Defining and Comparing Content Measures of Topological Relations

FRANCISCO GODOY AND ANDREA RODRÍGUEZ

Department of Information Engineering and Computer Science

University of Concepción and

Research Center for Web Computing

University of Chile

Edmundo Larenas 215, Concepción, Chile.

{fgodoyf,andrea}@udec.cl

Abstract

This work defines and compares three content measures that characterize topological relations between rectangular regions in a two-dimensional space. These content measures use simplified views of spatial objects in order to create an efficient mechanism for capturing the topological content of spatial configurations. The content measures are compared based on the correlation between two similarity rankings: (1) a similarity ranking defined in terms of the distance of content-measure values and (2) a similarity ranking defined in terms of the error of the geometric adjustment between pairs of objects. The correlation between similarity rankings is used as indicator of how well these content measures characterize topological relations. Such content measures provide mechanisms for creating efficient methods to describe and access information on the basis of the topological content of spatial configurations.

Keywords: topological relations, content measures, similarity function.

1 Introduction

The complexity and volume of spatial information available in current structured, semi-structured, and non-structured data repositories have made content-based retrieval a challenging and important area of investigation. The main idea of a content-based search of spatial information is to find instances in a data repository whose content description is most similar to the content of a user request. Fundamental to a solution to this type of problem is, therefore, to define an appropriate content description that characterizes and allows us to compare spatial information.

This study aims to define a systematic way to characterize spatial information, in particular, topological relations in spatial configurations. Unlike previous studies on content-based retrieval in image databases [1, 9, 18, 19, 38, 39, 42], this work focuses on the characterization of configurations that are seen as a combination of objects that stand in particular spatial relations to each other. In order to find desired configurations, systems must find object instances in spatial databases that satisfy the constraints defined by the spatial relations of a user request. This search of object instances is often done on the basis of information that consists of objects stored in relational tables and organized by thematic layers with spatial indexing methods. In these systems, queries are typically answered as cascaded *spatial joins* [2, 26, 29-31].

This paper describes three content measures that distinguish topological relations. By defining content measures of spatial relations, this work contributes to the definition of new mechanisms for spatial information organization and retrieval, so that queries with variable and large number of objects, such as queries expressed by sketches [4, 16], can be efficiently solved. The proposed content measures distinguish topological relations with a simplified view of spatial objects and, therefore, they try to minimize the computational cost of processing topological relations. Characterizing spatial relations between objects is useful for comparing configurations, since configurations are composed of a variable number of objects (i.e., a variable number of relations), and configurations can be seen as an aggregation of individual relations.

The problem of comparing spatial relations is not new [8, 22, 27, 32]; however, to the best of our knowledge, none of the previous studies have attempted to define a single content measure that distinguishes topological relations, making this content measure suitable for content-based indexing schemas that consider not only positional information, but also spatial relations [36]. Some of the previous studies combine multiple content components (e.g., angle, topology, and distance) [3, 28, 32], which may be highly sensitive to the way these components are combined. This work uses a quantitative approach to characterize objects' interrelations in terms of metric refinements of topology relations. In this sense, it follows closely the ideas derived from Egenhofer and Shariff's work [37] that
[Warning: Draw object ignored]

made metric refinements of topological relations; however, instead of defining measures to refine each of the topological relations, it pursues the definition of a single content measure that distinguishes topological relations.

The focus of the study are topological relations between regions [7, 10, 34, 35], since among different types of spatial relations, topological relations have been pointed out as the principal way in which people describe configurations [14]. Additionally, topological relations are invariant under continuous transformations of translation, rotation and scaling, which are desirable properties of content measures that describe configurations expressed by visual examples [4]. This work uses the simplified and common representation of objects (i.e., MBRs) in current spatial indexing schemas of Geographic Information Systems (GISs). Although this simplification of objects misses some details, it is broadly used and computational desirable, and it usually sufficient for finding objects in current GISs.

The proposed content measures are evaluated with an independent framework for comparing spatial configurations. Thus, this work makes a distinction between content measures and similarity functions. While content measures characterize and can be used for comparing spatial relations, a similarity function compares spatial relations without being able to say anything about the type of relations between objects. We will say that the content measures are good candidates for capturing and comparing topological information if the difference between values of content measures has a strong correlation with the independent similarity function. The similarity function used in this work takes ideas from image processing and uses principles of geometric adjustment between corresponding objects in spatial configurations [23].

The organization of this paper is as follows. Section 2 reviews related studies that address the description and comparison of spatial relations. Section 3 describes the characterization of MBRs as one-dimensional values, and Section 4 presents the three content measures that are proposed in this paper. Section 5 introduces the similarity function that is used for comparing content measures. Subsequently, Section 6 presents experimental results when comparing content measures. Conclusions and future work are given in Section 7.

