

2000-01-01

A Scene Similarity Metric for Matching Configurations of Image Objects

James Carswell

Technological University Dublin, james.carswell@tudublin.ie

Follow this and additional works at: <https://arrow.tudublin.ie/dmcart>



Part of the [Computer Sciences Commons](#)

Recommended Citation

Carswell, J. (2000) A scene similarity metric for matching configurations of image objects. *IAPRS*, Vol. XXXIII, Amsterdam, 2000.

This Article is brought to you for free and open access by the Digital Media Centre at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 License](#)
Funder: National Science Foundation

A SCENE SIMILARITY METRIC FOR MATCHING CONFIGURATIONS OF IMAGE OBJECTS

Peggy Agouris, Michela Bertolotto, James D. Carswell, Anthony Stefanidis

The University of Maine, USA
Dept. of Spatial Information Science and Engineering
{peggy,michelab,carswell,tony}@spatial.maine.edu

KEY WORDS: Image matching, Information extraction, Query, Raster, Spatial databases

ABSTRACT

This paper proposes a novel solution to querying image databases by matching raster features to imagery completely in the raster/spatial domain using the shape of single features together with their spatial relations as the matching primitives. The core feature-based matching module combines a revised least-squares matching algorithm, to accomplish the matching process on binary images, with a unique implementation of a feature library that organizes and links query objects with their images in the database, thus enabling fast and efficient real-time retrieval of relevant imagery. The purpose is to extend this previous work of ours, which matches on the shape of single features only, to include querying configurations of image objects. As such, both spatial reasoning, e.g. the topological (disjoint, touching, overlapping, etc.), directional (north, south, east, etc.), and metric (distance) relationships between individual features, together with feature shape matching are requirements for proper image retrieval. More specifically, we will define a methodology to evaluate similarity between a user-defined query scene and the retrieved images. To do this, it is required to define a function S_{met} that assesses the *similarity metric* between a query configuration Q and an image I in the database. Such a function will combine different similarity functions for individual object shapes, and the spatial relations between them. The result of this investigation will be a proposal for solving the problem of scene similarity querying in image databases that will provide the basis for future work on implementation and act as the framework for further extensions into temporal reasoning.

1 INTRODUCTION

Since the advent of digital scanners and sensors, beginning in the late 1970's, image database querying has become a major area of research. Most of the efforts during this time have focused on analyzing and comparing the more general properties of digital imagery. These include: color, in the form of histogram matching; texture, in the form of image coarseness and contrast matching and; composition, where an image is divided into homogeneous regions of color or texture and the relative positions of these regions analyzed [Carson et al., 1997; Flickner et al., 1995; Forsyth et al., 1996; Frankel et al., 1996; Pentland et al., 1996; Sclaroff et al., 1997].

The expression "image retrieval by content" in this paper means to retrieve images matching the higher level image characteristics, more specifically, actual shape information (synonymous with outlines or edges) of features and their configurations contained within the imagery, e.g. the outlines of a building or a particular arrangement of buildings. Furthermore, we are interested in doing so completely in the raster/spatial domains. This distinction is important as both the query sketch and the database images are stored in the same raw raster format without any vectorizing of the image-objects (feature digitizing) performed or other such image-object information known beforehand. The spatial component also implies that there has not been any processing applied to the imagery (such as Fast Fourier Transforms) that transforms the raster image into the frequency domain before further operations begin.

Our purpose is to extend the recently completed work on Image Query-by-Sketch [Agouris et al., 1999b], designed to match on single features completely in the raster domain, to include matching configurations of image objects. Thus we provide a comprehensive solution that fully describes the similarity between a given query scene and an image in the database.

2 MATCHING CONFIGURATIONS OF IMAGE-OBJECTS

An image is composed of spatial objects and their relations with the similarity between scenes described as a function of object similarity plus relation similarity [Goyal and Egenhofer, 2000]. But before scenes can be compared for their similarity properly, they have to be "normalized" so that all comparisons can begin from the same basic starting point. For example, for each "raw" image, one must find its "computational" counterpart; the ultimate "goal" in image understanding research. Therefore, the basic problem is to determine a methodology that will automatically enumerate

all the possible spatial relationships between image objects, or in other words, automatically transform each raw image into its computational image before their similarity can be properly compared.

