Modeling Spatial Relations for Image Retrieval by Conceptual Graphs

Carlos Hernández-Gracidas, L. Enrique Sucar and Manuel Montes-y-Gómez
Computer Science Department
National Institute of Astrophysics, Optics and Electronics
Puebla, Mexico
carloshg@ccc.inaoep.mx, esucar@inaoep.mx, mmontesg@inaoep.mx

***Abstract*—Content-based image retrieval is one attractive research field in computer vision, but also one facing critical problems. Approaches using the image as a whole and those focused on identifying relevant objects in the image, usually fail with several topics, for which they cannot provide a rich enough representation. Recent methods try to solve this problem by considering additional information. In this paper we present a retrieval method based on spatial relations and designed using conceptual graphs. Our results are competitive with state of the art approaches, and three important observations are made: first, this method shows spatial relations improve results; second, it corroborates having more samples improves retrieval; third, this method is better suited for queries focused in the contents of the image, rather than on its associated text.**

***Keywords*-Content-based image retrieval; conceptual graphs; spatial relations;**

I. Introduction

Image retrieval consists of obtaining a subset of relevant images from an image set, according to certain criteria. This retrieval may be performed mainly in one of two ways[1], [2], [3]: from a set of sample images, or from a textual query. We are particularly interested in retrieval based on sample images, which is frequently regarded as content-based image retrieval (CBIR). In this approach the tendency is to use the sample images to extract low-level features that are expected to describe the general idea behind the topic. A variation of this method consists of using these features to try to identify the objects in the image, and then associate labels to it. The main problem with using low-level features is that they have many difficulties for generalizing concepts and fail at distinguishing between two visually similar concepts.

Using high-level information, and particularly spatial setup to complement labels, is an interesting alternative to the simple use of low-level features, and this is the approach we follow in this paper. What we mean to establish is if by using spatial relations among regions found in a sample image, image retrieval is capable of providing good results. The objective is to compare structures consisting of both, labels and spatial relations, between pairs of images, and for this purpose, images must be segmented, annotated, and the spatial relations among these regions must be computed. After this, a similarity measure must be designed in order to perform retrieval.

One important problem usually related to this kind of retrieval is that with current automatic methods it is hard to obtain good segmentations and annotations, and consequently, results are severely affected by defective data. For this reason, we use manually segmented and annotated images, in order to perform our experiments; and, although we cannot refer to this information as "perfect data" (as we will explain in further sections) it should give an idea on what to expect from the exploration of this kind of image retrieval. We propose a new image retrieval method which takes annotations and spatial relations, and uses them to compare images, based on the conceptual graph comparison method introduced in [4]. We present a series of experiments on the IAPR-TC12 image database [5] and its manual segmentation and annotation sets [6], also using the topics defined for the ImageCLEF 2008 and their respective relevance judgments [7]. These experiments show conceptual graphs are a competitive representation for image retrieval, showing that the more sample images, the better the retrieval will be. We also confirm that this method performs better with queries searching for image contents. Finally, results clearly show that although totally discarding image labels for retrieval brings poor results, giving the highest weight to spatial relations is the configuration that provides the best results.

II. Related Work

The use of high-level features, and particularly of spatial information for image retrieval is far from new. In [8] they are based on the comparison of low-level features, and they include spatial relation by adding an extra step of spatial relations coding to the retrieval process. They use this method on a limited set of images containing objects, such as *grass*, *sky*, *clouds* and *trees*. In the case of [9], there is an interesting similarity with our approach, since they also represent images as graphs. They use graph isomorphism, topological similarity and other low-level region features to compare images.

Our work is based on the idea of conceptual graphs [10], [11], and particularly on [4]. Conceptual graphs are intended to express relations among concepts from natural language texts in the form of a network where both, the concepts found in the text, and the relations among these concepts, are represented by nodes in a graph. These graphs have the advantage of being computationally tractable and readable by humans. The use of semantic representations for image retrieval has already been explored in previous works, from which, particularly, conceptual graphs have been used to this end. One example of this is [12], where a multifaceted image representation is used for indexing and retrieval. In [13] we have another example of the use of conceptual graphs for image modeling and retrieval, where they even include spatial relations in their representation. Nevertheless, some differences must be pointed out:

- A desirable feature in conceptual graphs is that the original graphs are not completely related, i.e., the less relations among nodes, the more relevant information we will obtain. However, the spatial graphs we obtain, given their nature, are complete graphs, which means that every node in the graph is related to the remaining ones.
- In the case of the original conceptual graphs, only one graph is used to represent the relations. In our case, we use three different graphs (one per relation group), and for such reason, the formula for measuring similarity must include the information from these three graphs at the same time.
- It is also important to notice that in [4], the conceptual and relational similarities are combined by a multiplicative formula. In our case, however, an additive formula seems to be a more adequate solution for this fusion, given that we are interested on knowing how useful spatial relations are with respect to concepts.