2 Related Work

Many studies in the domain of image databases have compared objects' arrangements based on variations of *2D-strings*. 2D-strings represent configurations with a sequential structure for each encoded dimension [8, 24, 25]. Query processing using this structure is carried out as a string matching. Such string matching is possible only when users specify [Warning: Draw object ignored] queries by the schema of relations according to which 2D-strings are built, and images are composed of a predefined set of objects. A different string-based representation treats topology, orientation, and distance between objects' MBR as interval relations in two dimensions [27]. This type of representation defines a similarity function as inversely proportional to the number of changes that are needed to make two strings equivalent. In a similar way, a 3x3 matrix was used to determine the orientation relation as the proportional area in the quadrants defined by the orthogonal projection of a reference object's MBR [22]. Similarity between orientation relations is then defined by the inverse cost of transforming the matrix representation of a relation into the other matrix representation. In general, methods based on 2D-strings and their variations handle variations in scale and translation, but they are sensitive to rotation [18].

Using an object-oriented perspective, where configurations are sets of objects and sets of these objects' interrelations, some studies represent configurations and queries using Attribute Relation Graphs (ARGs) [3, 28, 32]. In these graphs, spatial relations are represented quantitatively by the distance and angle between centroids of objects, and qualitatively by the symbolic representation of spatial relations, such as the topological relations defined by Egenhofer and Franzosa [11, 13] or by Randell *et al.* [35]. For ranking configurations, a similarity function is defined, which depends on the representation type of spatial relations. For quantitative representations of spatial relations, such as the angle between MBRs, similarity is defined as the inverse of the difference between representations [3, 32]. Another approach considers the distance within a conceptual neighborhood [28]. For example, consider Figure 1 of conceptual neighbors of topological relations between regions derived from the concept of *gradual change* [12, 15]. Conceptual neighbors are relations connected by a line in this Figure, and they are considered to be more similar than relations that are not directly connected in the graph.

Related to the concept of *gradual change*, Bruns and Egenhofer [5] compared spatial scenes. Given two scenes (i.e., spatial configurations) of equal number of objects, they suggested that similarity could be determined by the minimum set of *gradual changes* that are needed to transform one scene into the other one. Although their work presents a sensible definition of similarity, it does not check whether or not this minimum set of changes is unique. Likewise, it does not discuss degrees of relevance that may affect different types of changes. Even if relevance weights were associated with these changes, it may be difficult to obtain a systematic strategy to determine these weights.

Focusing on topological relations, one study explores metric refinements of topological relations as they match with terms used in natural language, such as *going through* and *goes up to* [17]. This study defines ten quantitative measures that characterize topological relations based on metric properties, such as *length*, *area*, and *distance*. The

[Warning: Draw object ignored]

combination of these measures gives an indication of the topological relations and their associated terms in natural language. A disadvantage of using these measures for similarity assessment is the need to combine different measures to completely characterize a topological relation.

3 Characterizing MBRs

This work defines content measures of topological relations based on basic parameters that characterize individual MBRs and pairs of MBRs. Two basic parameters allow us to characterize MBRs as single dimensional values: areas and diagonals. They are simplified views of MBRs, and there is no one-to-one correspondence between these parameters and MBRs. In addition, they do not provide any information about the shape of the MBRs, which can vary from square to thin rectangles. Figure 2 shows two different MBRs with equivalent area (Figure 2a) and two different MBRs with equivalent diagonal (Figure 2b). Both examples in Figure 2 illustrate the effects that occur when areas and diagonals are used as the simplified representation of MBRs.

A second set of parameters allows us to characterize relations between pairs of MBRs. These parameters are the area and diagonal of the MBR that is created by the *union* of two MBRs (A **[Warning: Missing symbol F0C8]** B), the area of the MBR that is created by the *intersection* of two MBRs (A **[Warning: Missing symbol F0C7]** B), and the *minimum internal* (d_i) and *external* (d_e) *distances* between boundaries of MBRs (Figure 3). In this definition, distance is calculated between boundaries of MBRs. While *external distance* is defined as the minimum distance between MBRs that are *disjoint* (d_e([Warning: Missing symbol F064]A,[Warning: Missing symbol F064]B)), *internal distance* is the minimum distance between boundaries when one of the objects *is inside* or *contains* the other object (d_i([Warning: Missing symbol F064]A,[Warning: Missing symbol F064]B)). The distance between MBRs that stand in a *meet*, *overlap*, *covers*, or *covered_ by* relation is considered equal to zero.

4 Content Measures

The definition of the content measures characterizes topological relations considering metric characteristics of objects' arrangements such that the following two tasks can be accomplished:

1. Distinguishing the degree of *separation* and *overlapping* between MBRs. Figure

4 shows two pairs of objects *(A,B)* and *(C,D)* that are separated by the same
[Warning: Draw object ignored]

distance *d*. Although both configurations satisfy the same topological relation (i.e., *disjoint*), the pair (A,B) is considered more separated than the pair *(C,D)*, due to independence of scale [17].

1. Making explicit the influence of each object in the configuration. This influence affects the degree of separation and overlapping when considering small versus large objects satisfying the *disjoint* or *inside/contains* relations. Such influence of size in spatial relations creates asymmetric definitions of content measures, since the effect of the metric refinement on the topological relation between A and B is not necessarily the same as the effect on the relation between B and A (Figure 5). This type of asymmetry has been addressed by previous studies in the area of spatial reasoning, where distance has been defined as an asymmetric phenomenon [14].

