2002-01-01

Scale and Orientation-invariant Scene Similarity Metrics for Image Queries

James Carswell
Dublin Institute of Technology, jcarswell@dit.ie

Recommended Citation

Carswell, J. (2002) Scale and orientation-invariant scene similarity metrics for image queries. *International Journal of Geographical Information Science (IJGIS)*, 2002. doi:10.1080/13658810210148552.

SCALE- AND ORIENTATION-INVARIANT SCENE SIMILARITY METRICS FOR IMAGE QUERIES

Peggy Agouris [1], Michela Bertolotto, James D. Carswell, and Anthony Stefanidis

Dept. of Spatial Information Science and Engineering and
National Center for Geographic Information and Analysis (NCGIA)
{peggy,michelab,carswell,tony}@spatial.maine.edu
University of Maine
USA

KEY WORDS: Image analysis, Scene matching, Image queries, Raster Spatial databases

[1] **Corresponding Author:**

Dr. Peggy Agouris
5711 Boardman Hall #348
Orono, ME 04469-5711
USA
Tel: (207) 5812180
Fax: (207) 5812206
Email: peggy@spatial.maine.edu

SCALE- AND ORIENTATION-INVARIANT SCENE SIMILARITY METRICS FOR IMAGE QUERIES

ABSTRACT

In this paper we extend our previous work on shape-based queries to support queries on configurations of image objects. Here we consider spatial reasoning, especially directional and metric object relationships. Existing models for spatial reasoning tend to rely on pre-identified cardinal directions and minimal scale variations, assumption that cannot be considered as given in our image applications, where orientations and scale may vary substantially, and are often unknown. Accordingly, we have developed the method of *varying baselines* to identify similarities in direction and distance relations. Our method allows us to evaluate directional similarities without a priori knowledge of cardinal directions, and to compare distance relations even when query scene and database content differ in scale by unknown amounts. We use our method to evaluate similarity between a user-defined query scene and object configurations. Here we present this new method, and discuss its role within a broader image retrieval framework.

1. INTRODUCTION

Advancements in sensor technology resulted in substantial increases in the availability of reliable digital imagery for geospatial applications. The upgraded role of imagery within the geographic information production, management, and distribution cycle emphasizes the need for efficient image retrieval methods. Within the context of geospatial applications image retrieval solutions need to consider the actual content of images, namely objects depicted in them (synonymous in this paper with object outlines or edges), and the spatial relations of these objects (e.g. arrangements of buildings in a scene).

In contrast, the majority of image query efforts focused initially on analyzing and comparing images based on their more general properties. 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].

Furthermore, we have to consider the fact that despite substantial advancements [see e.g. Gruen et al. 1995b, 1997; Lukes, 1998 for overviews of on-going activities], fully automated object extraction from digital imagery remains an unsolved issue. Instead, semi-structured object outlines are more easily available. Accordingly, the ability to handle information (object outlines) in the raster domain, without requiring precise vectorization of image objects, remains important for image queries in geospatial applications. Considering how a raster-based approach can function equally well with semi-structured raster edges as well as with fully structured vectorized objects (as vector-to-raster conversion is trivial), we can see the full advantage of such an approach.

Our purpose in this paper is to extend our recently completed work on image query-by-sketch to handle multiple objects [Agouris et al., 1999b]. Our initial approach was designed to perform image queries using query elements and structured databases, all in raster format. Our work focused on single object queries, where the objective is to identify images containing a single query object (e.g. a cross-shaped building). Here we want to extend this framework to perform queries on spatial configurations of multiple objects. The work of [Egenhofer and Franzosa, 1991] on topological and directional relations, and its recent extension on scene similarity metrics [Goyal & Egenhofer, 2000; Blaser, 2000] present an ideal framework for such a task. However, these methods are not optimal for image-based applications, as they tend to rely on the assumption of pre-defined cardinal directions, and on minimal scale differences between query and database. These assumptions commonly fail when dealing with image collections, where orientation and scale may vary widely, and may even be unknown.

The novelty of this paper is related to the methods introduced here to assess the similarity between two scenes based on the relative positions of objects, and on the distances between these objects. The metrics introduced here bypass the need for the a priori knowledge of cardinal directions, and furthermore allow us to assess similarities even at the presence of large scale variations between two scenes. Both of these properties were lacking from currently existing scene similarity solutions and are important for image-related applications.

The paper is structured as follows: In Section 2 we present an overview of our image retrieval environment, and the role of a scene similarity metric within this environment. In Section 3 we introduce our innovations in scene similarity metrics, more specifically our new varying baseline methods for the comparison of relative positions of objects and of object distances. In Section 4 we provide a working example to demonstrate the application of our newly introduced methods. Conclusions and future plans are in Section 5.

2. IMAGE QUERY FRAMEWORK

Images depict objects and capture their relationships. This information exists implicitly within images, and has to become explicit in order to support complex database operations. In doing so, we essentially transform a *raw* image into its *computational* counterpart, a main goal of image understanding research. Among other issues, this transformation process involves two types of operations:

- identifying objects within these images, and
- modeling the spatial relationships of these objects.