A. Spatial Relations

Spatial relations are determined for an object with respect to another object of reference. These relations are useful to understand the relative position of an object in a scene. Three types of spatial relations are frequently defined in the literature [14]:

- Topological relations: These relations are characterized by the fact that they are preserved even when topological transformations, such as translation, rotation or scaling are applied to the image.
- Order relations: These relations are based on the definition of order. These relations are variable under rotation, but are preserved when scaling and translation are applied to the image.
- Metric relations: These relations are based on measures, such as distance and direction, and are affected by scaling, but unaffected by rotation and translation.

III. Our Image Retrieval Method

In [15], we presented a method to improve automatic image annotation by using spatial relations. Spatial relations were divided into three groups, namely: topological relations, horizontal relations, and vertical relations (see Table I). For each group, there will be one and only one relation between every pair of regions in an image, so the image is represented by a complete graph in each group.

			Directed	**Undirected**
Topological		1		Adjacent
		2		Disjoint
Order	Horizontal	3		Beside (left or right)
		4		H. aligned
	Vertical	5	Above	
		6	Below	
		7		V. aligned

Table I
Spatial relations used in this work.

A method for comparing conceptual graphs is introduced in [4], based on 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 the two compared graphs have in common, while S_r measures how similar the relations among the concepts in common are. The formulae are presented now:

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

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. Relational similarity measures how similar the relations among the same concepts in common are:

$$S_r = \frac{2m(G_c)}{m_{G_c}(G_1) + m_{G_c}(G_2)} \tag{2}$$

where $m(G_c)$ is the number of relations in the graph G_c, and $m_{G_c}(G_1)$ and $m_{G_c}(G_2)$ are the number of the arcs in the immediate neighborhood of the graph G_c in the graphs G_1 and G_2, respectively. These measures are combined by:

$$S = S_c \times (a + b \times S_r) \tag{3}$$

where a is given by

$$a = \frac{2n(G_c)}{2n(G_c) + m_{G_c}(G_1) + m_{G_c}(G_2)} \tag{4}$$

and $b = 1 - a$.

We present a new method for image retrieval based on this graph comparison method[4]. We represent images by means of conceptual graphs, and take advantage of spatial relations to show how the objects in the image interact. For S_c we use the same formula used in [4], considering the annotations given to the regions in an image as the concepts. We reformulate S_r as:

$$S_r = \frac{2m(G_{Tc}) + 2m(G_{Xc}) + 2m(G_{Yc})}{3m_{G_c}(G_1) + 3m_{G_c}(G_2)} \tag{5}$$

S_r is modified in order to consider the three relational graphs, so $m(G_{Tc})$, $m(G_{Xc})$ and $m(G_{Yc})$ represent the number of arcs (relations) in common between the two compared images, for topological, X and Y relations, respectively. In this case, given the completeness of graphs, we consider $m_{G_c}(G_1)$ and $m_{G_c}(G_2)$ as the total number of relations in the first and second image, respectively. We obtain the final similarity measure S by a simple additive formula considering S_c and S_r, and giving each a weight which depends on α:

$$S = \frac{\alpha S_c + (1-\alpha) S_r}{2} \tag{6}$$

In Figure 1 we show an example of how the images in the database are segmented and annotated, and how the spatial relations in the image can be represented by means of conceptual graphs. Rectangular nodes represent the labels, while elliptic nodes represent the spatial relations between a pair of labeled regions. In accordance with Table I all of the relations are undirected except for the vertical relations *above* and *below*.

For two images to be compared, they must be preprocessed by segmenting and annotating them. After this, spatial relations are computed in order to build their conceptual graphs. Once we have the conceptual graph for both images, they can be compared using the similarity formula (S). Finally, by repeating this process, one image can be compared against the whole image database, obtaining a ranked list. Figure 2 shows the diagram for this comparison.

IV. The Database

The image database used in our experiments is the IAPR-TC12 [5], consisting of 20,000 images of sports events, people, animals, cities and landscapes. There are some characteristics inherent to this image database that make it considerably difficult for annotation and retrieval, even using manual segmentation and annotation. These characteristics are:

1) The number of images is considerable (20,000 images).
2) Image diversity. Images range from archaeological ruins, houses, schools, monuments, etc, to family pictures. These pictures were taken all around the world mostly by tourists with different cameras and under quite different conditions; some of them were taken indoors, while others were taken outdoors.
3) The number of different objects is very high, and that is the reason why a set of initial labels according to a pseudo-ontology is used[6].

Some images from the IAPR-TC12 are shown in Figure 3.

The manually segmented and annotated IAPR-TC12 [6] was chosen for the experiments, given that it provides a reliable dataset that allows focusing more on the retrieval than on the effects of automatic segmentation and/or annotation.

