

# Markov Random Fields and Spatial Information to Improve Automatic Image Annotation

Carlos Hernández-Gracidas and L. Enrique Sucar

National Institute of Astrophysics, Optics and Electronics,  
Luis Enrique Erro #1, Sta. María Tonantzintla, Puebla, México,  
carloshg@ccc.inaoep.mx, esucar@inaoep.mx,  
WWW home page: <http://ccc.inaoep.mx>

**Abstract.** Content-based image retrieval (CBIR) is currently limited because of the lack of representational power of the low-level image features, which fail to properly represent the actual contents of an image, and consequently poor results are achieved with the use of this sole information. Spatial relations represent a class of high-level image features which can improve image annotation. We apply spatial relations to automatic image annotation, a task which is usually a first step towards CBIR. We follow a probabilistic approach to represent different types of spatial relations to improve the automatic annotations which are obtained based on low-level features. Different configurations and subsets of the computed spatial relations were used to perform experiments on a database of landscape images. Results show a noticeable improvement of almost 9% compared to the base results obtained using the k-Nearest Neighbor classifier.

**Key words:** Spatial relations, Markov random fields, automatic image annotation, content-based image retrieval.

## 1 Introduction

Considerable amounts of digitally stored visual information are available for their use in a number of different applications. Regarding this information, it is a frequent necessity to retrieve image subsets which fulfill certain criteria, in most of the cases concerning the visual contents of the image itself. Objects with specific physical characteristics, performing a given action or in a given position, are some examples of possible queries for image retrieval. Also, a desirable feature is the ability to retrieve images where the objects interact in a particular way, which is an even more complicated form of query. Unfortunately, most state of the art image retrieval systems are based on low-level features like color, texture, shape, or on the other hand, based on image captions assigned by humans. In the first case, retrieval is ineffective due to the lack of semantic information coming from the image; in the second case, often better results are obtained, but with the need of manual annotations; and for huge databases, manual annotation is a time consuming task which cannot always be performed correctly.

Content-based image retrieval (CBIR) is the use of computer vision to analyze the actual contents of images (by using their color, texture, shape or any other information derived from the images themselves), applied to the retrieval of images by their semantics. Spatial relations are useful to know the relative position of an object in a scene by using other objects in the same scene as reference.

It seems almost obvious that by applying spatial information CBIR will automatically improve results, but the interesting questions are: How to do it? Which of all the possible relations can be useful? It is important to notice that we do not suggest that spatial information will suffice to obtain an efficient image retrieval, on the contrary, this research is intended to encourage the use of it as a complementary source of key information. The number of fields where spatial relations could be applied is by itself an important motivation. A few examples are: medical imagery analysis, geographic information systems (GIS) and image retrieval.

In this paper we follow an approach based on Markov random fields (MRFs) to represent the information about the spatial relations among the regions in an image, so the probability of occurrence of a certain spatial relation between each pair of labels could be used to obtain the most probable label for each region, i.e., the most probable configuration of labels for the whole image. Spatial information extracted from training images is fused with “expert” knowledge to represent the information coming from the neighbors. Spatial relations are divided in this study in three groups: topological relations, horizontal relations and vertical relations. Experiments with each of these groups incorporated individually and with the three groups used at the same time were performed in order to determine their relevance in the final results. Different configurations were also used in the experiments. Results were obtained on a database of landscape images and they show a noticeable improvement of almost 9% compared to the base results obtained using the k-Nearest Neighbor (kNN) classifier. Since this work is proposed as an improvement to a basic classification algorithm, it is expected that if the annotation algorithm used provides better results, they can be improved as well and an even higher accuracy can be reached.

The structure of this paper is as follows. Section 2 reviews basic concepts on image segmentation, spatial relations and MRFs. Section 3 summarizes related work in this field. Section 4 presents the methodology followed. Section 5 describes how the experiments were performed and the results obtained. Finally, in section 6 we present our conclusions and the future research to be done.

## 2 Fundamentals

In this section we present definitions and basics of automatic image segmentation, automatic image annotation, spatial relations and Markov random fields.



**Fig. 1.** Example of the results of automatic image segmentation using Normalized cuts. Left: an oversegmented image. Right: important objects are incorrectly segmented in the image.