Accordingly, our approach to image queries reflects this duality. Our objective is to proceed by first identifying images that contain objects resembling the objects provided as query input, and subsequently, by analyzing the results, to identify these images in which the configurations of the objects best resemble the input configuration.

In our previous work we addressed the first part of this query, namely establishing a framework and relevant processes that would enable the comparison of a raster template to a collection of images to retrieve images containing objects that resemble the query template [Agouris et al., 1999a, 1999b; Carswell, 2000]. Here we present an extension of this work by introducing metrics to compare the configurations of multiple objects.

2.1 Query Environment

In our environment we assume that a typical query of an image database comprises metadata, semantic data, and a sketched configuration of image objects (Fig. 1). As shown in the figure, our comprehensive image database comprises actual images, and three related libraries, namely a metadata, a semantic, and a feature library. The metadata and semantic libraries are common indexing mechanisms, relating properties (e.g. metadata values) to image files. The feature library is an organized arrangement of outlines based on their mutual geometrical similarity as it is expressed via matching similarity coefficients. It is also an indexing structure, linking outline templates to image locations in which similar features appear. The reader is referred to [Agouris et al, 1999a] for a detailed description of the theoretical issues behind the feature library, and the interrelationships among these libraries.

Considering as an example that a user wishes to retrieve all images from the State of Maine over the last 2 years with jet airplanes, airplane hangers and runways that match a particular configuration (e.g. airplanes between the hangers and the runways), the query would:

- Process the metadata library and retrieve all images that match the specified metadata criterion (in this case the time of acquisition should be during the last 2 years, and the image footnote should be in the State of Maine);
- Follow the links from this subset of images to the semantic library where it would identify which of these images contained airports;
- Use this further reduced subset of imagery to identify the subset of feature library that will be searched for objects similar to the ones in the query sketch;
- Perform a matching-based similarity search to identify the library features that best match the query content;
- Retrieve all imagery linked to the best matched features;
- Determine which of these images contain all of the objects in the query configuration;
- Analyze the spatial relations of the query scene on this final subset of imagery and return a prioritized list of imagery as the query result.

In this paper we introduce metrics to support the performance of the last bullet of the above process, namely analyzing the spatial relations of objects considering the particularities of our image-driven applications (operation 9 in Fig. 1). Our previous work focused on the use of a single query object to retrieve images containing at least one object similar to the query (operations 1-8 and 10 of the ones depicted in Fig. 1). In the next section we provide a brief description of our object matching approach, as it is a key issue in our framework.

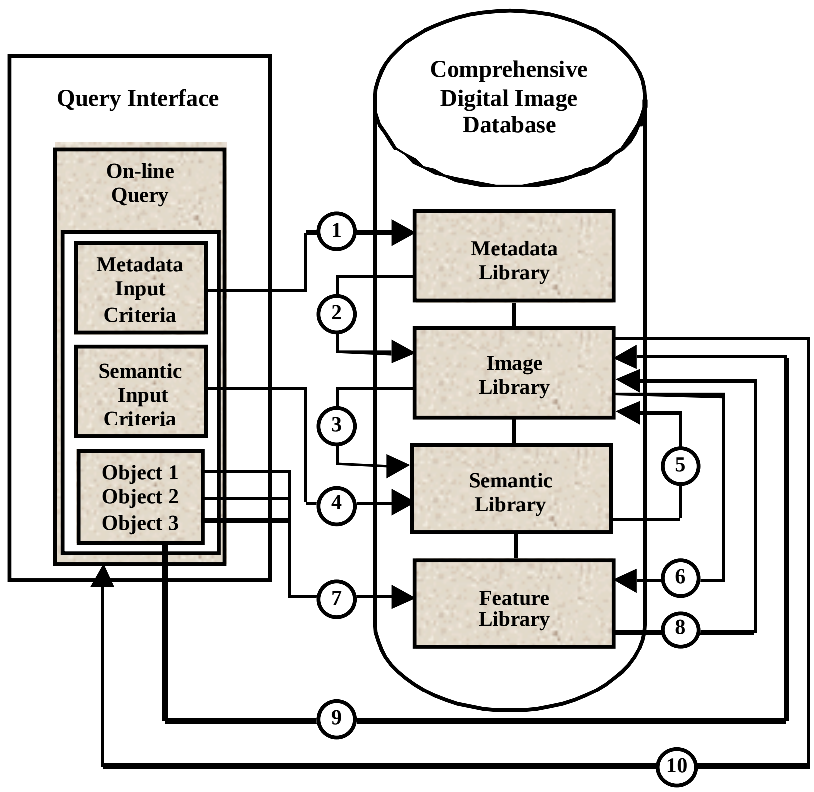

Figure 1: The Query Scene Similarity Matching Workflow

2.2 Shape Matching for Single Object Queries

Image matching is used to identify similar objects in two or more images. A very powerful method to perform this task is least squares matching (LSM). In traditional LSM, an image window is compared to another template window by analyzing the gray value differences between them and examining whether these differences could be minimized through a geometric transformation (e.g. affine transformation). The remaining gray value differences are indicators of the accuracy of the matching process. Limitations of this approach are that it can be a highly computation intensive operation, even for small patches of pixels, and that good initial approximations for positioning the template patch within the image are required for determining a correct match. Advantages of the approach include its very high accuracy potential (on the order of 0.1 pixel or better), and the ability to accommodate geometric variations between the two matched windows [Gruen & Agouris, 1994; Gruen et al., 1995].