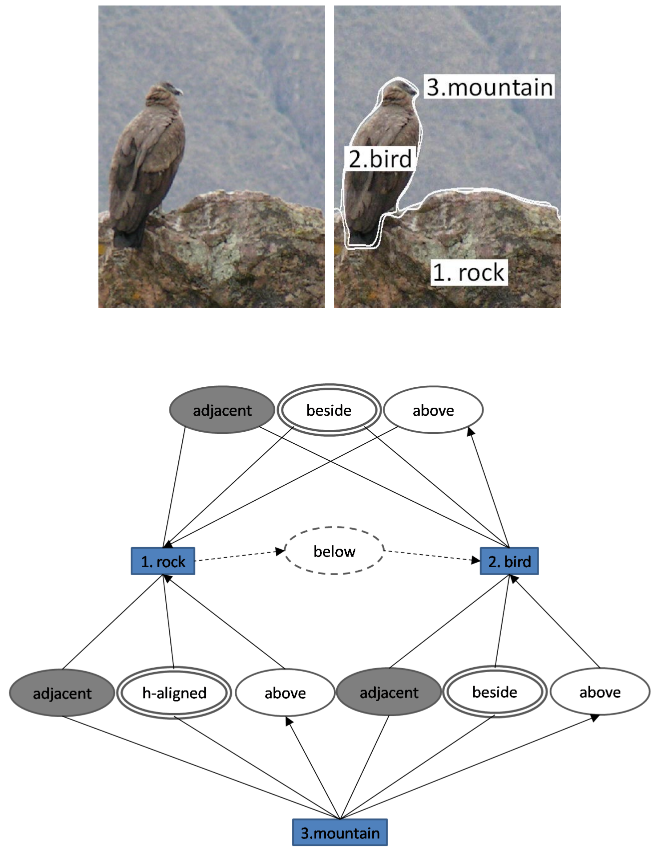


Figure 1. Top-left: One of the images in the IAPR-TC12. Top-right: The same image with its manual segmentations and annotations. Bottom: Conceptual graph of the spatial relations among the regions in which the image was divided. Topological relations are shown with filled nodes, horizontal relations appear in double-lined border, and vertical relations appear in single-lined border. The vertical relation *below* is not shown in all cases, since it is implied by *above*.

A. The Topics

For an objective evaluation of our method, we resorted to the 39 topics developed for the ImageCLEF 2008 photo retrieval task [7]. 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.

For the Image CLEF 2008, 39 retrieval topics are provided, and depending on the kind of information that is needed to accurately retrieve relevant images, they were previously classified by their creators as visual (20 topics) or non-visual (19 topics). Visual topics refer to the topics that are focused on the contents of the image, ie., the key information for the retrieval is supposed to be obtained from the sample images provided for the topic. Non-visual topics are those topics for which the key information is supposed to be obtained from the text accompanying the topic, and

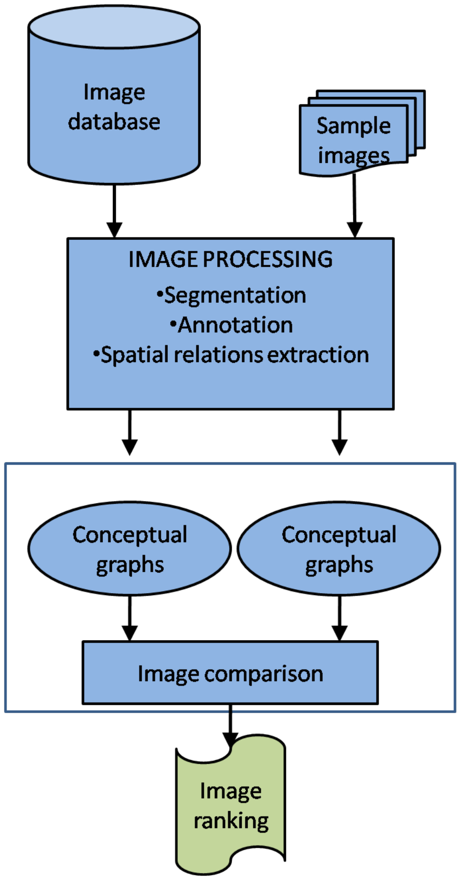


Figure 2. Diagram of the image retrieval by conceptual graphs.

Figure 3. Samples of the IAPR-TC12 image database.

for such reason we also call them textual topics.

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 have already been used in other research, a comparison is possible.

V. EXPERIMENTS

Three different experiments were designed, for which the purpose was:

1) Find out how useful spatial relations are for retrieval, with respect to concepts.
2) Corroborate if retrieval is improved if several sample images are available for a topic.
3) Observe the individual behavior of the method with visual and textual topics.

Experiments were performed using the three sample images provided for each topic. For each of the sample images, its conceptual graph was automatically obtained, and the same is done for each image in the database. We searched over the 20,000 images in the database, and obtained a ranked list, where the rank depends on the similarity measure described in Section III. After the list is retrieved, we keep the 1,000 more similar images. This is done for each of the 39 topics. Finally, we evaluate the retrieval performance, by computing MAP, P-20, and R measures using the relevance judgments [7]. 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. 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 give each label a weight similar to the idea of tf-idf [16], this weight is inverse to the number of times the label appears among the collection of images in the database. This is done considering that labels that are less frequent will give more information than those that are common to several images. Experiments proving this idea were performed but are not included.