4.1 Area-Based Content Measure

The first content measure F_a considers the normalization of the area of each MBR by the area of the union of the MBRs (Equation 1). Values of this function are larger than 0 and less or equal than 1.

[Warning: Draw object ignored] (1)

The content measures F_a is unable to distinguish *covered_by* and *covers* from *inside* and *contains*, respectively, since the union of MBRs is the same for all these relations. If F_a (A,B) + F_a (B,A) [Warning: Missing symbol F0B3] 1, the relation between A and B is *non-disjoint*, since the area of (A **[Warning: Missing symbol F0C8]** B) must be smaller than the sum of the area of A and the area of B. In the extreme case, the area of (A **[Warning: Missing symbol F0C8]** B) is equal to the sum of the area of A and the area of B when A *meets* B (Figure 6a). Note, however, that a *non-disjoint* relation does not imply that F_a (A,B) + F_a (B,A) [Warning: Missing symbol F0B3] 1 (Figure 6b); that is, this is not a double implication.

4.2 Diagonal-Based Content Measure

The second content measure F_d uses the diagonals instead of areas of MBRs (Equation 2). Values of this content measure are also larger than 0 and less than or equal to 1.
[Warning: Draw object ignored]

[Warning: Draw object ignored] (2)

As in the case of F_a, F_d does not distinguish between *inside* and *contains* from *covered_by* and *covers*, respectively, since the union of MBRs is the same for all cases. If $F_d\,(A,B) + F_d\,(B,A) < 1$, A and B must be *disjoint*. Consider Figure 7 where object A is fixed and different objects B with increasing diagonals a, b, and c are illustrated. In cases when the diagonal of the object B is equal to a or b (i.e., less than c), $F_d\,(A,B)$ + $F_d\,(B,A)$ must be less than 1, since the diagonal of (A **[Warning: Missing symbol F0C8]** B) is always equal to $d+c$, and $d+c$ is larger than $d+a$ or $d+b$. These cases represent *disjoint* relations. In the extreme case when the diagonal of B is equal to c and, therefore, $F_d\,(A,B) + F_d\,(B,A) = 1$, A may *meet* or may be *disjoint* from B.

4.3 Mixed Content Measure

With the goal of being able to distinguish more topological relations than the first two content measures, the last content measure F_m combines areas, diagonals, and distances (Equation 3). It considers that distance is a measure of *disjointness* while area is a measure of *overlapping*.

[Warning: Draw object ignored] (3)

Unlike the two first content measures, F_m distinguishes eight topological relations. The content measure, however, is unable to capture metric refinements of *meet* relations. Figure 8 shows the values of F_m according to the topological relations of Table 1. In this Figure, curves bound the topological relations *inside* and *contains*, which were experimentally determined by using extreme cases and defining their respective parametric equations.

4.4 Content Measures under Continues Transformations

As was mentioned above, topological relations are invariant under continuous transformations of scaling, translation, and rotation. Consequently, the behavior of content measures is analyzed when continuous transformations occur. In this analysis, objects' shapes in configurations do not change, but the scale or the frame of reference is modified.

It is easy to prove that all three content measures are invariant under changes in scale and translation. Translations do not modify the basic parameters (i.e., areas, diagonals, and distances) upon which the content measures are defined and, therefore, the content measures are invariant under continuous translations. In scaling, the scale factor that [Warning: Draw object ignored] is applied over individual objects is also applied to the union, intersection, or distance between objects. Consequently, the normalizations in Equations 1-3 cancel any scale factor applied to objects in configurations. For example, in Figure 9 the relations between A and B is the same as the relation between C and D.

Figure 10 illustrates the idea of continuous rotations. In Figure 10a, A is the rotation center of the configuration, so object A does not change its location, whereas object B changes its relative location with respect to A. Areas or diagonals of objects do not change; however, the distances as well as the areas or diagonals of the union or intersection of objects vary such that the content measures may change as well. We show experimentally the effect of rotation for the three content measures in the graph of Figure 10b, where content measures have been normalized. The graph indicates that although all three content measures are affected by rotation, rotation has the strongest impact on F_a. In cases when the rotation is 90[Warning: Missing symbol F0B0], 180[Warning: Missing symbol F0B0], or 270[Warning: Missing symbol F0B0], none of the content measures is affected.

In addition to analyzing the effect of rotation, translation, and scaling, an interesting analysis is to evaluate free movements of objects with respect to changes in values of content measures. Unlike translation where both objects continuously moves, this type of analysis considers free movements of one of the objects. Such an analysis reflects the homogeneity of the relation space. In a homogeneous space, distances in one part of the space (i.e., differences in content-measure values as points in a 2D space) could be correlated with distances in another part of the space. This type of analysis is important when defining a similarity function based on content measures or when applications deal with moving objects with imprecise positional information [33, 41]. In the case of a similarity function, a homogeneous space could easily define a similarity function in terms of distances in the space. In applications with moving objects, for example, one might need to design efficient mechanisms that do not store the complete sequence of movements and relations, but store those states that represent changes in the values of content measures.