In our framework we use a variation of LSM to compare the sketched object outline to the feature library, in order to identify the feature that best resembles the query sketch. In our variation of LSM we use edges instead of gray scale pixels as the metric for query scene to image similarity. By doing so, we avoid having to test full patches of pixels against each other. Instead, we reduce the patch into its information content by considering only those pixels that contain object outline information, i.e. only those that constitute image object edges. Edge information can be obtained from the images in our image library by performing common edge detection operations, or even by having operators manually digitizing object outlines. Since object edges are used instead of gray scale pixels, gray level observations are replaced by feature existence/absence observations. This resembles the comparison of gray values in traditional least-squares matching and the use of image gradients to identify shifts, rotations, and scalings.

The matching process then reduces to checking only if the input query template edge pixels overlay edge pixels from the binary edge-image; the answer then can only be either a “yes” or “no”. If yes, the corresponding template edge pixel “votes” to stay where it is. If no, then this template edge pixel votes to move. By summarizing and analyzing the voting patterns of all the pixels that make up an object's edge, a global decision as to where and how far to shift the query template is made. Following iterations we determine the best match between the template and a window. The least squares framework of our solution and the modeling of the relations between the two matched objects ensures invariance for scale, orientation, and shifts.

The details of this novel matching approach are beyond the scope of this paper. The reader may refer to [Agouris et al., 1999b] for a detailed analysis of our matching tool. What is important here is that this matching process provides us with accuracy measures for every matching result. These measures are expressed as a percentage, with a maximum of 100% indicating a perfect match and a minimum of 0% indicating no match. In-between values express variations in our matching accuracy, with higher values corresponding to better matches than lower ones. A perfect match implies that there exists a set of affine transformation parameters that could be used to transform the query template so that it matches perfectly the library feature and, correspondingly, objects in images in our image database.

Using this matching tool within our image query environment we can identify for a query sketch the most similar feature library member, and the corresponding image locations where this feature appears. Thus, responses to our query are images, locations within them where objects similar to our sketch appear, and percentages expressing how similar these objects are to our sketch. Object locations within the image are communicated as the minimum bounding rectangles (MBR) containing the object of interest.

2.3 Extension for Multi-Object Queries

Our approach to multi-object queries is a two-stage approach:

- first identify the images (and locations within them) where each of the query objects appears independently of the other objects, and

- subsequently examine these results to identify the configuration that best resembles the spatial properties of the query sketch.

Assuming that the query sketch contained *n* objects, the first issue is equivalent to performing *n* queries as described in 2.1 and 2.2. above, one for each object in the sketch. The second part requires the use of a scene similarity metric that takes into account spatial and topological relationships between objects. The importance of topology, orientation, and distance in assessing spatial similarity of scenes is well-documented (see [Egenhofer & Franzosa, 1991; Goyal & Egenhofer, 2000, Shariff, 1996]). A combinatorial expression for all these properties has been introduced by [Blaser, 2000] to function within a general query-by-sketch environment. Based on this work we define a function S_{met} that assesses the similarity metric between a query configuration Q and an image I in the database. The function combines different similarity metrics for individual object shapes and relations between them like topolgy, orientation, and distance. More specifically S_{met} is defined in our case as:

$$S_{met}(Q,I) = S_{sh}(Q,I) \cdot w_{sh} + S_{top}(Q,I) \cdot w_{top} + S_{or}(Q,I) \cdot w_{or} + S_{dist}(Q,I) \cdot w_{dist} \qquad \text{(Eq 1)}$$

The elements of this formula are as follows:

- 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 $obj_1, \ldots, obj_n$ indicate the n objects in Q,, then $S_{sh}(Q,I) = [\Sigma \ match\%(obj_i)]/n$, where $math\%(obj_i)$ is the matching percentage between object obj_i in Q and the corresponding object in I. We can further constrain this formula if we so wish by imposing acceptability constraints. For example we can require that for each $i=1,\ldots,n$ $math\%(obj_i)>t$, with t a given threshold value. This would make us consider an object obj_i in Q as “found” in I if and only if the corresponding object in I matches to it more than a preset threshold (e.g. 50%).

- 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.
- 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.
- 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.
- $w_{sh}, w_{top}, w_{or}, w_{dist}$ are weight coefficients establishing the relative importance of their corresponding similarity metrics for the overall scene similarity assessment. By minimizing for example the first three coefficients, we search for configurations resembling the query sketch only in terms of distances between the objects, regardless of their shape, topology, and orientation.

All above similarity metrics are in the range [0,1] with higher values corresponding to higher similarity. By enforcing $\sum_{j \in J} w_j = 1$, $J = \{sh, top, or, dist\}$ we ensure that the overall scene similarity metric S_{me} will also have a value in the range [0,1]. Equation 1 is based on the approach of [Blaser, 2000] for general scene similarity metrics. In our case however, and considering the particularities of image databases, we introduce new metrics for orientation and distance similarities, and make use of image matching techniques to provide shape similarity measures.