A. Concepts vs Spatial Relations

Given that one of the interests of this research is to determine if spatial information provides better results than the simple use of annotations, we performed an experiment varying α from 0 to 1, with increments of 0.1. The results of these experiments over the 39 topics are shown in Figures 4, 5 and 6. It is noticeable, specially with MAP and P-20, that small values of α (approximately between 0.1 and 0.3) are the ones that provide the best performance (experiments fusing results from the three images give the highest scores with $\alpha = 0.1$), compared to the use of only conceptual similarity ($\alpha = 1$), which is the typical approach. This clearly means that the inclusion of spatial relations in the retrieval process, and even giving more weight to them than to the concepts, actually improves results. However, there seems to be a limit when $\alpha \to 0$, since $alpha = 0$ produces the worst results. This apparent limit took us to perform additional experiments with increments of 0.01 between 0 and 0.1, which are also shown in the graph. The 0-limit, on the other hand, indicates that concepts should not be completely omitted for retrieval.

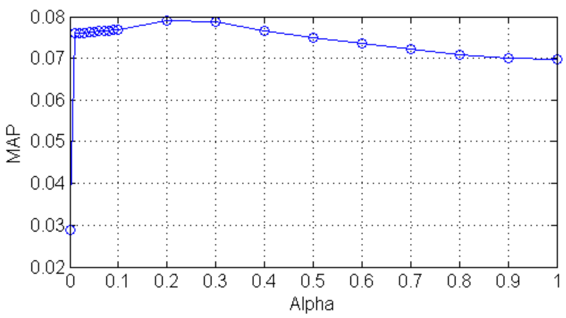


Figure 4. MAP, varying α from 0 to 1. Average results for the 39 topics. The values shown are the result of computing the mean of the execution using the three sample images individually.

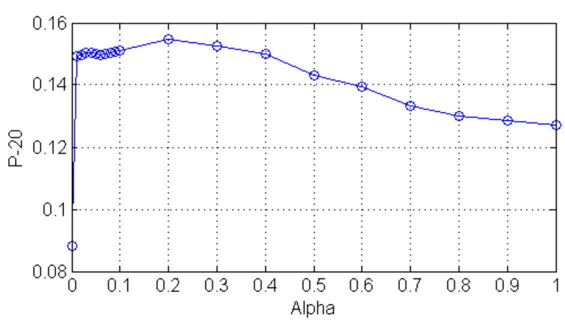


Figure 5. P-20, varying α from 0 to 1. Average results for the 39 topics. The values shown are the result of computing the mean of the execution using the three sample images individually.

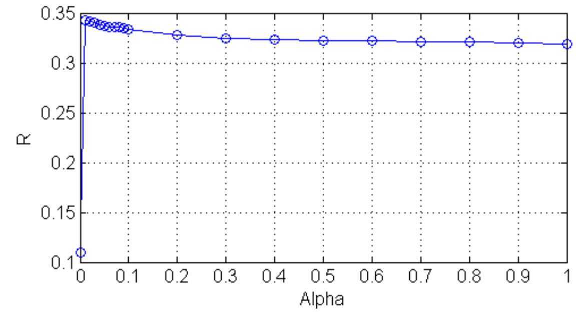


Figure 6. R, varying α from 0 to 1. Average results for the 39 topics. The values shown are the result of computing the mean of the execution using the three sample images individually.

B. *List Fusion for Retrieval*

Since we had three different images for each topic, there was a need to know if by combining these images, or the resulting lists using these images, could improve results. We performed an experiment with different simple list fusions, where, rather than finding the best fusion method, we were interested in understanding how useful a fusion would be. Independent retrievals were performed, obtaining the lists L_1, L_2 and L_3, one for each sample image. These lists were then fused, and the fusion variations used are:

- List addition (LA). The similarity value given by each list is added in order to obtain a final list L_f, where, for each element $L_f(i)$, $L_f(i) = L_1(i)+L_2(i)+L_3(i)$.
- Maximum value (MV). The similarity value is obtained by computing the maximum value from the three lists for the same image. This means that for each element $L_f(i)$, $L_f(i) = max(L_1(i), L_2(i), L_3(i))$
- Normalized list addition (NLA). Same as LA, but similarity values are normalized in order to avoid a particular image to be considered as more relevant than the others. Normalization is performed per list and not globally before the addition.
- Normalized maximum value (NMV). Same as MV, but similarity values are normalized as well.

The results obtained with the different schemata are shown in Figures 7, 8 and 9. We also show in each graph the performance with the individual images, for reference. The experiments shown are the ones with $\alpha = 0.1$, where we obtained our best scores.

It is clear from the graphics, that for most measures, the performance of any fusion schema is higher than the performance with a single image, having our best results with the simple LA method.