To illustrate the behavior of the content measures for different transition states of moving objects, consider Table 2 where two objects are moved continuously from *disjoint* to *inside,* and values of content measures are given for the 6 possible topological relations that occur along the movement. Figure 11 and Figure 12 complement Table 2 with graphs that describe the changes of content measures for two different continues movements. In these graphs, values on the x-axis are the constant variations of objects' positions, and values on the y-axis are the normalized values of content measures (i.e., values between 0 and 1).

[Warning: Draw object ignored]

The first type of movement (Figure 11a) consists in passing an object B through another object A, such that the relations between A and B in sequential order are *disjoint*, *meet*, *covers*, *contains*, *covers*, *meet*, and *disjoint*. In Figure 11b variations of content measures F_a and F_d follow quadratic curves with two break points: when the relation changes from *overlap* to *covered_ by* and, conversely, when it changes from *covered_ by* to *overlap*. While the object is *covered_ by* or *inside* the other object, both content measures are constant and equal to 1. Both quadratic curves are continuous from *disjoint* to *overlap* or from *overlap* to *disjoint*, being the curve that represents changes of F_d less quadratic than curve of F_a. Changes of content measure F_m, in contrast, are lineal with 6 break points: the transition from *disjoint* to *meet*, the transition from *meet* to *covered_ by*, the transition from *covererd_ by* to *inside*, and their corresponding converse transitions.

A second movement is presented in Figure 12. In this movement, object B approaches and passes object A, all while maintaining a *disjoint* relation with A. Figure 12b shows that content measures do not always change continuously when disjoint objects continuously change their locations. The content measure F_a has different break points, whereas content measures F_d and F_m change continuously, with changes of F_m being linear with respect to the distance between objects.

5 A Framework for Comparing Content Measures

An intuitive way to define similarity between spatial configurations is as the inverse of the difference between configurations. Distance is a typical measure of difference, whose metric property of triangle inequality is useful for defining data organization and access methods [6, 21]. In image processing, distance can be used for evaluating the quality of adjustment between images. Two images that are thought to represent the same space are considered completely adjusted if the distances between control points in an image and control points in the other image are zero. Since images may suffer deformations, transformations of rotation, scaling, and translation are applied to the control points such that these points can adjust [23].

This work follows the strategy of image adjustment for defining a similarity function between spatial configurations. Unlike image adjustment, however, this work deals with configurations that are composed of spatial objects, that is, points, lines, and regions. Consequentially, this work applies transformations of rotation, translation, and scaling while preserving the shapes of objects and their topological interrelations. Analogous to image adjustments, control points are used to adjust configurations. These control points are extracted from the geometric representation of objects' MBRs. For example, in a first instance, the extreme four vertices of MBRs are the control points in a configuration. [Warning: Draw object ignored]

Figure 13 shows the graphic schema of a pair of objects in a reference system (x, y) that is transformed into a reference system (X, Y), and Equation 4 is the general expression to make that this transformation consider rotation, translation, and scaling of objects without producing deformations.

[Warning: Draw object ignored] (4)

Equation 5 expresses Equation 4 for the eight control points (four vertices for each MBR) in a configuration like the one illustrated in Figure 13a. This Equation 5 rewrites Equation 4 to handle 4 unknowns (i.e., a, b, c, d) such as a system that can be solved by the least squares approximation of the form $[A]$[Warning: Missing symbol F0B4]$[B]$[Warning: Missing symbol F0BB]$[C]$ (Equation 6) [40].

[Warning: Draw object ignored] (5)

[Warning: Draw object ignored] (6)

In this approach to adjusting configurations, points associated with an object's vertices are made correspond to vertices of a target object. Since the right correspondence is unknown, different combinations of vertices (i.e., points) were analyzed, and the combination with the minimum error (i.e., distance) after the adjustment was considered correct. For example, eight ways to assign vertices are possible in a configuration with 8 points: four arising from a rotation of [Warning: Missing symbol F070]/2 radians (Figure 14a), and four from a rotation of [Warning: Missing symbol F070]/2 radians in a mirror effect (Figure 14b).

The adjustment error is determined as a function of the position difference of corresponding points. This error is normalized by the sum of the diagonals of the unions of original and target MBRs, respectively (Figure 15, Equation 7). This normalization allows us to compare configurations of pairs of objects independently of scale.

(7)

This approach to comparing configurations is sensitive to the way objects are represented. Therefore, MBRs' representation with eight points was also analyzed and experimentally compared in the next Section. A difference between using eight instead of four points per MBR is that with eight points, sixteen different correspondences of points need to be checked before finding the best adjustment. These sixteen different possibilities arise from eight rotations of [Warning: Missing symbol F070]/4 radians and eight rotations of [Warning: Missing symbol F070]/4 radians in a mirror effect (Figure 16). Us-

[Warning: Draw object ignored]

ing eight points for MBRs' representation overcomes problems that are presented in the representation with four points and rotations of [Warning: Missing symbol F070]/2 radians. For example, Figure 17 illustrates the difference of an adjustment with four-points or eight-points representation and [Warning: Missing symbol F070]/4 radian rotation.