## 2.1 Automatic Image Segmentation

Segmenting an image is partitioning that image into several regions, which are determined by their local features like color and texture. In general, automatic segmentation algorithms like Normalized cuts [13] usually tend to produce erroneously delimited regions with results like the ones shown in Figure 1. In the first image, the elephant and the grass are oversegmented, providing more segments than the necessary; in the second case, important objects in the image are incorrectly segmented, making of this an almost useless segmentation. These errors affect directly the performance of automatic annotation algorithms. However, it is important to mention that the emphasis of our work is not on segmentation improvements.

## 2.2 Automatic Image Annotation

Automatic image annotation (AIA) is a process that has been commonly used to support image retrieval, though results are not quite accurate at this time. Automatic image annotation is the task of automatically assigning annotations or labels to images or segments of images, based on their local features. Given the size of most image databases, image annotation is frequently performed by automatic systems, and this task, though necessary, is currently poorly performed given the difficulty and complexity of the extraction of adequate features which allow to generalize and distinguish an object of interest from others with similar visual properties. Erroneous labeling of regions is a common consequence of the lack of a good characterization of the classes by low-level features.

## 2.3 Representing Spatial Relations

Spatial relations provide relevant high-level information about the way elements interact in a scene. They are useful to know the relative position of an object in a scene with respect to other reference objects.

Given two objects of the classes **A** and **B**, it is feasible to think that, depending on their kind, object of the class **A** can relate to object of the class **B**

in a limited number of ways, and that some spatial relations will not be valid between these two objects. If we assume that most of the automatically assigned annotations are correct, then it is feasible to correct the mistaken ones. This binary method of evaluating region annotations by their spatial relations, classifying them as *valid* or *not valid*, can be extended by means of probabilities. In this way, the more frequent a relation between objects of the class **A** and objects of the class **B**, the higher the associated probability will be; conversely, the less frequent a relation between **A** and **B**, the closer its probability value will be to 0. This probability measure allows to obtain a global optimal configuration, i.e., the set of annotations (one for each region) which according to the spatial relations among regions, provides the highest probability of being globally correct in the image. It is important to notice that the spatial relations are independent from the annotations, and consequently they are not affected by the quality of such annotations. The more correct annotations we have, the more reliable our corrections will be. Same as with the number of annotations, the more objects, and the more relations among these objects, the more information about coherence of these annotations can be inferred.

Spatial relations are divided into:

1. Topological relations: They are preserved under rotation, scaling and translation. Examples of them are: *overlapped* and *contained by*.
2. Order relations: They are preserved under scaling and translation but change under rotation. They are based on the definition of order. Some examples are: *above-below* and *left-of-right of*.
3. Metric relations: They change under scaling but are unaffected by rotation and translation. They measure distance and direction. Some examples are: *2 miles away* and *30 meters around*.
4. Fuzzy relations: They are measured in vague terms, and consequently, are difficult to quantify. Examples of them are: *near* and *far*.

Spatial relations provide important information in domains such as GIS, Robotics and CBIR; where the location of an object implies knowledge about the geographic position of a certain place, the possible path to follow by a robot or the contents of an image to be retrieved.

## 2.4 Markov Random Fields

Markov Random Fields [10, 15] are probabilistic models which combine *a priori* knowledge given by some observations, and knowledge given by the interaction with neighbors.

Let  $F = \{F_1, F_2, \dots, F_n\}$  be random variables on a set  $S$ , where each  $F_i$  can take a value  $f_i$  in a set of labels  $L$ . This  $F$  is called a random field, and the instantiation of each of these  $F_i \in F$  as an  $f_i$ , is what is called a configuration of  $F$ , so, the probability that a random variable  $F_i$  takes the value  $f_i$  is denoted by  $P(f_i)$ , and the joint probability is denoted as  $P(F_1 = f_1, F_2 = f_2 \dots, F_n = f_n)$ .

A random field is said to be an MRF if it has the properties of *positivity* and *markovianity*. The joint probability can be expressed as

$$P(f) = \frac{e^{-U_p(f)}}{Z} \quad (1)$$

where  $Z$  is called the partition function or normalizing constant, and  $U_p(f)$  is called the energy function.