C. *Per topic performance*

Regarding the individual performance of the method with respect to each separate topic, we present in Figures 10, 11,

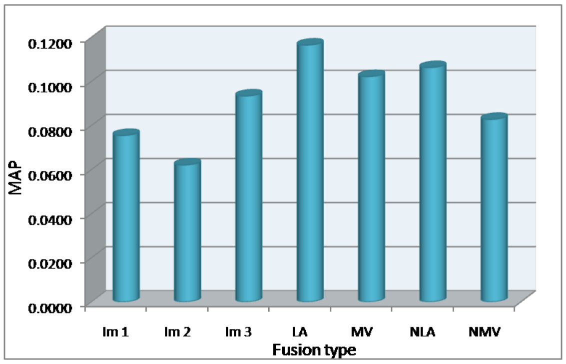


Figure 7. Comparison of the different fusion types and the individual performance with a single image. We show MAP, fixing $\alpha = 0.1$.

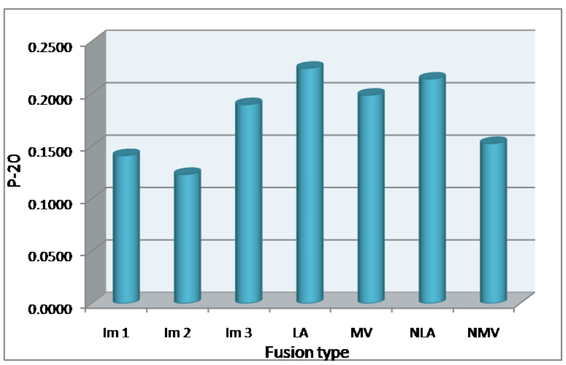


Figure 8. Comparison of the different fusion types and the individual performance with a single image. We show P-20, fixing $\alpha = 0.1$.

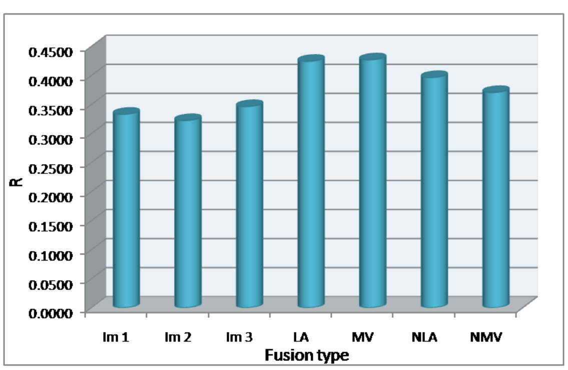


Figure 9. Comparison of the different fusion types and the individual performance with a single image. We show R, fixing $\alpha = 0.1$.

12 and 13 performance measured by AP, P-5 (precision at the first 5 documents), P-20, and R, respectively, for the LA fusion and fixing $\alpha = 0.1$. As a reference, we also show the mean for the 39 topics, the textual topics and the visual topics. Textual and visual means are separately computed over the textual and visual topics, respectively; in order to show how good the method is individually with each type of topic. Results show our method is more adequate for retrieval based on the image contents, since segmentation and annotation are aimed to describe the contents of the image from the visual point of view. It also shows that there are some topics that turned out to benefit from this method, while there are others that are not well suited for it, and consequently, results are extremely poor on them. Performance of P-5 and P-20, gives some evidence that automatic feedback based on the retrieval of the first images could be of benefit, since clearly, for several topics, the first retrieved images have a high probability to be useful if they are considered as additional samples.

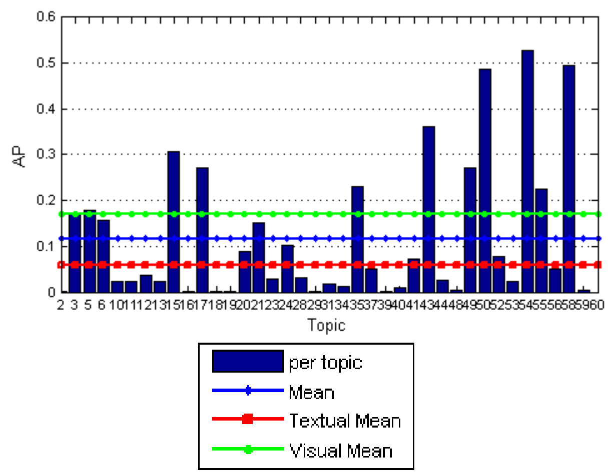


Figure 10. Per topic performance of retrieval. Results of the LA fusion with $\alpha = 0.1$ are shown. AP versus visual and textual mean.

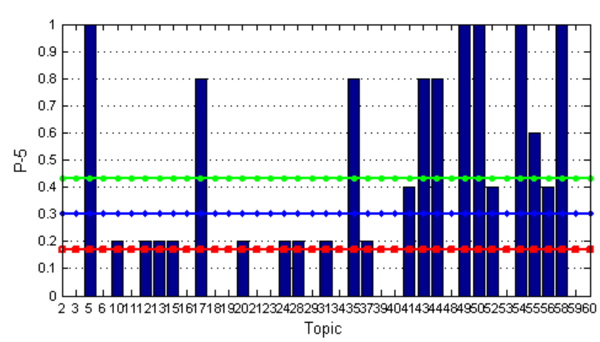


Figure 11. Per topic performance of retrieval. Results of the LA fusion with $\alpha = 0.1$ are shown. P-5 versus visual and textual mean. The same legends as in Figure 10 are used.