6 Comparing Content Measures

The analysis consists in applying the three content measures to a set of configurations that are composed of two MBRs (i.e., configurations with one topological relation). Then, these configurations are combined to create all possible pairs of different configurations. For such pairs of configurations, the geometric adjustment between configurations and the distance between configurations' content measures were determined. Finally, the correlation between the adjustment error and the distances of the content-measure values are used for comparing content measures (Figure 18).

6.1 Data Set

The experiments were carried out with a data set created with all possible MBRs that fix in boxes of 2x2, 3x3, 4x4, 5x5, and 6x6 cells. For example, Figure 19 shows the nine MBRs derived from a 2x2 box.

For each set of MBRs, configurations composed of two different MBRs were created. From this set of initial configurations, the experiments considered a subset of configurations, where none of these configurations are equivalent under transformations of scaling, rotation, and translation. For example, from the nine MBRs of the 2x2 box, it is possible to create a set of 36 configurations, which is then reduced to just eight different configurations (Figure 20).

Using the final set of configurations, comparisons between different configurations were performed. The number of comparisons depends on the number of configurations in each set (Table 3). This table includes a cell box of 7x7, which case was not used in the experiments for its computational cost.

6.2 Comparison

The correlations between the distance of content-measure values and the error of the geometric adjustment are presented in Table 4, where the geometric adjustment was determined by using two representations of MBRs: (1) four points per MBR and (2) eight points per MBR. The results indicate that the correlation when using eight points

[Warning: Draw object ignored]

per MBR was better than the correlation when using four points. This tallies with the fact that a more accurate representation of objects provides more information for geometric adjustment. In all cases, the content measures F_m gives better correlations than measures F_a and F_d due to its ability to distinguish more topological relations.

An analysis that tries to explain the variations of correlation among different sets of configurations considers the distribution of topological relations in the sets (Figure 21).

As the size of the cell box increases, the number of configurations as well as the number relations *disjoint*, *overlap*, and *inside&contains* also increases. This type of distribution is in agreement with situations in real geographic information systems, where the most frequent relation is *disjoint* [20]. In order to understand the effect of each relation in the three content measures, the correlation between content-measure values and geometric adjustments was re-calculated by eliminating topological relations one-by-one. Given that the set of configurations in a cell box of 2x2 is small and does not include all types of topological relations, configurations derived from a cell box of 2x2 were ignored in this analysis. Results of correlations for content measures F_a, F_d, and F_m are shown in Figure 22, 23, and 24, respectively.

Figure 22 shows that the *disjoint* relation has a positive effect on the correlation between the content measure F_a and the geometric adjustment. The relation *overlap*, in contrast, negatively affects the content measure, since in all cases, the correlation after eliminating the *overlap* relation was larger than the correlation with this relation.

As in the case of the content measure F_a, the *disjoint* relation positively affects the content measure F_d. The other *non-disjoint* relations have a similar behavior, which indicates an even capacity of this content measure to characterize *non-disjoint* relations (Figure 23).

Finally, disjoint relations also have a positive effect on the content measure F_m (Figure 24). The highest correlation was found when eliminating relation *meet*, which indicates the negative effect of these relations. As indicated in Section 3, F_m does not distinguish among metric differences of the *meet* relations, so *meet* relations have a negative effect on this content measure.

7 Conclusions and Future Work

This paper describes the definition and comparison of three content measures for topological relations: (1) F_a based on the areas of the union of MBRs and the area of individual MBRs, (2) F_d based on the diagonal of the union of MBRs and the diagonal of individual MBRs, and (3) F_m based on the area of the intersection of MBRs, area of individual [Warning: Draw object ignored]

MBRs, diagonal of individual MBRs, and distances between MBRs. These measures were compared by using their correlations with the errors of the geometric adjustment between configurations. Experimental results indicate that the relation best characterized by the content measures is the *disjoint* relation. The correlation between content measures and similarity function shows that F_m better distinguishes topological relations, followed by F_d and F_a.

The possibility to compare content measures and analyze the content measures' behavior with respect to different topological relations has allowed us not only to evaluate these defined content measures, but also to define a strategy for comparing new content measures.

Left for future work is the study of how to combine content measures of topological relations for comparing complex spatial configurations with more than two objects. An issue in defining such combination is the degree of homogeneity of the relation space. In such a space, small differences may not be equivalent depending on the location in the space, such that a traditional combination of distance values may not be adequate for defining a similarity value. Another natural extension to this work is the use of volume for defining content measures in a 3D space.

Acknowledgments

This work has been funded by Nucleus Millenium Center for Web Research, Grant P01-029-F, Mideplan, Chile.

[Warning: Draw object ignored]

References

1. G. Amato, G. Mainetto and P. Savino. "An Approach to a Content-Based Retrieval of Multimedia Data," *Multimedia Tools and Applications*, Vol. 7(1/2): 5-32, 1998.