The optimal configuration is found by minimizing the energy function  $U_p(f)$  obtained by

$$U_p(f) = \sum_c V_c(f) + \lambda \sum_o V_o(f) \quad (2)$$

$V_c(f)$  and  $V_o(f)$  are called the potential functions, where  $V_c(f)$  stands for the information coming from the neighboring nodes,  $V_o(f)$  represents the information coming from the observations, and  $\lambda$  is a constant used to weight the observations and the information from the neighbors, giving each a relative importance with respect to the other. The optimal configuration is obtained when the value of  $U_p(f)$  with minimal energy is found for every random variable in  $F$ .

An MRF can also be seen as an undirected graph  $G = (V, E)$ , where each vertex  $v \in V$  represents a random variable, and each edge  $u, v \in E$  determines that nodes  $u$  and  $v$  are neighbors.

### 3 Related Work

In this section we give a general perspective on how spatial relations are used in previous works and how they are applied on image retrieval and other tasks.

One of the first attempts to consider spatial information (combined with temporal information) is introduced by Allen [1]. Several topological models have been developed, from which the most used are the 4-Intersection Model [5] and its extensions, such as the 9-Intersection Model [5] and the Voronoi-Based 9-Intersection Model [4].

A deductive system for deriving relations in images is introduced by [14] where a set of rules is proposed to deduce new relations from a basic set of relations. This approach can be used to extend queries when images are searched in a retrieval system. This is intended to complement text-based search systems, but assumes spatial relations are somehow existent in the image since it provides no form of computing them from images. The system of rules is shown to be complete for the 3D case but incomplete for 2D. Studies like [17, 18] focus on the problem of using spatial relations in CBIR related tasks. Basic image retrieval systems using some kind of spatial information are shown in [7, 11, 12]. In [7] another method to extend textual information is proposed. They suggest a way to complement image annotations by semi-automatically adding spatial information about annotated objects in the image. A human needs to be involved since the objects are assumed to be well segmented and well annotated; then spatial relations are computed and annotations are complemented with this information. They provide a study about the relative relevance of spatial relations

based on the way people tend to use them. In [11] a system for image retrieval using spatial information as a complementary element is shown. They implement a web crawler which uses textual information from the web pages retrieved and from the image names, and complements it with low-level features like color and high-level features like spatial information. However, as in other similar works, human interaction is necessary to recognize objects and adequately label them so the images can be stored in the database, which significantly limits the usefulness of this search tool. Queries are performed on the database by using hand sketches or sample images and human interaction is also required in this process to determine objects and their relevance in the image before the query is processed. In [12] a retrieval system based on a set of basic spatial relations is proposed using a matching system to measure similarity between a pair of images and using automatic image segmentation and annotation. They propose the use of six spatial relations and show experiments using a limited set of labels and images where objects like grass, clouds, trees and sky are contained.

The already existent methods are insufficient to apply them directly in the solution of the problem of CBIR, nor are they suitable for our purpose of improving AIA. The reason is that these methods focus mostly on topological relations and other important spatial relations which provide interesting information are usually discarded. In the few cases where non-topological relations are considered, they are used in a very basic way, like simple image comparison.