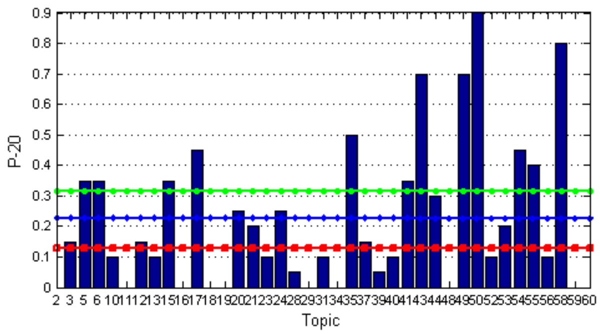


Figure 12. Per topic performance of retrieval. Results of the LA fusion with $\alpha = 0.1$ are shown. P-20 versus visual and textual mean. The same legends as in Figure 10 are used.

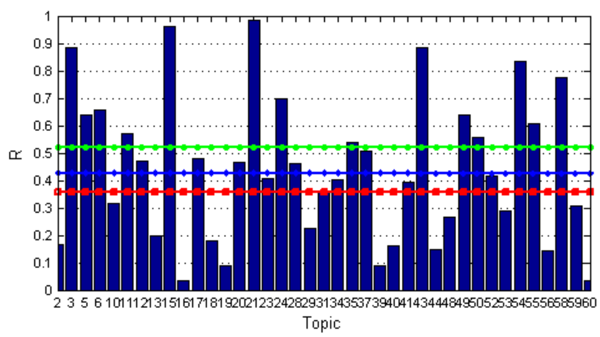


Figure 13. Per topic performance of retrieval. Results of the LA fusion with $\alpha = 0.1$ are shown. R versus visual and textual mean. The same legends as in Figure 10 are used.

Table II) shows examples of the retrieval using our method for 4 of the 39 topics. For each topic we show the textual query, the visual query (i.e., the 3 sample images), and the top-5 retrieved images using our method. The first two examples are our best results (measured with AP, and the last two examples are among our worst results, also measured with AP. These results were obtained using LA and fixing $\alpha = 0.1$.

VI. Analysis

Considering manual segmentation and annotations were used for the experiments, our results might be considered not as good as expected. However, several factors must be taken into account in order to understand the results. First of all, although manually treated data are expected to be more reliable, it cannot be considered as perfect. As an example of this, we have topic 2 (a church, cathedral or a mosque with three or more towers), which shows a poor retrieval, even when this is a visual topic (examples of the retrieval for this topic are shown in Table II). After analyzing the images in the database, we found that in most of the relevant images, these edifications were wrongly labeled as *castles*. This might be considered as a bad annotation, but according to the knowledge of the annotator, and even the kind of *churches* known by them, most of these buildings do not look as *churches*, and then, given that the suggestion is to assign the closest label, they labeled them in almost all cases as castles. As we can observe, the backgrounds of an annotator clearly affect annotation results.

We also observe that simply using the labels found in the sample images is not enough in a number of cases, where a more general concept is intended to be expressed. An example of this is topic 5 (one or more animals swimming in a body of water), where the general concept *animal* is not generalized from the three sample images, containing a whale, a bird and fish. Generalization, by taking advantage of the pseudo-ontology is a possible path to follow.

Using only high-level features is not enough for an important number of topics. Topic 6 (a straight road or highway in the United States of America) is an example of this, given that, although most of the retrieved images have similar contents to the sample and relevant images, the knowledge of which were taken in the United States of America cannot be included unless textual information is also considered. The same happens with Topic 40 (tourist sights in bad weather conditions), where "tourist sights" is a concept that cannot be defined using a label, and the information about the bad weather is not so clear from the sample images (examples of the retrieval for this topic are also shown in Table II).

An example of how important having a suitable label in the list of possible labels and having it in the query as well, is topic 50 (indoor photos of a church or cathedral), where P-20 is considerably high with respect to the average of the 39 topics, given that the label *church interior* is actually defined in the pseudo-ontology. An example of the opposite is topic 60 (one or more salt piles in a salt pan), where the most similar available label was *mountain*, which does not capture the meaning of the topic, and its poor retrieval rate proves it as inadequate.

Other factors such as human mistakes, not finding the best label, the size and shape of the object, shadows, image illumination, etc., should also be considered as relevant.

Finally, in Tables III and IV, we summarize a comparison of our method with the ones participating at ImageCLEF 2008 [7], in terms of MAP. For practical reasons we only show the 5 best results for both variations and show the hypothetical position of our method, compared to the total runs. From this, we find that we would rank in the 12^{th} position, compared to the 33 submitted visual methods; and in the 82^{nd} place, compared to the 400 submitted textual methods (with the best of them involving manual interaction).