Contrary to orientation and distance, topological relations are rather unaffected by variations in image scale and orientation. Of course, extreme scale variations may cause two disjoint objects for example to appear merged into a single blob. Extreme oblique views may have similar effects, distorting the apparent topological relations of objects. However, such exceptional cases are considered beyond our interests. Therefore, we assume the use of the well-established hierarchical topological models introduced by [Egenhofer & Franzosa, 1991; Egenhofer & Al-Tahe, 1992] to describe topological relations in our scene similarity metric. According to this model the binary topological relations between simply connected regions range progressively from disjoint to meets, overlap, covers/covered, and contains/contained. The topological similarity index between two scenes describes how far their contents are in this arrangement. The reader is referred to the above mentioned references for a detailed description of these metrics.

3. THE VARYING BASELINES APPROACH FOR SCENE SIMILARITY

In this section we introduce two similarity metrics to compare the orientation and distance content of scenes for use in our query environment. These new metrics, which are at the core of this paper, are based on the use of new matrices, namely the *position relation* matrix, and the *distance ratio* matrix. We refer to both approaches under the term *"varying baselines approach"*, as they inherently employ the establishment of baselines connecting objects, and the comparison of properties derived from these baselines.

3.1 Position Relation Matrix for Orientation Similarity Metrics

The direction (orientation) between features in a spatial database is required to further refine their spatial relationships. When querying, it is not enough to return all the scenes where feature A is disjoint from feature B (although this would be a good first approximation). It is also important to consider the directional relationship between them.

To make sense of direction, a reference frame must first be established. In general there are three types of reference frames:

- intrinsic, where the reference frame is in respect to the orientation of the feature itself, e.g. front or back, left or right of a building;
- deictic, where the reference frame is relative to each individual looking at the scene, e.g. what is "in front" for me might be "to the left of" someone else and;
- extrinsic, where the reference frame is established independently of the orientation of the features or the observers, e.g. north, south, east, west.

For configurations of spatial objects, in a GIS or digital image, that represent real positions and orientations of the environment, it has been customary to use extrinsic reference systems.

Traditional methods to determine direction between spatial entities have been simplified to determining the direction between their point approximations. For example, the position of the centroid of a feature is determined and compared to the position of the centroid of a neighboring feature. This gives the precise bearing between the two points, e.g. feature A is 45°15'33"W of feature B. But because areal features may have non-symmetric shapes, their centroids may not even lie within their respective boundaries. To overcome this case, the minimum bounding rectangle (MBR) was introduced to approximate the feature's shape, and the centers of these MBRs are compared to determine direction. Again though, only the direction between two points is being compared.

The latest research into determining the direction between spatial entities uses a direction relation matrix [Goyal and Egenhofer, 2000]. In this approach, an extended MBR is placed around feature A and the percentage of the area of feature B that falls within the various regions is recorded in the direction relation matrix *dir(A,B)*. Pairs of direction relation matrices are then compared and their differences analyzed to determine their respective similarity.

Although the direction relation matrix can distinguish between most configurations of query/image objects, it has the drawback that it depends upon an extrinsic reference frame, i.e. both the query and the image must be orientated to the cardinal directions of north, south, east and west. Unfortunately, when dealing with raw raster imagery, as in our case, there are no exterior orientation parameters known a-priori. Therefore, another approach must be implemented for testing direction similarity between the query and image scenes.

To overcome the lack of exterior orientation information in the query and image scenes we propose to use a new matrix, the *position relation matrix*. This matrix reflects an intrinsic reference frame, and models query/image object orientations in respect to *"left-of"* or *"right-of"* the features themselves.

Let us assume to have a scene with *n* objects $obj_1, \ldots, obj_n$. For each object obj_i (such that $1 \le i < n$), an extended, imaginary baseline is identified, connecting the centroid of obj_i to the centroid of each object obj_j (such that $i < j \le n$). Object obj_i is arbitrarily considered as the *"top"* object and obj_j the *"bottom"* object, and for every other object obj_k in the scene (such that $1 \le k \le n, k \ne i, k \ne j$), it is determined whether obj_k lies left-of or right-of this line. Fixing the same objects 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 and query scenes. These relations are tabulated in the position relation matrix.