2. L. Becker, A. Giesen, K. Hinrichs and J. Vahrenhold. "Algorithms for Performing Polygonal Map Overlay and Spatial Join on Massive Data Sets," R.H. Güting, D. Papadias, and F. Lochovsky, editors, in *Advances in Spatial Databases, 6th International Symposium, SSD '99*, Hong Kong, China, Springer, 1999.
3. S. Berretti, A. Del Bimbo and E. Vicario. "The Computational Aspect of Retrieval by Spatial Arrangement," in *International Conference on Pattern Recognition*, Barcelona, España, 2000.
4. A. Blaser. "Sketching Spatial Queries", Department of Spatial Information Science and Engineering, University of Maine, Orono, ME, Ph.D. Thesis, 2000.
5. T. Bruns and M.J. Egenhofer. "Similarity of Spatial Scenes," M. Kraak and M. Molenaar, editors, in *Seventh International Symposium on Spatial Data Handling (SDH '96)*, Delft, The Netherlands, 1996.
6. E. Chávez, G. Navarro, R. Baeza-Yates and J.L. Marroquín. "Proximity Searching in Metric Spaces," *ACM Computing Surveys*, Vol. 33(3): 273-321, 2001.
7. E. Clementini, J. Sharma and M.J. Egenhofer. "Modeling Topological Spatial Relations: Strategies for Query Processing," *Computers and Graphics*, Vol. 18(6): 815-822, 1994.
8. G. Costagliola, G. Tortora and T. Arndt. "A Unifying Approach to Iconic Indexing for 2-D and 3-D Scenes," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 4(3): 205-222, 1992.
9. C. Diamantini and M. Panti. "A Conceptual Indexing Method for Content-Based Retrieval," in *10th International Workshop on Database & Expert Systems Applications*, Firenze, Italy, IEEE Press, 1999.
10. M.J. Egenhofer. "Reasoning about Binary Topological Relations," O. Günter and

[Warning: Draw object ignored]

H.-J. Schek, editors, in *Second Symposium on Large Spatial Databases*, Zurich, Switzerland, Springer-Verlag, 1991.

11. M.J. Egenhofer and R. Franzosa. "Point-set topological spatial relations," *International Journal of Geographical Information Systems*, Vol. 5(2): 161-174, 1991.

12. M.J. Egenhofer and K. Al-Taha. "Reasoning About Gradual Changes of Topological Relationships," A. Frank, I. Campari, and U. Formentini, editors, in *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, Pisa, Italy, Springer-Verlag, 1992.

13. M.J. Egenhofer, E. Clementini and P. Di Felice. "Topological Relations Between Regions with Holes," *International Journal of Geographical Information Systems*, Vol. 8(2): 129-142, 1994.

14. M.J. Egenhofer and D. Mark. "Naive Geography," A. Frank and W. Kuhn, editors, in *COSIT '95*, Semmering, Austria, Springer-Verlag, 1995.

15. M.J. Egenhofer and D. Mark. "Modeling Conceptual Neighborhoods of Topological Line-Region Relations," *International Journal of Geographical Information Science*, Vol. 9(5): 555-565, 1995.

16. M.J. Egenhofer. "Query Processing in Spatial-Query-By-Sketch," *Journal of Visual Languages and Computing*, Vol. 8(4): 403-424, 1997.

17. M.J. Egenhofer and R. Shariff. "Metric Details for Natural-Language Spatial Relations," *ACM Transactions on Information Systems*, Vol. 16(4): 295-321, 1998.

18. E. El-Kwae and M. Kabuka. "A Robust Framework for Content-Based retrieval by Spatial Similarity in Image Databases," *ACM Transactions on Information Systems*, Vol. 17(2): 174-198, 1999.

19. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petrovik, D. Steele and P. Yanker. "Query by Image and Video Content: The QBIC System," *IEEE Computer*, Vol. 28(9): 23-32, 1995.

[Warning: Draw object ignored]

20. J. Florence and M.J. Egenhofer. "Distribution of Topological Relations in Geographic Datasets," in *ACSM/ASPRS*, Baltimore, MD, 1996.
21. V. Gaede and O. Günter. "Multidimensional Access Methods," *ACM Computing Surveys*, Vol. 30(2): 170-231, 1998.
22. R. Goyal and M.J. Egenhofer. "Similarity of Direction Relations," C. Jensen, et al., editors, in *Seventh International Symposium on Spatial and Temporal Databases*, Los Angeles ,CA, Springer-Verlag, 2001.
23. D. Hearn and M.P. Backer, Computer Graphics. Prentice Hall International Editions: 1994.
24. S.-Y. Lee and F.-J. Hsu. "Spatial Reasoning and Similarity Retrieval of Images Using 2D C-Strings Knowledge Representation," *Pattern Recognition*, Vol. 25(3): 305-318, 1992.
25. S.-Y. Lee, M.-C. Yang and J.-W. Chen. "Signature File as a Spatial Filter for Iconic Image Database," *Journal of Visual Languages and Computing*, Vol. 3: 373-397, 1992.
26. N. Mamoulis and D. Papadias. "Integration of Spatial Join Algorithms for Processing Multiple Inputs," in *ACM Conference on the Management of Data*, Philadelphia, PA, 1999.
27. D. Papadias, D. Arkoumanis and N. Karacapilidis. "On the Retrieval of Similar Configurations," T. Poiker and N. Chrisman, editors, in *8th International Symposium on Spatial Data Handling, SDH '98*, Vancouver, International Geographical Union, 1998.
28. D. Papadias, N. Mamoulis and V. Delis. "Algorithms for Querying Spatial Structure," A. Gupta, O. Shmueli, and J. Widom, editors, in *24^{th} VLDB Conference*, New York, NY, 1998.
29. D. Papadias, N. Mamoulis and Y. Theodoridis. "Constraint-Based Processing of Multiway Spatial Joins," *Algorithmica*, Vol. Special Issue on Algorithms for GIS, 2001.