A typical query of an image database consists of metadata, semantic data, and a sketched configuration of image-objects [Agouris et al., 1999a]. The metadata criterion could include information on image scale or location, etc., the semantic criterion could specify what type (or purpose, etc.) of objects should be contained within the images returned from the query, and the sketched configuration of image-objects will determine scene matching similarity (Figure 1).

For example, if the user decided to retrieve all images with airplanes, airplane hangers and runways that match to a particular configuration, the query would:

- Process the metadata library and retrieve all images that match to the specified metadata criterion (e.g. scale, or location, etc.).
- Follow the links from this subset of images to the semantic library where it would identify which of these images contained airplanes, hangers, and runways of any size and shape;
- Use the semantic input criteria to further narrow down the list of images to search within for the required query shapes;
- Use this further subset of imagery to select, through feature linking, a refined subset of features within the feature library to match individual objects in the query sketch against;
- Retrieve all the imagery linked to the best matched features in this refined subset from the feature library;
- Determine which of these images were returned more than once, i.e. which images contain all of the objects in the query configuration;
- Process the spatial relations of the query scene on this final subset of imagery and return a prioritized list of imagery as the query result.

Configurations can be matched within the query environment proposed in [Carswell, 2000] by adding an extension to the query building interface. What is required is that the individual objects within the sketch are recognized as such, and the pixels that make up their respective edges stored independently.

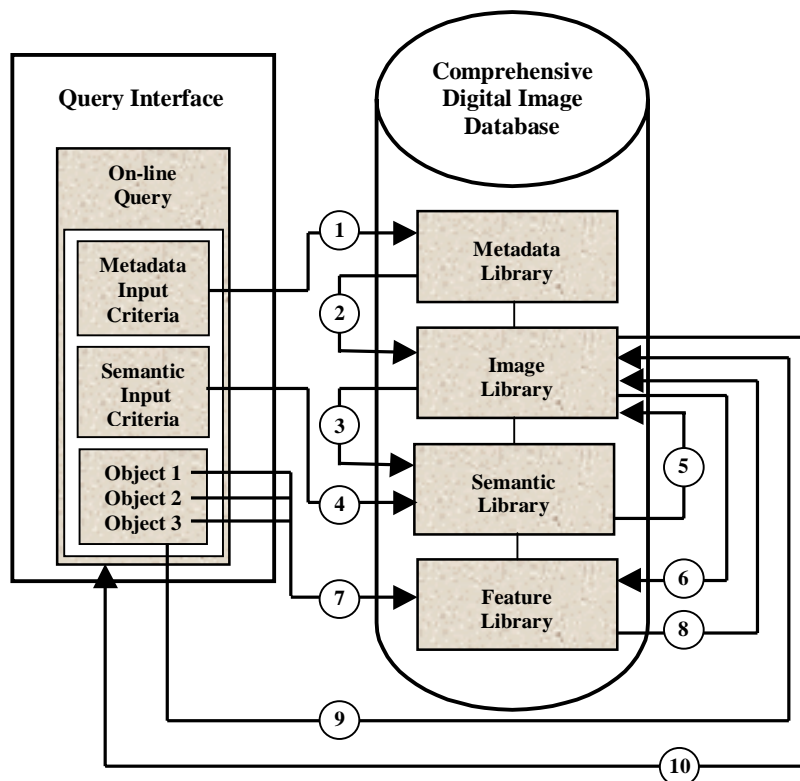


Figure 1 : The Query Scene Similarity Matching Workflow

When a feature gets matched to an image, its centroid coordinates within the image are recorded as well as the top left and bottom right coordinates of the query feature's minimum bounding rectangle, after scaling and rotation have taken place. Spatial relations on the image are determined through the use of the matched query features MBRs instead of the actual image-objects because the image is in raster (non-vectorized) format and therefore no a-priori information is known about any image-object (in particular, their boundaries). Also, it is straight forward to determine the MBR

containing the pixels composing the translated/rotated/scaled query object and using their MBRs allows for spatial relations to be built and queried in real-time.