The position relation matrix *P* for a query scene of *n* objects has *n(n-1)/2* rows (one for each line connecting two objects) and *n* columns (one for each object). Each row corresponds to a pair (obj_i, obj_j) of objects with $1 \le i < n, i < j \le n$, and each column corresponds to an object in the scene. An element p_{hk} in *P* (with *h* the index of the row of *P* corresponding to a pair (obj_i, obj_j) such that $1 \le i < n, i < j \le n$, and $1 \le k \le n, k \ne i, k \ne j$) is set to -1 (negative 1), if the MBR of object obj_k in the image scene lies left-of the line connecting the MBR centroids of objects obj_i and obj_j. Conversely, p_{hk} is set to 1 if the MBR of object obj_k lies right-of this line. Furthermore, p_{hk} is set equal to 0 if the MBR of object obj_k intersects this extended line. Obviously, for each element p_{hk} in the matrix such that $k = i$ or $k = j$, p_{hk} is equal to 0. As an example, Figure 2 shows a scene containing 4 image objects. The 6x4 position relation matrix corresponding to Fig. 2 is:

$$
\begin{array}{c}
\\ AB \\ AC \\ AD \\ BC \\ BD \\ CD
\end{array}
\begin{array}{c}
\begin{array}{cccc} A & B & C & D \end{array} \\
\begin{bmatrix}
0 & 0 & 1 & 1 \\
0 & -1 & 0 & 1 \\
0 & -1 & -1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & -1 & 0 \\
1 & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

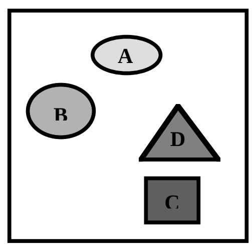

Figure 2: Example Image Scene

By design, the information contained in this matrix is independent of orientation. The role of cardinality is indeed nullified by the selection of different baselines (corresponding to each row) and the analysis of the scene content accordingly. It should also be noted that, by partitioning the space in two equal regions ("left-of" and "right-of") every time we populate the matrix, we also avoid the unfavorable bias that may be introduced in models where the NE, NW, SE and SW areas are larger in size than the N, E, S, and W regions. In this sense our approach preserves the unbiased nature of the traditional 4-region models like the projection-based models.

One can easily see that the matrix content also conveys general information about the distribution of objects in a scene. More than two 0 elements in a row indicate the alignment of three or more objects in a scene. In an extreme example, where all 4 objects would be aligned, the above matrix would become null. In the case where the objects are scattered in our scene, with no three of them aligned, each matrix row would contain exactly two 0 elements.

In our query environment, the shape similarity process produces candidate image objects that resemble the query ones. Combinations of these candidate image objects create candidate image scenes that have to be compared to the query scene. Accordingly, a position relation matrix is constructed for the query scene and for each candidate image scene returned from the initial query. Our objective is to identify the query/image scene combination that is most similar among the available options. This corresponds to identifying the most similar combination of position relation matrices among the options. The similarity of two such matrices is described by their normalized correlation coefficient as:

$$C_{or} = \frac{\sum (I - \bar{I})(Q - \overline{Q})}{\sqrt{\sum (I - \bar{I})^2 (Q - \overline{Q})^2}} \qquad \text{(Eq. 2)}$$

where I and Q are the position relation matrices for the image and query scenes respectively, with $\bar{I}$ and $\overline{Q}$ being the averages of their respective elements. 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.

3.2 Distance Ratio Matrix

Similar to the extrinsic reference frame approach, distance information is also an important part of a scene similarity comparison. 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. Thus, we consider a scale-independent approach, making use of a new matrix, the *distance ratio matrix*. Let's consider a scene with n objects. The distance ratio matrix for this scene is a square matrix D of rank $n(n-1)/2$. Its rows and columns correspond to baselines formed between object pairs (e.g. AB, AC, AD, BC etc.). The matrix elements are describing distance ratios for these baselines. Specifically, for every row of D, we pick the corresponding entry to serve as the unit distance, and populate the row by the ratio of all other distances over this unit distance. More formally, each entry d_{hk} in D corresponds to the ratio of the k^{th} distance over the h^{th} distance. There are two properties of this matrix that become immediately apparent. First, its diagonal elements are equal to 1, as they correspond to the ratio of a distance to itself. Second, d_{hk} is equal to $1/d_{kh}$. To better illustrate the design of the distance ratio matrix, let's consider the scene of Figure 3. In Figure 3 we have an example scene of 3 objects. The 3x3 position relation matrix corresponding to Fig. 3 is:

$$\begin{array}{c} \\ AB \\ AC \\ BC \end{array} \begin{array}{c} \begin{array}{ccc} AB & AC & BC \end{array} \\ \left[\begin{array}{ccc} 1 & 1.74 & 2.30 \\ 0.57 & 1 & 1.32 \\ 0.44 & 0.76 & 1 \end{array}\right] \end{array}$$

For example, element (1,2) of this matrix corresponds to the ratio of AC over AB, and element (3,1) corresponds to the ratio of BC over AB.

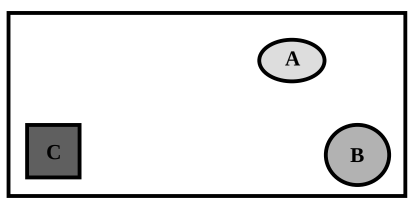

Figure 3: Example Image Scene for Distance Ratio Matrix

Similar to the use of position relation matrices, distance ratio matrices are formed for a query and candidate scenes and they are compared via their normalized correlation coefficients:

$$C_{dist} = \frac{\sum (I - \bar{I})(Q - \overline{Q})}{\sqrt{\sum (I - \bar{I})^2 (Q - \overline{Q})^2}} \qquad \text{(Eq. 3)}$$

where I and Q are the distance ratio matrices for the image and query scenes respectively, with $\bar{I}$ and $\overline{Q}$ being the averages of their respective elements. This coefficient is then scaled between 0 and 1 to give a total scene distance matching percentage between the query and image object configurations independent of any arbitrary scene rotations.

