphs. In: Proceedings of the Mexican international conference on articial intelligence, MICAI 2000
28. Paul K (1995) Decision level data fusion for routing of documents in the trec3 context: a best case analysis of worst case results. In: Proceedings of the third text retrieval conference (TREC-3)
29. Picard RW, Minka TP (1995) Vision texture for annotation. In: Multimedia systems
30. Rathi V, Majumdar AK (2002) Content based image search over the world wide web. In: Proceedings of the Indian conference on computer vision, graphics and image processing
31. Ren W, Singh M, Singh S (2002) Image retrieval using spatial context. In: Proceedings of the 9th international workshop on systems, signals and image processing, pp 44–49
32. Rui Y, Huang TS, Chang S (1999) Image retrieval: current techniques, promising directions and open issues. J Vis Commun Image Represent 10(1):39–62
33. Shaw JA, Fox EA, Shaw JA, Fox EA (1994) Combination of multiple searches. In: Proceedings of the second text retrieval conference (TREC-2), pp 243–252
34. Sistla AP, Yu C, Haddack R (1994) Reasoning about spatial relationships in picture retrieval systems. In: Proceedings of the 20th international conference on very large data bases, pp 570–581
35. Smeulders AWM, Member S, Worring M, Santini S, Gupta A, Jain R (2000) Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22:1349–1380
36. Sowa JF (1984) Conceptual structures: information processing in mind and machine. Addison-Wesley Longman Publishing Co., Inc., Boston
37. Vogt CC, Cottrell GW (1999) Fusion via a linear combination of scores. Inf Retr 1(3):151–173
38. Wu S, McClean S (2006) Performance prediction of data fusion for information retrieval. Inf Process Manag 42(4):899–91
39. Yuan J, Li J, Zhang B (2007) Exploiting spatial context constraints for automatic image region annotation. In: MULTIMEDIA '07: proceedings of the 15th international conference on multimedia. ACM, New York, pp 595–604
40. Zhang Q, Yau SS (2004) On intractability of spatial relationships in content-based image database systems. Commun Inf Syst 4(2):181–190
41. Zhou X, Chen J, Li Z, Zhao R, Zhu J (2005) A model for topological relationships based on euler number. In: Proceedings of the ISPRS Hangzhou workshop

Carlos Arturo Hernández-Gracidas obtained the Ph.D. with specialty in computer science degree in 2010 at the National Institute of Astrophysics, Optics and Electronics, in Puebla, Mexico. In 2005 he obtained a M.Sc. with specialty in computer science, also at the National Institute of Astrophysics, Optics and Electronics. He studied the bachelor in computer science at the Autonomous University of Puebla in Mexico, obtaining the degree in 2001. He has worked in object recognition using the appearance-based model approach. His current work is centered on the use of computer vision and methods of representation of spatial knowledge applied to content-based image retrieval. In his research he has explored the use of artificial intelligence, data fusion and information modeling.



Luis Enrique Sucar has a Ph.D in computing from Imperial College, London, UK, 1992; a M.Sc. in electrical engineering from Stanford University, California, USA, 1982; and a B.Sc. in electronics and communications engineering from ITESM, Monterrey, México, 1980. He is currently a Senior Researcher at INAOE, Puebla, Mexico. Dr. Sucar is Member of the National Research System, the Mexican Science Academy, AAAI, SMIA and Senior Member of the IEEE. He has served as president of the Mexican AI Society, has been member of the Advisory Board of IJCAI, and is Associate Editor of the journals Computación y Sistemas and Revista Iberoamericana de Inteligencia Artificial. His main research interests are in graphical models and probabilistic reasoning, and their applications in computer vision, robotics and biomedicine.

Manuel Montes-y-Gómez obtained his Ph.D. in Computer Science from the Computing Research Center of the National Polytechnic Institute of Mexico. Currently, he is a full-time lecturer at the Computational Sciences Department of the National Institute of Astrophysics, Optics and Electronics, located in Puebla, Mexico. His research is on automatic text processing. He is author of more than 100 international papers in the fields of information retrieval, question answering, information extraction and text mining. Dr. Montes has been visiting professor at the Polytechnic University of Valencia (Spain), at the University of Geneva (Italy), and from August 2010 he is visiting professor at the University of Alabama (USA). In addition, Dr. Montes is member of the Mexican National System of Researchers, the Mexican Society of Artificial Intelligence, the Mexican Association for Natural Language Processing and the International Web Intelligence Consortium.