Once the query scene has been built, the individual component objects that comprise the scene are separable and the 9 intersection matrix [Egenhofer and Franzosa, 1991] can be determined for each pair of objects in the scene. Determining the position relation matrix and the distance ratio matrix for the query scene follows this. Next, the search will commence by first retrieving all the images that match to the individual objects of the query scene. From this subset of images, only those images that have a match above a given threshold for all of the objects in the query scene will be further considered. From these remaining images, the topological/directional/distance relations between the matched objects will be computed and compared to that of the query scene.

2.1 Position Relation Matrix

To overcome the lack of exterior orientation information in the query and image scenes, we propose to use an intrinsic reference frame where the query/image-object orientation is in respect to "left-of" or "right-of" the features themselves. To do this, a position relation matrix is built for each query/image scene. Since a scene comprised of two (or less) objects is trivial to discern, i.e. a scene could be rotated to suit any orientation, our approach assumes three (or more) objects in the scene (Figure 2).

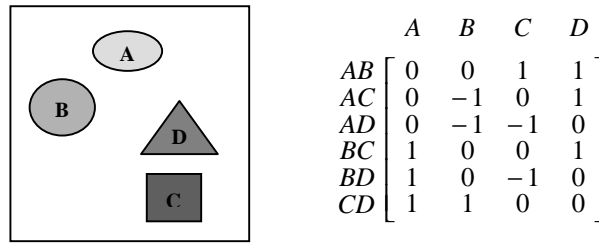


Figure 2 : Example Image Scene and Corresponding Position Relation Matrix

To build the position relation matrix for the scene depicted in Figure 2, an extended, imaginary line connecting the centroid of Feature A to the centroid of Feature B is drawn. Feature A is arbitrarily considered as the "top" feature and Feature B the "bottom" feature, and for every other feature (Features C through n) in the scene it is determined whether they lie left-of or right-of this line. Fixing the same features in both the query and image scenes to be either top or bottom renders any rotations in the scenes immaterial. The calculation of left-of or right-of is a simple matter considering we know (from the feature matching algorithm) the pixel coordinates of each feature's MBR in the image scene and each features actual boundary outline in the query scene. For example, if the MBR of Feature C in the image scene lies left-of this line, a -1 (negative 1) is placed in the position relation matrix. If Feature C's MBR lies right-of this line, a 1 is placed in the matrix and if Feature C's MBR lies somewhere along this extended line, a 0 is placed in the matrix.

After the remaining features of the scene are likewise added to the matrix, the imaginary line between Feature A and Feature B is deleted and redrawn between Feature A and Feature C with the process of determining left-of and right-of for the remaining features in the scene repeated. This entire procedure is repeated for every combination of features in the scene, i.e. there are $n(n-1)/2$ combinations of extended, imaginary lines that need to be tested against. For example, in the image scene of Figure 2, with 4 image-objects, there are 6 combinations of imaginary lines to be tested against, resulting in a 6x4 position relation matrix.

A position relation matrix is constructed for the query scene and each image scene returned from the initial query. A normalized correlation coefficient for each query/image scene is then calculated using Equation 1, describing their respective similarities. This coefficient is then scaled between 0 and 1 to give a total scene position matching percentage between the query and image-object configurations independent of any arbitrary scene rotations.

$$NC = \frac{\sum (I - \bar{I})(Q - \bar{Q})}{\sqrt{\sum (I - \bar{I})^2 \sum (Q - \bar{Q})^2}} \quad (1)$$

where I and Q are the position relation matrices for the image and query scenes respectively. Similar to the extrinsic reference frame approach, distance is also required in order to distinguish properly between these configurations of spatial entities.