It is important to mention that in literature, compared to the textual methods, visual approaches are typically less accurate; and given the fact that our method is based on information that is exclusively obtained from the image

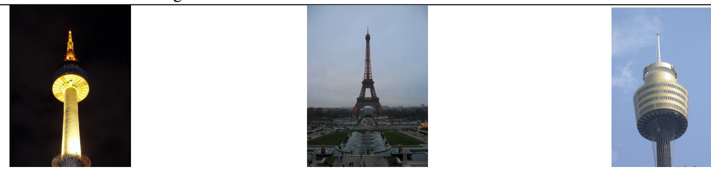
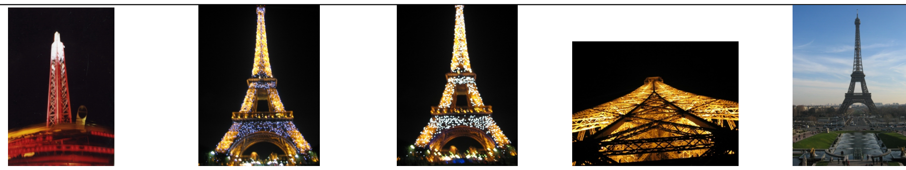
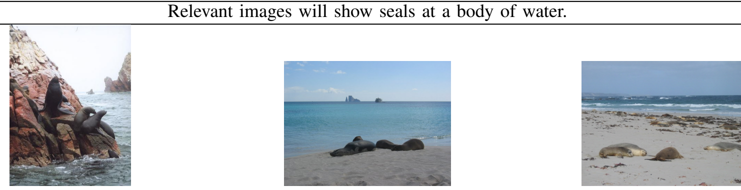
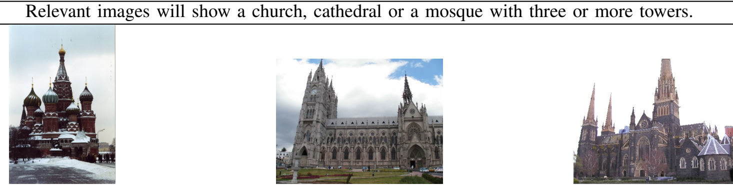
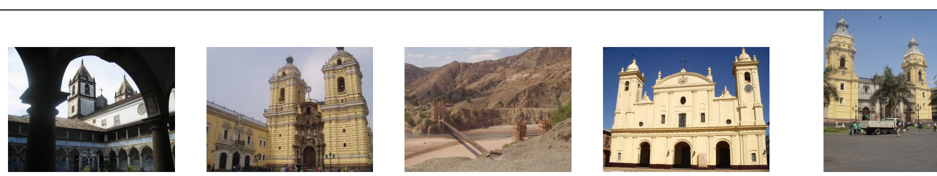
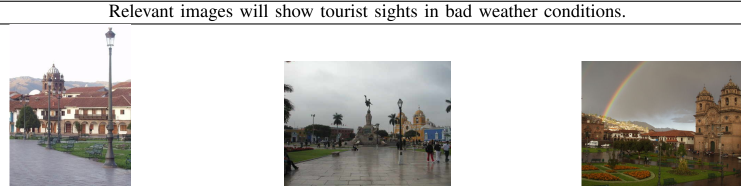
Text query	Relevant images will show famous television and telecommunication towers.
Visual query	
Top-5 retrieved	
Text query	Relevant images will show seals at a body of water.
Visual query	
Top-5 retrieved	
Text query	Relevant images will show a church, cathedral or a mosque with three or more towers.
Visual query	
Top-5 retrieved	
Text query	Relevant images will show tourist sights in bad weather conditions.
Visual query	
Top-5 retrieved	

Table II

SOME EXAMPLES OF THE RETRIEVAL FOR 4 OF THE 39 TOPICS. FOR EACH TOPIC WE SHOW THE TEXTUAL QUERY, THE VISUAL QUERY (I.E., THE 3 SAMPLE IMAGES), AND THE TOP-5 RETRIEVED IMAGES USING OUR METHOD. THE FIRST TWO EXAMPLES ARE OUR BEST RESULTS (MEASURED WITH AP, AND THE LAST TWO EXAMPLES ARE AMONG OUR WORST RESULTS, ALSO MEASURED WITH AP. THESE RESULTS WERE OBTAINED USING LA AND FIXING $\alpha = 0.1$.

Position	Group	MAP
1	NTU	0.2103
2	NTU	0.1875
3	XRCE	0.1502
4	XRCE	0.1329
5	XRCE	0.1317
12	**Our Method**	**0.1170**

Table III
COMPARISON OF THE MAP WITH RESPECT TO THE 33 VISUAL METHODS PRESENTED AT IMAGECLEFF 2008 [7] (THE POSITION SHOWN FOR OUR METHOD IS HYPOTHETICAL, SINCE WE DID NOT ACTUALLY PARTICIPATE). WE ONLY SHOW THE 5 BEST METHODS IN TERMS OF MAP.

Position	Group	MAP
1	PTECH	0.4283
2	DCU	0.3514
3	DCU	0.3158
4	Meiji	0.3011
5	Budapest-AC	0.2988
82	**Our Method**	**0.1170**

Table IV
COMPARISON OF THE MAP WITH RESPECT TO THE 400 TEXTUAL METHODS PRESENTED AT IMAGECLEFF 2008 [7] (THE POSITION SHOWN FOR OUR METHOD IS HYPOTHETICAL, SINCE WE DID NOT ACTUALLY PARTICIPATE). WE ONLY SHOW THE 5 BEST METHODS IN TERMS OF MAP.