## 4 Improved Image Annotation

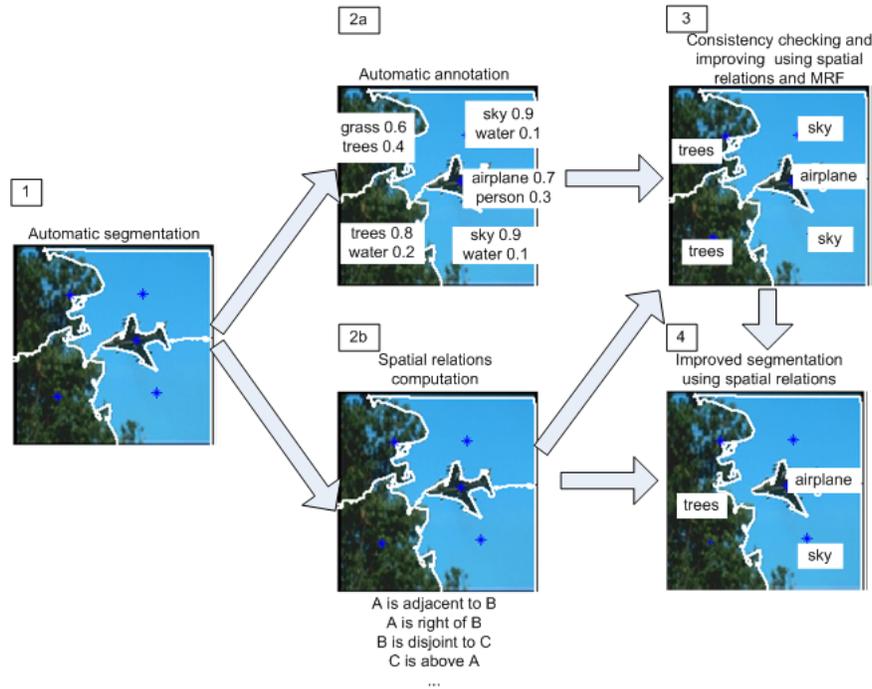
In this work we make use of an automatic segmentation system to divide images into regions, and an automatic annotation system to assign potential labels to each region. These regions and labels are validated by means of the spatial relations among the regions themselves, and if that is the case, modified to harmonize within the image. We claim that we can improve AIA by iteratively validating and correcting these intermediate processing steps.

The methodology, depicted in Figure 2, is the following:

1. The image is automatically segmented (using Normalized cuts).
2. The obtained segments are assigned with a list of labels and their probabilities (computed with the kNN algorithm). Concurrently, the spatial relations among the same regions are computed.
3. The labels are checked for consistency by using spatial information based on MRFs. The labels with Maximum A-Posteriori Probability (MAP) are obtained for each region.
4. Adjacent regions with the same label are joined.

As mentioned before, this is an iterative process, and steps 2 to 4 may be repeated until the system stabilizes.

The spatial relations considered in this work are shown in Table 1. Considering the image as a graph, where each node of the graph represents a region, and the spatial relations are represented by edges joining these nodes, then the



**Fig. 2.** Block diagram of the proposed methodology. In 1 the image is segmented with the Normalized cuts algorithm; in 2a each region is assigned with a list of labels and an associated probability for each label, at the same time, in 2b spatial relations among regions are computed; in 3 the labels are improved with the use of MRFs and spatial information; finally, in 4, and using these improved labels and the adjacency relations, if two or more adjacent regions have the same label they are joined.

relations can also be divided into: *directed* and *undirected*. Table 1 also shows this separation. Observe the distinction made between the horizontal relation *beside* (which is a generalization of *left* and *right* in order to convert these two originally directed relations into a single undirected relation), and the use of *above* and *below* as separate relations. The reason for this decision is that for the kind of images to be analyzed (landscapes, as in the Corel data set), knowing if an object is left or right of another object is irrelevant, but knowing if it is above or below is considered to give important information about the coherence of the image annotations. This is not the case for every application domain, since, for example, in the case of medical imagery, knowing if a certain organ is left, right, above or below another, certainly gives crucial information for the interpretation of the image.

These relations can be grouped in 3 sets: topological relations, horizontal relations and vertical relations. In the case of horizontal relations and vertical relations we separate order relations into these two groups, given the fact that

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

**Table 1.** Spatial relations used in this work. Relations are divided as: topological, horizontal and vertical; and also as directed and undirected.

an object can be related to another in both ways (for example, an object **A** can be at the same time above and beside another object **B**).

An important feature of these three sets is that for each there will be one and only one relation between every pair of regions, since in each group the relations are exclusive among the rest (an object **A** cannot be above and below an object **B** at the same time). In our notation we use the term XOR ( $\oplus$ ) to represent this characteristic, meaning that only one of the values in the corresponding group is taken into account, but no more than one at the same time. It must be remembered that in this particular case, XOR does not mean a logic XOR, but a notation for value exclusion.

Given that for each group there will be one and only one relation, we can infer that each group defines by itself a complete graph, i.e., a graph where all regions are connected to all others by exactly one edge representing a relation.