[Warning: Draw object ignored]

30. A. Papadopoulos, P. Rigaux and M. Scholl. "A Performance Evaluation of Spatial Join Processing Strategies," R.H. Güting, D. Papadias, and F. Lochovsky, editors, in *Advances in Spatial Databases, 6th International Symposium, SSD '99*, Hong Kong, China, Springer, 1999.

31. H.-H. Park, G.-H. Cha and C.-W. Chung. "Multi-Way Spatial Joins Using R-Trees: Methodology and Performance Evaluation," R.H. Güting, D. Papadias, and F. Lochovsky, editors, in *Advances in Spatial Databases, 6th International Symposium, SSD '99*, Hong Kong, China, Springer, 1999.

32. G. Petrakis and C. Faloustos. "Similarity Searching in Medical Image Databases," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 9(3), 1997.

33. D. Pfoser and C. Jensen. "Capturing the Uncertainty of Moving-Object Representations," R.H. Güting, D. Papadias, and F. Lochovsky, editors, in *Advances in Spatial Databases-6th International Symposium, SSD '99*, Hong Kong, China, Springer-Verlag, 1999.

34. D. Pullar and M.J. Egenhofer. "Toward Formal Definitions of Topological Relations Among Spatial Objects," in *Third International Symposium on Spatial Data Handling*, Sydney, Australia, 1988.

35. D. Randell, Z. Cui and A. Cohn. "A Spatial Logic Based on Regions and Connection," B. Nebel, C. Rich, and W. Swarthout, editors, in *3rd International Conference on Knowledge Representation and Reasoning*, St. Charles, IL, Morgan Kaufmann, 1992.

36. A. Rodríguez and F. Godoy. "A Content-Based Approach to Searching and Indexing Spatial Configurations," M.J. Egenhofer and D. Mark, editors, in GIScience 2002, Springer-Verlag, 2002.

37. R. Shariff. "Natural-Language Spatial Relations: Metric Refinaments of Topological Properties", Department of Spatial Information Science and Engineering, University of Maine, Orono, ME, Ph.D. Thesis, 1996.

38. J. Smith and S.F. Chang. "Integrated Spatial Query and Feature Image Query,"

[Warning: Draw object ignored]

Multimedia Systems, Vol. 7: 129-140, 1999.

39. J.R. Smith and S.-F. Chang. "Visually Searching the Web for Content," *IEEE Multimedia*, Vol. 4(3): 12-20, 1997.

40. Stanford Exploration Project. "More Equations than Unknowns," http://sepwww.stanford.edu/s access in 2003.

41. O. Wolfson, B. Xu, S. Chamberlain and L. Jiang. "Moving Objects Databases: Issues and Solutions," M. Rafanelli and M. Jarke, editors, in *Statistical and Scientific Database Management*, Capri, Italy, IEEE Computer Society Press, 1998.

42. A. Yoshitaka and T. Ichikawa. "A Survey on Content-Based Retrieval for Multimedia Databases," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 11(1): 81-93, 1999.

[Warning: Draw object ignored]

List of Figures

[Warning: Draw object ignored]

[Warning: Image not found]

Figure 5: Asymmetric property of topological relations.

[Warning: Draw object ignored]

[Warning: Image not found]

(a) (b)

Figure 6: Characterizing values of content measure F_a: (a) *meet* relation when $F_a(A,B) + F_a(B,A) = 1$ and (b) *overlap* relation when $F_a(A,B) + F_a(B,A) < 1$.

[Warning: Draw object ignored]

[Warning: Image not found]

Figure 7: Characterizing values of content measure F_d.

Figure 8: Possible values of F_m that are classified into eight topological relations.

[Warning: Image not found]

[Warning: Image not found]

Figure 9: Scaling of a pair of objects.

[Warning: Image not found] [Warning: Draw object ignored]

(a) (b)

Figure 10: Rotation dependence of content measures: (a) illustrative case and (b) variation graph.

[Warning: Image not found] [Warning: Draw object ignored]

(a) (b)

Figure 11: Content measures versus objects' movements: (a) movement (b) variation graph.

[Warning: Draw object ignored]

[Warning: Image not found] [Warning: Draw object ignored]

(a) (b)

Figure 12: Content measures versus objects' movements: (a) movement (b) variation graph.

[Warning: Draw object ignored]

[Warning: Image not found]

(a) (b)

Figure 13: Control points: (a) original configuration and (b) target configuration.