(both, labels and spatial relations among regions) it could be considered as a visual method, independently from the fact that images are manually segmented and annotated; and the per-topic experiments also corroborate it performs better with visual queries. This might be another reason why the performance is not as good as expected.

Something that must be taken into account is that the state of the art methods we are comparing our results to, try to use sophisticated retrieval strategies, while our method is on the other hand basic, using simple fusion mechanisms and not taking advantage of image captions or the annotation pseudo-ontology. We expect that by using a more advanced list fusion, combining this method with textual information, and considering the pseudo-ontology, our results will have an important improvement. This potential gives us strong basis to suggest this as an innovative and promising image retrieval method.

VII. CONCLUSIONS

In this paper we modeled images and the relations among the objects in the image by means of conceptual graphs. We used this model for image retrieval and experimented different basic fusion schemata. This retrieval proved to be competitive and promising for future research. We can conclude, observing the results, that the use of spatial relations is a determinant factor in retrieval, obtaining the best results when most of the weight is given to them with respect to the concepts ($\alpha \approx 0.1$). However, on the other hand, when only spatial relations are used, they tend to be insufficient to represent the meaning of images and topics, and this produces poor results.

We observe a competitive performance of our method with respect to state of the art retrieval methods. We emphasize that finding the best fusion is not the goal of this research, and for such reason, more complex fusion methods are not explored. However, it is evident that even with the simple fusion schemata considered, there is a significant improvement in results, which shows that the more sample images, the better the results could be expected to be.

From the analysis we conclude that further work on the use of the pseudo-ontology for generalizing the concepts, must be done. Particularly, more complex fusion methods and the combination of this method with the captions associated to the images, are future work that we expect to significantly improve results.

REFERENCES

[1] R. Datta, D. Joshi, J. Li, James, and Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 39, pp. 1–60, 2008.

[2] A. Goodrum, "Image information retrieval: An overview of current research," *Informing Science*, vol. 3, p. 2000, 2000.

[3] A. W. M. Smeulders, S. Member, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1349–1380, 2000.

[4] M. Montes, A. Gelbukh, A. López-López, and R. Baeza-Yates, "Flexible comparison of conceptual graphs," in *Proceedings of the 12th International Conference and Workshop on Database and Expert Systems Applications*. Springer, 2001, pp. 102–111.

[5] M. Grubinger, "Analysis and evaluation of visual information systems performance," Ph.D. dissertation, School of Computer Science and Mathematics, Faculty of Health, Engineering and Science, Victoria University, Melbourne, Australia, 2007.

[6] H. Escalante, C. Hernández-Gracidas, J. González, A. López, M. Montes, E. Morales, L. Sucar, L. Villaseor, and M. Grubinger, "The segmented and annotated iapr-tc12 benchmark," *Computer Vision and Image Understanding, in press*, 2009.

[7] T. Arni, M. Sanderson, P. Clough, and M. Grubinger, "Overview of the imageclef 2008 photographic retrieval task," *Working Notes of the CLEF*, 2008.

[8] W. Ren, M. Singh, and S. Singh, "Image retrieval using spatial context," in *Proceedings of the 9th International Workshop on Systems, Signals and Image Processing*, 2002, pp. 44 – 49. [Online]. Available: citeseer.ist.psu.edu/535521.html

[9] V. Rathi and A. K. Majumdar, “Content based image search over the world wide web,” in *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, 2002. [Online]. Available: citeseer.ist.psu.edu/553765.html

[10] A. Collins and M. Quillian, “Retrieval time from semantic memory,” *Journal of Verbal Learning and Verbal Behavior*, vol. 8, no. 2, pp. 240–247, April 1969. [Online]. Available: http://dx.doi.org/10.1016/S0022-5371(69)80069-1

[11] C. Peirce, *The Collected Papers of Charles Sanders Peirce*. Harvard University Press, 1931.

[12] M. Bekhatir, Y. Chiaramella, and P. Mulhem, “A signal/semantic framework for image retrieval,” in *JCDL 005: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 2005, pp. 368–368.

[13] M. Mechkour, C. Berrut, and Y. Chiaramella, “Using a conceptual graph framework for image retrieval,” in *Proc. of the MMM95 (Multimedia Modeling) conference*, 1995, pp. 127–142.

[14] M. J. Egenhofer, “A formal definition of binary topological relationships,” in *Proceedings of the 3rd International Conference, FODO 1989 on Foundations of Data Organization and Algorithms*, 1989, pp. 457 – 472.

[15] C. Hernández-Gracidas and L. Sucar, “Markov random fields and spatial information to improve automatic image annotation,” in *Proceedings of the Pacific-Rim Symposium on Image and Video Technology*, 2007.

[16] K. S. Jones, “A statistical interpretation of term specificity and its application in retrieval,” *Journal of Documentation*, vol. 28, pp. 11–21, 1972.