2.2 Distance Ratio Matrix

For the purpose of this research, where no scale information on the query/image scene is provided a-priori, it is necessary to analyze the relative distances between image-objects. For example, a square matrix of rank $n(n-1)/2$ (where n is the number of objects in the scene) can be built for every query/image scene. Each combination of image-object connection is set to a distance of 1 unit and the distances to the other objects in the scene determined relative to this unit distance. Assuming an example scene of 3 objects, a 3x3 distance ratio matrix would be built (Figure 3).

The ratio approach to calculating the relative distances between objects does not require absolute scale information and is possible of course because the pixel coordinates of the image-object centroids are returned from the feature matching algorithm. The similarity between various query and image scenes, or more specifically their respective distance ratio matrices, is then determined through analysis of their normalized correlation coefficients (Equation 1) scaled between 0 and 1.

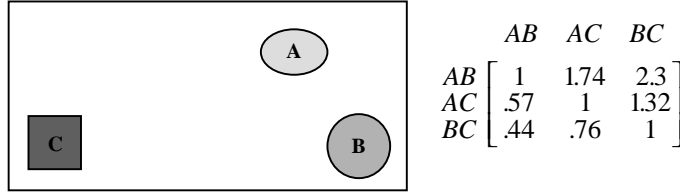


Figure 3 : Example Image Scene and Corresponding Distance Ratio Matrix

3 S_{MET} - A SCENE SIMILARITY METRIC

The next step will be to provide an evaluation of the percentage of similarities to allow for a ranking of the retrieved images. To do this, it is required to define a function S_{met} that assesses the similarity metric between a query configuration Q and an image I in the database. Such a function will combine different similarity functions for individual object shapes, and the topological, orientation and distance relations between them. More specifically S_{met} is defined as follows (Equation 2):

$$S_{met}(Q, I) = S_{sh}(Q, I) \cdot w_{sh} + S_{top}(Q, I) \cdot w_{top} + S_{or}(Q, I) \cdot w_{or} + S_{dist}(Q, I) \cdot w_{dist} \quad (2)$$

where:

- S_{sh} is a function measuring the degree/percentage of shape similarity between the objects in Q and the corresponding objects in I . For example, assuming that obj_1, \dots, obj_n indicate the objects in Q ,

$$S_{sh}(Q, I) = \frac{\sum_{i=1}^n match\%(obj_i)}{n}$$
, where $match\%(obj_i)$ is the matching percentage between object obj_i in Q and the corresponding object in I . We might further constrain this formula by imposing that for each $i=1, \dots, n$ $match\%(obj_i) > \epsilon$, with ϵ a given threshold value; for example, we only consider an object obj_i in Q as “found” in I if the corresponding object in I matches to it more than 50%;

- w_{sh} is a given weight value establishing the “importance” of shape similarity with respect to the other similarity criteria considered;
- S_{top} is a function measuring the degree/percentage of similarity between the set of topological relations characterizing the set of objects in Q and the topological relations among the corresponding objects in I ;
- w_{top} is a given weight value establishing the “importance” of similarity in topological relations with respect to the other similarity criteria considered;
- S_{or} is a function measuring the degree/percentage of similarity between the set of orientation relations characterizing the set of objects in Q and the orientation relations among the corresponding objects in I ;
- w_{or} is a given weight value establishing the “importance” of similarity in orientation relations with respect to the other similarity criteria considered;
- S_{dist} is a function measuring the degree/percentage of similarity between the set of distance relations characterizing the set of objects in Q and the distance relations among the corresponding objects in I ;
- w_{dist} is a given weight value establishing the “importance” of similarity in distance relations with respect to the other similarity criteria considered, and;

- $\sum_{j \in J} w_j = 1, J = \{sh, top, or, dist\}$

4 A WORKING EXAMPLE

In this section we show an example of how four different query scenes (Figure 4), sketched by the user, match to a given image I (Figure 5). A summary of the results of the similarity metric calculations for this example can be found in Table 1. The query scenes comprise differing configurations of the same three object shapes; i.e. an outline of an airplane, airplane hanger, and runway.

More specifically, let:

- $obj_1 = \text{airplane}$,
- $obj_2 = \text{airplane hanger}$,
- $obj_3 = \text{runway}$.