There are some obstacles for the direct use of MRFs in the solution of this problem. The first one is that in traditional MRFs, every edge determining vicinity must be an undirected edge, and some important relations are directed by nature, and though they can be generalized in an undirected fashion, important information may be lost in the process (like in the *above* and *below* cases), so it would be desirable to be able to keep such information; the second obstacle is the presence of more than one relation between each pair of regions, since MRFs are defined by at most one edge connecting any pair of nodes. These points force to extend the use of MRFs in order to adequate them to be used in this kind of application.

If we structure the spatial relations as follows, we can provide a new energy function  $U_p(f)$  using these different relations at the same time. Each  $R_{ij}$  represents a spatial relation in one of the groups:

$$\begin{aligned}
R_{ij}(T) &\in \{1, 2\} \text{ --- } \textit{Adjacent}, \quad \textit{Disjoint} \\
R_{ij}(H) &\in \{3, 4\} \text{ --- } \textit{Beside}, \quad \textit{Horizontally Aligned} \\
R_{ij}(V) &\in \{5, 6, 7\} \text{ --- } \textit{Below}, \quad \textit{Above}, \quad \textit{Vertically Aligned}
\end{aligned}$$

Using these three groups of relations, the energy function is:

$$U_p(f) = \alpha_1 V_T(f) + \alpha_2 V_H(f) + \alpha_3 V_V(f) + \lambda \sum_o V_o(f) \quad (3)$$

where  $V_T$ ,  $V_H$  and  $V_V$  are potential functions computed from the topological, horizontal and vertical relations, respectively. They can be obtained as inversely proportional to the sum of the probabilities in each group of relations.

In the energy formula in Equation 2, only lambda ( $\lambda$ ) is used for the purpose of weighting the relative value of  $V_c$  and  $V_o$ . The use of three alphas ( $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$ ) in Equation 3, allows to give a different weight to each group of relations, with the premise that they do not have the same relevance. Given the fact that the best configuration will be the one giving the lowest energy value, the more relevant a relation, the higher its associated alpha value.

We define now the energy functions. In these functions, each  $P_{kc}$  represents the probability of relation  $k$ . The XOR operators depend on the value taken by the  $R_{ij}$  related.

$$\begin{aligned} V_T(f) &= \frac{1}{\sum_c P_{1c}(f) \oplus P_{2c}(f)} \\ V_H(f) &= \frac{1}{\sum_c P_{3c}(f) \oplus P_{4c}(f)} \\ V_V(f) &= \frac{1}{\sum_c P_{5c}(f) \oplus P_{6c}(f) \oplus P_{7c}(f)} \end{aligned}$$

In order to compute each  $P_{kc}$ , a combination of information extracted from training images and “expert” knowledge is used. First, training images which are already segmented and manually labeled, are examined to determine the presence of spatial relations and their frequency. For each relation,  $k$ , a matrix  $Rel_k(i, j)$  of  $D \times D$  components is created, where  $D$  is the number of possible labels for a region. An equivalent matrix  $E_k(i, j)$  is created for each relation incorporating a priori knowledge of each relation based on subjective estimates. These two estimates are linearly combined as follows [9]:

$$P_{kc}(f) = \frac{Rel_k(i, j) + \delta E_k(i, j)}{NR(i, j) + \delta 100} \quad (4)$$

where  $\delta$  is a constant used to determine the relevance of expert knowledge with respect to the training data, and  $NR(i, j)$  is the number of times labels  $i$  and  $j$  appeared together in an image. The use of  $E_k(i, j)$  also serves as a smoothing technique to minimize the number of 0’s in the original  $Rel_k(i, j)$  matrices. To obtain the “best” configuration, the MRFs are solved by computing the MAP using simulated annealing [8] with temperature ( $T$ ) decremented as follows:

$$T = \frac{T}{\log(100 + j)} \log(100) \quad (5)$$

where  $j$  is the number of the iteration in the MRFs.

airplane	grass	mountains	sky
bird	ground	pilot	snow
boat	horse	road	trees
church	house	rock	water
cow	lion	sand	
elephant	log	sheep	

**Table 2.** The set of labels for CoreLA.

## 5 Experiments and Results

To evaluate our method we used the CoreL database, and particularly, the subset CoreLA<sup>1</sup> developed by [3] and consisting of 205 images. This data set was divided into 137 training images and 68 test images. This database portraits landscapes, containing mainly elements in a set of 22 different possible annotations, which are shown in Table 2.