The ratio approach presented here to calculating the relative distances between objects does not require absolute scale information and is calculated using the pixel coordinates of the imageobject centroids returned from the feature matching algorithm. Using this information, a distance ratio matrix D is easily built for every query/image scene and the above analysis allows us to measure their similarity. This similarity measure is becoming invalid only in cases where severe variations in image orientation (e.g. highly oblique photography compared to a vertical query scene) may distort substantially the apparent distances between objects. However, such extreme cases are beyond the scope of our work, and are rather rare in common GIS processes.

4. A WORKING EXAMPLE

In this section we will show an example of how the scene similarity metrics introduced in this paper can be used to compare four different query scenes (Figure 3) to a given image I (Figure 4). 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 = airplane,
- obj_2 = airplane hanger,
- obj_3 = runway.

Q1 **Q2** **Q3** **Q4**

Figure 3: Four Sample Query Scenes

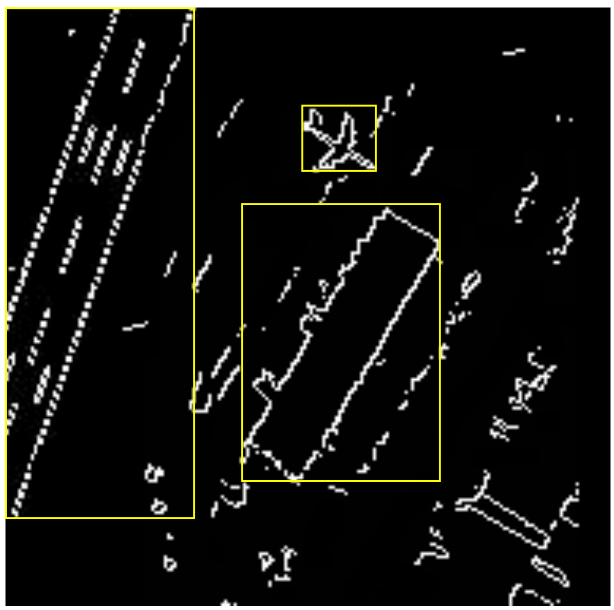

Figure 4: Sample Edge Image with Superimposed MBRs

For each of the four query scenes, $Q_{1,}$ $Q_{2,}$ $Q_{3,}$ and $Q_{4,}$ we will calculate the value of S_{met} using the scene similarity metric provided by Equation 1. As the viewer can easily see, the four query scenes are different configurations of the above mentioned three objects. Accordingly, the shape similarity measures S_{sh} among individual query objects and corresponding image features will be similar for all four query scenes. Our feature matching algorithm returned the following matching percentages:

- $match\%(obj_1)$ = 87% (the matching percentage between the airplane sketch and the edge outline in the upper middle box in the image of Fig. 4)
- $match\%(obj_2)$ = 70%
- $match\%(obj_3)$ = 90%

Thus the overall value for shape similarity for all query scenes is: $S_{sh}(Q_i, I) = \frac{87+70+90}{3} = 82.3\%$, (where $i=1,...4$).

In Section 4.1 we will establish the position relation matrix and the distance ratio matrix for the image scene. In Sections 4.2 - 4.5 we will establish these matrices for each of the query scenes, and we will estimate the overall scene similarity between each query and the given image. In section 4.6 we will compare these results to identify the best match.

4.1 Spatial Relations for the Image Scene

For the image *I*, spatial relations are derived using minimum bounding rectangles (MBRs). 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 image *I*:

- Position Relation Matrix:

$$\begin{array}{c} \\ 12 \\ 13 \\ 23 \end{array}\begin{array}{c} \begin{array}{ccc} obj_1 & obj_2 & obj_3 \end{array} \\ \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \end{array}$$

Distance relations between objects in image *I*:

- Distance Ratio Matrix:

$$\begin{array}{c} \\ 12 \\ 13 \\ 23 \end{array}\begin{array}{c} \begin{array}{ccc} 12 & 13 & 23 \end{array} \\ \begin{bmatrix} 1 & 1.41 & 1.45 \\ 0.71 & 1 & 1.03 \\ 0.69 & 0.97 & 1 \end{bmatrix} \end{array}$$

In the following sections, we will 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,* and therefore:

- $S_{top}(Q_1, I) = 100\%$

Furthermore, the pixel coordinates (row,col) of the three objects are: obj_1=(40,59); obj_2=(180,117); obj_3=(152,305).