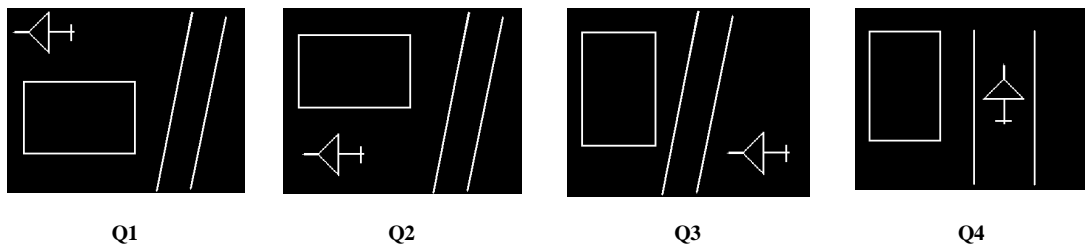


Figure 4 : Four Sample Query Scenes

For each of the four query scenes, Q_1, Q_2, Q_3 , and Q_4 , we calculate the value of S_{met} using Equation 2. First we calculate the value of S_{sh} (which is the same for all query scenes). For these three objects, the feature matching algorithm returned the following matching percentages:

- $match\%(obj_1) = 87$
- $match\%(obj_2) = 70$
- $match\%(obj_3) = 90$

Thus the value for shape similarity is the following: $S_{sh}(Q_i, I) = \frac{87+70+90}{3} = 82.3\%, i = 1, \dots, 4$

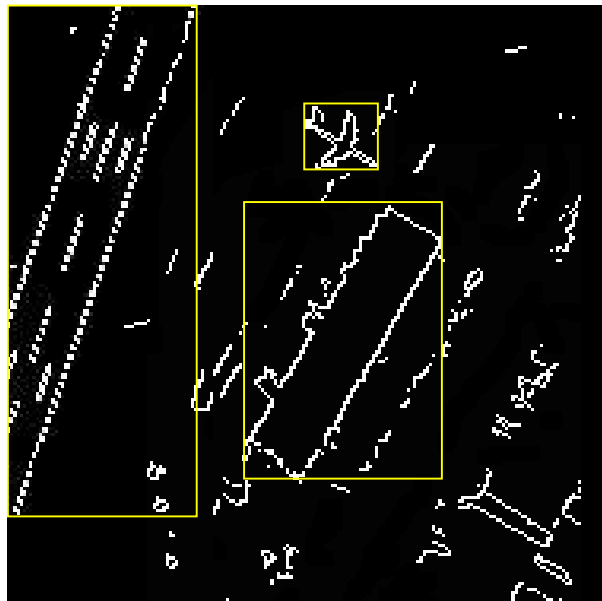


Figure 5 : Sample Edge-Image with Superimposed MBRs

4.1 Spatial Relations for the Image Scene

For the image I , spatial relations are derived using minimum bounding rectangles (MBRs), while for user-defined query objects, more accurate methods are used. Spatial relations in I therefore are characterized as follows.

Topological relations between objects in I :

- all objects are *disjoint*;
- Pixel coordinates (row,col) $obj_1 = (44,105)$
- Pixel coordinates (row,col) $obj_2 = (106,105)$
- Pixel coordinates (row,col) $obj_3 = (71,22)$

Direction relations between objects in I :

- Position Relation Matrix
- | | obj_1 | obj_2 | obj_3 |
|----|---------|---------|---------|
| 12 | 0 | 0 | -1 |
| 13 | 0 | 1 | 0 |
| 23 | -1 | 0 | 0 |

Metric relations between objects in I :

- Distance Ratio Matrix
- | | 1 2 | 1 3 | 2 3 |
|----|-----|------|------|
| 12 | 1 | 1.41 | 1.45 |
| 13 | .71 | 1 | 1.03 |
| 23 | .69 | .97 | 1 |

In the following sections, we calculate the values of S_{top} , S_{or} , and S_{dist} for all four query scenes Q_1 , Q_2 , Q_3 , and Q_4 .