[Warning: Draw object ignored]

[Warning: Image not found]

(a)

[Warning: Image not found]

(b)

Figure 14: Combinations of four vertices under: (a) rotations of [Warning: Missing symbol F070]/2 radians and (b) rotations of [Warning: Missing symbol F070]/2 radians flipped over the y-axis.

Figure 15: Geometric adjustment with 4 points per MBR.

[Warning: Image not found]

[Warning: Draw object ignored]

[Warning: Image not found]

(a)

(b)

[Warning: Image not found]

Figure 16: Combinations of eight vertices under: (a) rotations of [Warning: Missing symbol F070]/2 radians and (b) rotations of [Warning: Missing symbol F070]/2 radians flipped over y-axis.

[Warning: Draw object ignored]

[Warning: Image not found]

(a)

[Warning: Image not found]

(b)

Figure 17: Effect of using (a) four- or (b) eight-points for MBR representation on configuration adjustment.

[Warning: Draw object ignored]

[Warning: Image not found]

Figure 18: Strategy for comparing content measures: content measures versus geometric adjustment.

[Warning: Draw object ignored]

[Warning: Image not found]

Figure 19: Nine possible MBRs of a 2x2 Cell-Box.

[Warning: Draw object ignored]

[Warning: Image not found]

Figure 20: The eight different configurations in a box of 2x2 cells.

[Warning: Draw object ignored]

Figure 21: Distribution of topological relations in sets of configurations.

[Warning: Draw object ignored]

[Warning: Draw object ignored]

Figure 22: Correlations of content measure F_a when relations are eliminated from the data set.

[Warning: Draw object ignored]

[Warning: Draw object ignored]

Figure 23: Correlations of content measure F_d as relations are eliminated from the data set.

[Warning: Draw object ignored]

[Warning: Draw object ignored]

Figure 24: Correlations of content measure F_m as relations are eliminated from data set.

Topological Relation	**Value Range for ($F_m(A,B)$, $F_m(B,A)$)**
disjoint	($F_m(A,B) > 1$, $F_m(B,A) > 1$)
meet	($F_m(A,B) = 1$, $F_m(B,A) = 1$)
overlap	($\mid F_m(A,B) \mid < 1$, $\mid F_m(B,A) \mid < 1$)
equal	($F_m(A,B) =$ -1 , $F_m(B,A) =$ -1)
covers & covered_ by	($\mid F_m(A,B) \mid < 1$, $F_m(B,A) =$ -1) or ($F_m(A,B) =$ -1 , $\mid F_m(B,A) \mid < 1$)
inside & contains	($\mid F_m(A,B) \mid < 1$, $F_m(B,A) <$ -1) or ($F_m(A,B) <$ -1 , $\mid F_m(B,A) \mid < 1$)

Table 1: Possible values of F_m according to topological relations.

Transitions	$F_a(A,B)$, $F_a(B,A)$	$F_d(A,B)$, $F_d(B,A)$	$F_m(A,B)$, $F_m(B,A)$
[Warning: Image not found]	(0.33,0.11)	(0.54,0.33)	(1.28 , 1.45)
[Warning: Draw object ignored] [Warning: Image not found]	(0.40,0.13)	(0.62,0.38)	(1.00,1.00)

[Warning: Draw object ignored]

[Warning: Draw object ignored] [Warning: Image not found]	(0.60,0.20)	(0.67,0.42)	(1.00,1.00)
[Warning: Draw object ignored] [Warning: Image not found]	(0.75,0.25)	(0.81,0.50)	(0.67,0.00)
[Warning: Draw object ignored] [Warning: Image not found]	(1.00,0.33)	(1.00,0.62)	(0.33,-1.00)
[Warning: Draw object ignored] [Warning: Image not found]	(1.00,0.33)	(1.00,0.62)	(0.19,-1.22)

Table 2: Values of content measure for different state transitions of two MBRs.

Cell Box	**# MBRs**	**# Different Configurations**	**# Comparisons**
2x2	9	8	56
3x3	36	78	6.006
4x4	100	359	128.522

[Warning: Draw object ignored]

5x5	225	1290	1.662.810
6x6	441	3550	12.598.950
7x7	784	8773	76.956.756

Table 3: Number of Comparisons for different cell boxes.

	Four Points			Eight Points		
Cell Box	$\boldsymbol{F}_a$	$\mathbf{F}_d$	$\mathbf{F}_m$	$\mathbf{F}_a$	$\mathbf{F}_d$	$\mathbf{F}_m$
2x2	0.43	0.48	0.58	0.48	0.55	0.65
3x3	0.46	0.55	0.62	0.46	0.56	0.64
4x4	0.45	0.55	0.65	0.46	0.56	0.67
5x5	0.44	0.54	0.64	0.45	0.55	0.67
6x6	0.43	0.53	0.64	0.45	0.54	0.67
Mean	**0.44**	**0.53**	**0.63**	**0.46**	**0.55**	**0.66**

Table 4: Correlations between content measures and similarity function when using four points or eight points in the representation of MBRs.

[Warning: Draw object ignored]