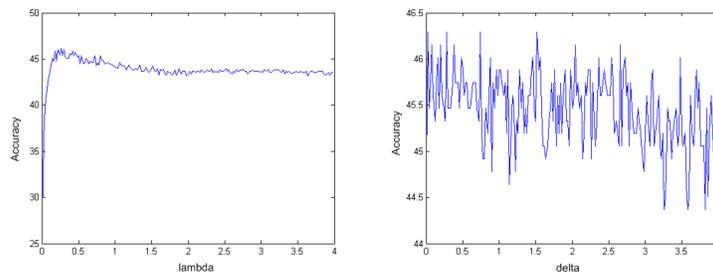
The advantage of using this database is that besides being already segmented, it counts with labeled regions, so the time-consuming task of hand-labeling the regions for the experiments was already performed, allowing for experimentation only on image annotation. The experiments were performed using as annotation system a kNN classifier as implemented in [6], which is a simple but efficient instance-based learning algorithm.

We must mention the special case when a region is *unknown* (label 0). These values affect our spatial approach since no spatial information can be obtained from them. If the region being examined is *unknown*, values of  $P_{1c}$  to  $P_{4c}$  are set to  $\frac{1}{2}$ , and values of  $P_{5c}$  to  $P_{7c}$  are set to  $\frac{1}{3}$ , to reflect the lack of a priori knowledge coming from neighbors in such cases.

Experiments were performed dividing them in four groups: tests using only one group of spatial relations individually, and tests using the three groups of spatial relations simultaneously. For each of these groups three variations were also tested: the use of no smoothing technique, the use of a simple Laplacian smoothing, and the use of expert knowledge as initialization and smoothing technique.

Several parameters had to be set:  $\lambda$ ,  $\delta$ ,  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ ,  $T$  and the number of iterations ( $n$ ) for the MRFs. Figure 3 shows the effect of changing  $\lambda$  and  $\delta$  values using our approach; considering the three groups of spatial relations and fixing the remaining parameters to their optimal values. We can see how important the prior observations are, since setting  $\lambda=0$  makes annotation rates fall to approximately 25.3%. However, a value of  $\lambda$  set too high makes rates oscilate around 43.5%. The highest annotation rates were reached with  $\lambda$  set to values close to 0.25, and this is the value used for the experiments. Experimental tests showed the necessity of an initialization either by an expert or by a Laplacian smoothing; although it is not clear what the ideal value for  $\delta$  would be (as it is shown, also in Figure 3). When  $\delta$  was set to a value of 0.25, the best annotation rates were

<sup>1</sup> Available at <http://www.cs.ubc.ca/~pcarbo/>



**Fig. 3.** Left: the variation of annotation accuracy with values of  $\lambda$  in the interval  $(0,4)$  and incremented by 0.02. Right: the variation of annotation accuracy with respect to  $\delta$  values in the same interval and with the same increments.

reached, so, this is the value used for the experiments. The idea behind using  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  was to be able to give a different relative weight to each relation group with the premise that they will have different relevance in the annotation result (which is partially confirmed by our experiments). However, setting these values requires more investigation, and for the experiments they were equally fixed to a value of 1, leaving for a future research their estimation using a more sophisticated heuristic. Values for  $T$  and  $n$  were also set by experimentation, fixing them to 116 and 400, respectively. The approximate execution time with this value of  $n$  is of 35 seconds for the 68 test images.

Experimental results are shown in Table 3. To calculate the accuracy of the method, we ran 10 times each test and proceeded to eliminate “peaks”, i.e., we discarded the highest result and the lowest result and obtained the average of the 8 remaining results.

These experiments show no significant difference between an expert initialization and a Laplacian smoothing, probably because these values were not well estimated by the expert or because the training images contained already sufficient information.