Direction relations between objects in Q_1:

- Position Relation Matrix

$$\begin{array}{c} \\ 12 \\ 13 \\ 23 \end{array}\begin{array}{c} \begin{array}{ccc} obj_1 & obj_2 & obj_3 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

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

$$C_{or} = \frac{\sum(I-\bar{I})(Q-\overline{Q})}{\sqrt{\sum(I-\bar{I})^2(Q-\overline{Q})^2}} = \frac{-2.6447}{\sqrt{2.4489 * 2.8889}} = -0.99$$

where I and Q are the position relation matrices for the image and query scenes respectively, and

$$\therefore S_{or}(Q_1, I) = \frac{-.99+1}{2} = .5\%$$

<u>Distance relations between objects in Q_1 are calculated similarly to above:</u>

- Distance Ratio Matrix

$$\begin{array}{c c} & \begin{array}{ccc} 12 & 13 & 23 \end{array} \\ \begin{array}{c} 12 \\ 13 \\ 23 \end{array} & \begin{bmatrix} 1 & 1.78 & 1.25 \\ 0.56 & 1 & 0.70 \\ 0.80 & 1.42 & 1 \end{bmatrix} \end{array}$$

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

$$C_{dist} = \frac{\sum(I-\bar{I})(Q-\overline{Q})}{\sqrt{\sum(I-\bar{I})^2(Q-\overline{Q})^2}} = \frac{0.6288}{\sqrt{0.8871 * 1.1421}} = 0.6247$$

$$\therefore S_{dist}(Q_1, I) = \frac{.6247+1}{2} = 81.2\%$$

Finally, assuming equal weighting for all four components in Eq. 1 (i.e. shape similarity is as important as topology, orientation, and distances), the value of S_{met} for Q_1 becomes:

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

4.3 Spatial Relations for Query Scene 2

<u>Topological relations between objects in Q_2:</u>

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,* and therefore:

- $S_{top}(Q_2, I) = 100\%$

Furthermore, the pixel coordinates (row,col) of the three objects are: obj_1=(239,82); obj_2=(103,113); obj_3=(153,305).

<u>Direction relations between objects in Q_2:</u>

- Position Relation Matrix

$$\begin{array}{c} \\ 12 \\ 13 \\ 23 \end{array} \begin{array}{c} \begin{array}{ccc} obj_1 & obj_2 & obj_3 \end{array} \\ \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \end{array}$$

Therefore, the value of S_{or} for Q_2 is $S_{or}(Q_2, I) = \frac{1+1}{2} = 100\%$. Indeed, since Q_2 is a 180 degree rotation of *I,* all position relations are 100% similar to the position relations among the corresponding objects in *I*.

<u>Distance relations between objects in Q_2:</u>

- Distance Ratio Matrix

$$\begin{array}{c} \\ 12 \\ 13 \\ 23 \end{array} \begin{array}{c} \begin{array}{ccc} 12 & 13 & 23 \end{array} \\ \begin{bmatrix} 1 & 1.71 & 1.42 \\ 0.58 & 1 & 0.83 \\ 0.70 & 1.20 & 1 \end{bmatrix} \end{array}$$

Therefore, the value of S_{dist} for Q_2 is $S_{dist}(Q_2, I) = \frac{0.7337+1}{2} = 86.7\%$

Using the above information, the value of S_{met} for Q_2 becomes:

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

4.4 Spatial Relations for Query Scene 3

Topological relations between objects in Q_3:

All objects in Q_3 are disjoint, i.e. topological relations in Q_3 are 100% similar to topological relations among the corresponding objects in I, therefore:

- $S_{top}(Q_3, I) = 100\%$

Furthermore, the pixel coordinates (row,col) of the three objects are: obj_1=(223,309); obj_2=(127,83); obj_3=(143,211).

Direction relations between objects in Q_3 are:

- Position Relation Matrix

$$\begin{array}{cc} & \begin{array}{ccc} obj_1 & obj_2 & obj_3 \end{array} \\ \begin{array}{c} 12 \\ 13 \\ 23 \end{array} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \end{array}$$

Therefore, the value of S_{or} for Q_3 is $S_{or}(Q_3, I) = \frac{.59 + 1}{2} = 79\%$.

Distance relations between objects in Q_3 are:

- Distance Ratio Matrix

$$\begin{array}{cc} & \begin{array}{ccc} 12 & 13 & 23 \end{array} \\ \begin{array}{c} 12 \\ 13 \\ 23 \end{array} & \begin{bmatrix} 1 & 0.52 & 0.53 \\ 0.94 & 1 & 1.02 \\ 0.90 & 0.98 & 1 \end{bmatrix} \end{array}$$

Thus, $S_{dist}(Q_3, I) = \frac{-.326 + 1}{2} = 33.7\%$

Using the above information, the value of S_{met} for Q_3 becomes:

- $S_{met}(Q_3,I)$=82.3 * .25 + 100 * .25 + 79 * .25 + 33.7 * .25 = **73.8%**

4.5 Spatial Relations for Query Scene 4

Topological relations between objects in Q_4 are the following:

- obj_1 and obj_2 are disjoint;
- obj_2 and obj_3 are disjoint;
- obj_1 is contained by obj_3.

Therefore, the topological relation between the first two pairs of objects in Q_4 is 100% similar to the topological relations between the corresponding objects in *I*, while the topological relation between the third pair of objects is 0% similar. The value of S_{top} is:

- $S_{top}(Q_4, I) = \frac{100+100+0}{3} = 66.7\%$

Furthermore, the pixel coordinates (row,col) of the three objects are: obj_1=(139,251); obj_2=(129,85); obj_3=(165,251).