4.2 Spatial Relations for Query Scene 1

Topological relations between objects in Q_1 :

All objects in Q_1 are disjoint, i.e., topological relations in Q_1 are 100% similar to topological relations among the corresponding objects in I , therefore:

- $S_{top}(Q_1, I) = 100\%$
- Pixel coordinates (row,col) $obj_1 = (40,59)$
- Pixel coordinates (row,col) $obj_2 = (180,117)$
- Pixel coordinates (row,col) $obj_3 = (152,305)$

Direction relations between objects in Q_1 :

- Position Relation Matrix
- | | obj_1 | obj_2 | obj_3 |
|----|---------|---------|---------|
| 12 | 0 | 0 | 1 |
| 13 | 0 | -1 | 0 |
| 23 | 1 | 0 | 0 |

Therefore, the value of S_{or} for Q_1 is the following:

- $NC_{1or} = \frac{\sum(I - \bar{I})(Q - \bar{Q})}{\sqrt{\sum(I - \bar{I})^2 \sum(Q - \bar{Q})^2}} = \frac{-2.6447}{\sqrt{2.4489 * 2.8889}} = -.99$
- $\therefore S_{or}(Q_1, I) = \frac{-.99 + 1}{2} = 5\%$

Metric relations between objects in Q_1 are calculated similarly to above:

- Distance Ratio Matrix
- | | 1 2 | 1 3 | 2 3 |
|----|-----|------|------|
| 12 | 1 | 1.78 | 1.25 |
| 13 | .56 | 1 | .70 |
| 23 | .80 | 1.42 | 1 |

Therefore, the value of S_{dist} for Q_1 is the following:

- $NC_{dist} = \frac{\sum (I - \bar{I})(Q - \bar{Q})}{\sqrt{\sum (I - \bar{I})^2 \sum (Q - \bar{Q})^2}} = \frac{.6288}{\sqrt{.8871 * 1.1421}} = .6247$
- $\therefore S_{dist}(Q_1, I) = \frac{.6247 + 1}{2} = 81.2\%$

Finally, assuming equal weighting, i.e. that the value of all w_j (with $j \in \{sh, top, or, dist\}$) is .25, the value of S_{met} for Q_1 becomes:

- $S_{met}(Q_1, I) = 82.3 * .25 + 100 * .25 + .5 * .25 + 81.2 * .25 = \mathbf{66\%}$

4.3 Spatial Relations for Query Scene 2

Topological relations between objects in Q_2 :

All objects in Q_2 are disjoint, i.e., all topological relations in Q_2 are 100% similar to the topological relations among the corresponding objects in I , therefore:

- $S_{top}(Q_2, I) = 100\%$
- Pixel coordinates (row,col) $obj_1 = (239,82)$
- Pixel coordinates (row,col) $obj_2 = (103,113)$
- Pixel coordinates (row,col) $obj_3 = (153,305)$

Direction relations between objects in Q_2 :

- Position Relation Matrix

	obj_1	obj_2	obj_3
12	0	0	-1
13	0	1	0
23	-1	0	0

Therefore, the value of S_{or} for Q_2 is the following:

- $S_{or}(Q_2, I) = \frac{1+1}{2} = 100\%$

Since Q_2 is a 180° rotation of I , all position relations are 100% similar to the position relations among the corresponding objects in I .

Metric relations between objects in Q_2 :

- Distance Ratio Matrix

	1 2	1 3	2 3
12	1	1.71	1.42
13	.58	1	.83
23	.70	1.20	1

Thus,

- $S_{dist}(Q_2, I) = \frac{.7337 + 1}{2} = 86.7\%$

Finally, assuming equal weighting, i.e. that the value of all w_j (with $j \in \{sh, top, or, dist\}$) is .25, the value of S_{met} for Q_2 becomes:

- $S_{met}(Q_2, I) = 82.3 * .25 + 100 * .25 + 100 * .25 + 86.7 * .25 = \mathbf{92.3\%}$

The spatial relations for the remaining query scenes (i.e. 3 and 4) were calculated similarly to the first two scenes and due to space limitations are not detailed here. The results of all the similarity metric calculations can be seen in Table 1.