Individual experiments using only one group of relations at a time, show that the more significant results are obtained when vertical relations are used, which proves the usefulness of these undirected relations on this particular domain, with an improvement of 8.12% and a relative improvement of 22.06%. Experiments show that the individual use of each group of spatial relation certainly improves results, but the highest accuracy rates are reached when they are used simultaneously, showing in the best case an improvement of 8.82% and a relative improvement of 23.97%.

As a way of comparing our results with other methods, we show in Table 4 a comparison with three state of the art automatic annotation methods. These results are obtained with the code developed by [3]. It can be observed that our base result is similar in performance to those we are comparing to, but at the same time, our proposed method is at least 9% more accurate than any of them.

Algorithm	Relation group	Smoothing	Accuracy	Improvement	Rel. improvement
kNN	None	None	36.81%	-	-
MRFs	Topological	None	42.72%	5.91%	16.04%
		Laplacian	43.51%	6.70%	18.19%
		Expert info.	43.25%	6.44%	17.49%
	Horizontal	None	41.72%	4.91%	13.34%
		Laplacian	43.08%	6.27%	17.02%
		Expert info.	43.58%	6.76%	18.38%
	Vertical	None	43.73%	6.92%	18.80%
		Laplacian	44.93%	8.12%	22.06%
		Expert info.	44.88%	8.07%	21.92%
	All	None	43.29%	6.47%	17.58%
		Laplacian	45.41%	8.60%	23.37%
		<b>Expert info.</b>	<b>45.64%</b>	<b>8.82%</b>	<b>23.97%</b>

**Table 3.** Results obtained with the use of MRFs and spatial information with the different groups of relations and smoothing types for the test images in the CoreLA database. The best results were obtained when the three groups of spatial relations are used together (in bold).

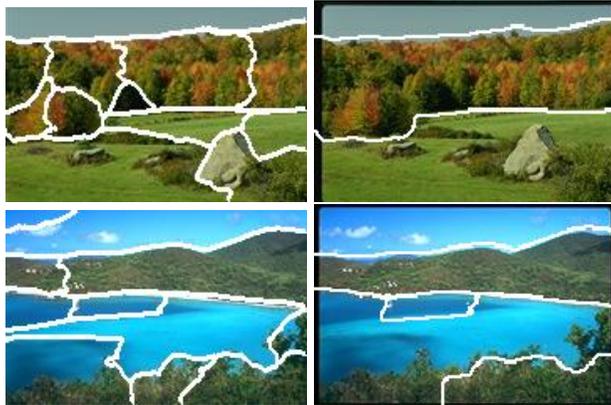
Although improving segmentation was not our main goal, we performed a simple experiment with interesting results. After improving the initial annotations with our method, adjacent regions with the same label were joined. A couple of examples of this are shown in Figure 4. We found that several other image segmentations were improved the same way.

## 6 Conclusions and Future Work

We proposed a novel methodology to improve AIA based on spatial relations and MRFs. It combines several types of spatial relations under an MRF framework, where the potentials are obtained using subjective estimates and real data. We concluded that it is feasible to apply spatial relations to improve automatic image annotation systems. Our experiments show that an important number of

Algorithm	Accuracy
gML1[3]	35.69%
gML1o[3]	36.21%
gMAP1[3]	35.71%
gMAP1MRF[3]	35.71%
kNN	36.81%
<b>MRFs AREK</b>	<b>45.64%</b>

**Table 4.** Comparison of our results with other methods. The last two rows show our base result and our best result using MRFs and all the spatial relations with expert knowledge (MRFs AREK).



**Fig. 4.** Some examples of the result of joining regions after their annotations are improved. Left: the originally segmented images. Right: the resulting segmentations after joining adjacent regions with the same labels.

the labels are corrected by using MRFs and the spatial relations among regions. In experiments with the CoreLA database, a significant improvement of almost 9% and a relative improvement of almost 24% were obtained with respect to the original annotations. Further experiments must be performed to clarify the relevance of each relation group and also to evaluate the advantage of using expert estimations. Also, a more sophisticated way of determining optimal values for parameters, like the use of evolutionary methods [2], is a possible future line of research. The iteration of steps 2 to 4 of the method as we suggested, should also provide better results.