Direction relations between objects in Q_4 are:

- Position Relation Matrix

$$\begin{array}{c} \\ 12 \\ 13 \\ 23 \end{array} \begin{array}{c} \begin{array}{ccc} obj_1 & obj_2 & obj_3 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

Therefore, the value of S_{or} for Q_4 is $S_{or}(Q_4, I) = \frac{.62+1}{2} = 81\%$.

Metric relations between objects in Q_4 are:

- Distance Ratio Matrix

$$\begin{array}{c} \\ 12 \\ 13 \\ 23 \end{array}\begin{array}{c} \begin{array}{ccc} 12 & 13 & 23 \end{array} \\ \begin{bmatrix} 1 & 0.16 & 1.02 \\ 6.40 & 1 & 6.53 \\ 0.98 & 0.15 & 1 \end{bmatrix} \end{array}$$

Thus, $S_{dist}(Q_4, I) = \frac{-.3+1}{2} = 35\%$

Using the above information, the value of S_{met} for Q_4 becomes:

- $S_{met}(Q_4,I)$= (82.3 * .25) + (66.7 * .25) + (81 * .25) + (35 * .25) = **66.3%**

4.6 Comparison of results

The results of all 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 degree rotation of the image, sketched at a significantly reduced scale. This result demonstrates the ability of our varying baseline approach to handle 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

The approach introduced in this paper allows us to compare the orientation and distance content of scenes regardless of pre-defined cardinal directions, and scale variations. This is an significant advantage that becomes especially important for image query applications, where scale and orientation may vary by large and often unknown amounts. In this paper we also presented a framework where these metrics can complement our previous work on shape-based queries to provide a comprehensive approach for scene-based image queries. Furthermore, the relationships used to populate the position relation and distance ratio matrices are easy to calculate. Indeed, the "left-of"/"right-of" and distance information used to populate these two matrices is readily available once the individual object MBRs are defined. This makes the approach presented here computationally efficient in addition to being versatile in terms of cardinality and scale.

The working example of Section 4 presented a practical illustration of the calculation of the similarity metrics between four query scenes and a given image. The results of this example demonstrated how our approach overcame substantial differences in cardinality and scale to identify the most similar scene. Our future plans include the analysis of gradual changes in the overall scene similarity metric in spatiotemporal applications to monitor the movement of objects in a scene. Furthermore, we plan to investigate the use of the position relation matrix to model the distribution of point data in larger scenes.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation through CAREER grant number IIS-9702233 and Digital Government grant number DGI-9983445, by the National Imagery and

Mapping Agency through NURI grant number NMA202-98-1-1113, and by CNR through 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.

Blaser A., 2000. *Sketching Spatial Queries*. Ph.D. Dissertation Thesis, University of Maine, Orono, Maine.

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 Thesis, University of Maine, Orono, Maine.

Egenhofer, M. and Al-Taha, K., 1992. Reasoning About Gradual Changes of Topological Relationships. In: A. Frank (Editor), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, Pisa, Italy. Springer-Verlag, Berlin, pp. 196-219.

Egenhofer, M. and Franzosa, R., 1991. Point-Set Topological Spatial Relations. *International Journal of Geographical Information Systems*, 5(2): 161-174.

Egenhofer, M. and Mark, D., 1995. Naive Geography, Spatial Information Theory - A Theoretical Basis for GIS. In: A. Frank and W. Kuhn (Editors), Lecture Notes in Computer Science. Springer-Verlag, Berlin, Serrering, Austira, pp. 1-15.

Egenhofer, M. and Sharma, J., 1992. Topological Consistency, 5th International Symposium on *Spatial Data Handling,* Charleston, SC, pp. 335-343.

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* (in press).

Gruen, A. and Agouris, P., 1994. Linear Feature Extraction by Least Squares Template Matching Constrained by Internal Shape Forces. *International Archives of Photogrammetry & Remote Sensing*, 30(Part 3/1): 316-323.

Gruen, A., Agouris, P. and Li, H., 1995. Linear Feature Extraction with Dynamic Programming and Globally Enforced Least Squares Matching. In: A. Gruen, O. Kuebler and P. Agouris (Editors), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Birkhaeuser Verlag, pp. 83-94.

Gruen A., O. Kuebler, and P. Agouris (eds.), 1995b. *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhaeuser Verlag, Basel.

Gruen A., E. Baltsavias, and O. Henricsson (eds.), 1997. *Automatic Extraction of Man-Made Objects from Aerial and Space Images II*, Birkhaeuser Verlag, Basel.

Lukes G. (ed.), 1998. *Proceedings DARPA Image Understanding Workshop,* Monterey, CA. Morgan Kaufmann.

Pentland, A., Picard, R.W. and Scarloff, S., 1996. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*.

Sclaroff, 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, Peurto Rico, pp. 2-9.

Shariff, A.R., 1996. Natural Language Spatial Relations: Metric Refinement of Topological Properties, University of Maine, Orono, ME, USA.