From these results it can be seen that Query Scene 2 is the best-matched configuration for the given image. This agrees with what a human observer would choose as the best matched configuration since Query Scene 2 is plainly a 180° rotation of the image, sketched at a significantly reduced scale. This result demonstrates the ability of the S_{met} approach to differentiate between arbitrary rotations and scaling of varying query/image scenes in addition to the capacity to distinguish between configurations and shapes of individual image-objects.

	Query Scene 1	Query Scene 2	Query Scene 3	Query Scene 4
S_{sh}	82.3	82.3	82.3	82.3
S_{top}	100	100	100	66.7
S_{or}	.5	100	79	81
S_{dist}	81.2	86.7	33.7	35
S_{met}	66	92.3	73.8	66.3

Table 1 : S_{met} Results for the Four Query Scenes

5 CONCLUSIONS

In this paper we propose a novel solution to querying image databases by matching configurations of raster features to imagery completely in the raster/spatial domain using the shape of single features together with their spatial relations as the matching primitives. We defined a methodology to evaluate similarity (both of objects and relations) between a user-defined query scene and the retrieved images that is insensitive to both scene/image scale and rotation. To do this, we introduced the function S_{met} that assesses the *similarity metric* between a query configuration Q and an image I in the database. Such a function combined the different similarity functions describing individual object shapes and the topological, orientation and distance relations between them.

A working example presented a practical illustration of how the S_{met} function calculated the similarity metric between four query scenes and a given image. The results of this example showed how each of the spatial relations between objects comprising a query scene affects the assessment of scene similarity. Depending on the method used for determining topology, (e.g. using rotated MBRs), or direction (e.g. intrinsic vs. extrinsic reference frames), different results for S_{met} can be obtained. Also, by modifying the weights assigned to each of the relations independently, another set of results could be returned. Furthermore, by loosening up object/relation constraints and by analyzing matching percentages, we can detect temporal changes in some areas, such as: the elimination of some objects, changes in object shape, and change in location.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation through CAREER grant number IIS-9702233 and by National Imagery and Mapping Agency through NURI grant number NMA202-98-1-1113. Michela Bertolotto's work was partially supported by CNR grant number 106701-00/97/10003.

REFERENCES

- Agouris, P., Carswell, J. and Stefanidis, A., 1999a. An Environment for Content-Based Image Retrieval from Large Spatial Databases. *ISPRS Journal of Photogrammetry & Remote Sensing*, Elsevier, 54(4): 263-272.
- Agouris, P., Carswell, J. and Stefanidis, A., 1999b. Sketch-Based Image Queries in Topographic Databases. *Journal of Visual Communication and Image Representation*, 10(2): 113-129.
- Carson, C., Belongie, S., Greenspan, H. and Malik, J., 1997. Region-Based Image Querying, *IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, pp. 42-49.
- Carswell, J., 2000. Using Raster Sketches for Digital Image Retrieval. Ph.D. Dissertation, University of Maine, Orono, Maine, USA.
- Egenhofer, M. and Franzosa, R., 1991. Point-Set Topological Spatial Relations. *International Journal of Geographical Information Systems*, 5(2): 161-174.
- Flickner, M. et al., 1995. Query by Image and Video Content: The QBIC System. *IEEE Computer*, 28(9): pp. 23-32.
- Forsyth, D.A. et al., 1996. Finding pictures of objects in larges collections of images, *ECCV 96 Workshop on Object Representation*.
- Frankel, C., Swain, M. and Athitsos, W., 1996. WebSeer: An image search engine for the world wide web. TR-96-14, Department of Computer Science, University of Chicago.
- Goyal, R. and Egenhofer, M., 2000. Cardinal Directions Between Extended Spatial Objects. *IEEE Transactions on Knowledge and Data Engineering* (to appear).
- Pentland, A., Picard, R.W. and Scarloff, S., 1996. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*.
- Scarloff, S., Taycher, L. and La Cascia, M., 1997. ImageRover: A Content-Based Image Browser for the World Wide Web, *IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, pp. 2-9.