Using a different annotation algorithm is an alternative that might provide better results. An interestingly different approach to the use of MRFs for more than one relation would be the use of interacting MRFs [16], and finding the way this interaction should be performed represents a motivating challenge. The application in other image domains may show the generality of the method here proposed, and should also confirm our hypothesis that the importance of some relations varies depending on the domain. Medical imagery, GIS, automatic robot navigation, are some of the potential future applications of our method.

## References

1. Allen, J. F.: Maintaining Knowledge About Temporal Intervals. *Communications of the ACM*. **26** 11 (1983) 832–843
2. Bäck, T.: *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford Univ. Press., Oxford, UK (1996)

3. Carbonetto, P.: Unsupervised Statistical Models for General Object Recognition. The Faculty of Graduate Studies, Department of Computer Science, MSc Thesis, The University of British Columbia, West Mall Vancouver, BC Canada (2003)
4. Chen, J., Li, Z., Li, C., Gold, C.M.: Describing Topological Relations with Voronoi-Based 9-Intersection Model. *International Archives of Photogrammetry and Remote Sensing* **32** 4 (1998) 99–104
5. Egenhofer, M.J., Sharma, J., Mark, D.M.: A Critical Comparison of the 4-Intersection and 9-Intersection Models for Spatial Relations: Formal Analysis. In: McCaster R.B., Armstrong M. (eds.): Eleventh International Symposium on Computer-Assisted Cartography, Auto-Carto 11, Minneapolis, Minnesota, USA (1993) 1–11
6. Escalante, H.J., Montes, M., Sucar, L.E.: Word Co-occurrence and Markov Random Fields for Improving Automatic Image Annotation. In: Rajpoot N.M, Bhalerao A. (eds.): Eighteenth British Machine Vision Conference, BMVC '07, Vol. 2, Warwick, UK (2007) 600–609
7. Hollink, L., Nguyen, G., Schreiber, G., Wielemaker, J., Wielinga, B., Worring, M.: Adding Spatial Semantics to Image Annotations. Fourth International Workshop on Knowledge Markup and Semantic Annotation, SemAnnot '04, Hiroshima, Japan: Springer LNCS (2004)
8. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by Simulated Annealing. *Science*. **220** 4598 (1983) 671–680
9. Neapolitan, R.E.: Probabilistic Reasoning in Expert Systems. Wiley, New York, NY, USA (1990)
10. Preston, C.J.: Gibbs States on Countable Sets. Cambridge University Press, Cambridge, UK (1974)
11. Rathi, V., Majumdar, A.K.: Content Based Image Search over the World Wide Web. In: Chaudhuri, S., Zisserman, A., Jain, A.K., Majumder, K.L. (eds.): Third Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP '02, Ahmadabad, India (2002)
12. Ren, W., Singh, M., Singh, S.: Image Retrieval Using Spatial Context. In: Wellstead, P. (ed.): Ninth International Workshop on Systems, Signals and Image Processing, IWSSIP '02, Vol. 18, Manchester, UK, (2002) 44–49
13. Shi, J., Malik, J.: Normalized Cuts and Image Segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **22** (2000) 888–905
14. Sistla, A. P., Yu, C., Haddack, R.: Reasoning about spatial relationships in picture retrieval systems. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.): Twentieth International Conference on Very Large Data Bases, VLDB '94, Santiago, Chile (1994) 570–581
15. Spitzer, F.L.: Random Fields and Interacting Particle Systems. Mathematical Association of America, (1971)
16. Wellington, C., Courville, A., Stentz, A.: Interacting Markov Random Fields for Simultaneous Terrain Modeling and Obstacle Detection. In: Thrun, S., Sukhatme, G.S., Schaal, S. (eds.): Robotics: Science and Systems, RSS '05, Boston, USA (2005) 1–8
17. Zhang, Q., Yau, S.S.T.: On Intractability of Spatial Relationships in Content-Based Image Database Systems. *Communications in Information and Systems*. **4** 2 (2004) 181–190
18. Zhang, Q., Chang, S., Yau, S.S.T.: On Consistency Checking of Spatial Relationships in Content-Based Image Database Systems. *Communications in Information and Systems*. **5** 3 (2005) 341–366