



# Article Hierarchical Model for the Similarity Measurement of a Complex Holed-Region Entity Scene

Zhanlong Chen 💩, Rongrong Zhu, Zhong Xie and Liang Wu \* 💩

Department of Information Engineering, China University of Geosciences, Wuhan 430074, China; chenzl@cug.edu.cn (Z.C.); zhurongrong2017@gmail.com (R.Z.); xiezhong@cug.edu.cn (Z.X.)

\* Correspondence: wuliang@cug.edu.cn; Tel.: +86-133-1719-3026

Received: 13 October 2017; Accepted: 27 November 2017; Published: 29 November 2017

Abstract: Complex multi-holed-region entity scenes (i.e., sets of random region with holes) are common in spatial database systems, spatial query languages, and the Geographic Information System (GIS). A multi-holed-region (region with an arbitrary number of holes) is an abstraction of the real world that primarily represents geographic objects that have more than one interior boundary, such as areas that contain several lakes or lakes that contain islands. When the similarity of the two complex holed-region entity scenes is measured, the number of regions in the scenes and the number of holes in the regions are usually different between the two scenes, which complicates the matching relationships of holed-regions and holes. The aim of this research is to develop several holed-region similarity metrics and propose a hierarchical model to measure comprehensively the similarity between two complex holed-region entity scenes. The procedure first divides a complex entity scene into three layers: a complex scene, a micro-spatial-scene, and a simple entity (hole). The relationships between the adjacent layers are considered to be sets of relationships, and each level of similarity measurements is nested with the adjacent one. Next, entity matching is performed from top to bottom, while the similarity results are calculated from local to global. In addition, we utilize position graphs to describe the distribution of the holed-regions and subsequently describe the directions between the holes using a feature matrix. A case study that uses the Great Lakes in North America in 1986 and 2015 as experimental data illustrates the entire similarity measurement process between two complex holed-region entity scenes. The experimental results show that the hierarchical model accounts for the relationships of the different layers in the entire complex holed-region entity scene. The model can effectively calculate the similarity of complex holed-region entity scenes, even if the two scenes comprise different regions and have different holes in each region.

**Keywords:** holed-region entity; complex holed-region entity scene; hierarchical model; position graph; multi-level chord-length function

# 1. Introduction

Although the collection and maintenance of spatial data is the most time-consuming activity in GIS engineering in the real world, data for the same objects are often repeatedly collected by different departments [1]. Several differences are present in the same objects on maps produced from different sources that are due to mapping errors, different applications, and interpretations [2–4]. These identical entities are usually inconsistent in positioning accuracy, shape, and attribute information. To reduce the cost of GIS data collected by the GIS application departments and to evaluate the quality of the existing map data, it is necessary to use standard map data as a reference to measure the similarity of two map data. Similarity and matching are not only widely used to update data from different sources [5] but also in data retrieval [6] and classification [7]. These topics have attracted a great deal of research attention over the past decades [8,9].

In a real-life situation, matching and similarity are used in spatial applications [10,11], not only to match simple objects described as points [12], lines [13], and polygons [14,15] but also to match complex geometric structures [16,17]. Many geographic phenomena have discontinuities in the form of internal cavities [18]. A real-world geographic entity containing inner boundaries, such as a region containing several lakes or an island lake, is abstracted as a holed-region. In the simple feature specification for the Structured Query Language (SQL) of Open Geospatial Consortium (OGC), the existence of a holed-region is implied when defining a region as a simple feature in a geo-database, "a region is a planar surface, defined by one exterior boundary and zero or more interior boundaries; interior boundary defines a hole in the region, …, the exterior of a region with one or more holes is not connected. Each hole defines a connected component of the exterior" [19]. Meanwhile, the scene containing several island lakes is abstracted as a complex holed-region entity scene can be regarded as a collection of a random number of regions that contain holes.

Numerous methods that measure the similarity of two holed-regions exist, and most of them focus on topological relationships, such as the topological relationships between two holed-regions [20], a hierarchical model based on the CBM to treat complex regions in topological queries [21], the 23 topological relationships between a simple region and a region with a hole [22], the 152 topological relationships between two single holed-regions [23], and the topological similarity between a hole-free region and a multi-holed-region [18]. Compared to a simple holed-region, a complex holed-region entity scene usually has a more complicated distribution pattern. Such characteristics make a similar evaluation between two complex holed-region entity scenes extremely difficult. To the best of our knowledge, similarity measurements between two complex holed-region similarity measurement methods and models cannot be directly used for complex holed-region entity scenes because they lack inherited relationships between different layers.

In this paper, we present a hierarchical model to measure the similarity of two complex holed-region entity scenes. In our model, a complex holed-region entity scene is decomposed into three layers—a complex scene, a micro-spatial-scene (i.e., a spatial scene comprising a collection of holes and their spatial relationships), and a simple entity (i.e., a hole).

The relationships between the adjacent layers are considered as sets of relationships. Each level of similarity measurements is nested with the adjacent one. Entity matching is performed from top to bottom, while the similarity results are calculated from local to global. The hole distribution is considered as a constraint satisfaction problem (CSP) [24] that comprises the shape descriptors and the directional relationships of the holes. In addition, position graphs [25] describe the distribution of the micro-spatial-scene, and a force diagram [26,27] is used to measure the similarity between the graphs. The extended Hausdorff distance [28] represents the relationships between the holed-regions and the scene center. Because of topological and semantic limitations of the entity matching process [29], we introduced the Hungarian algorithm [6] to resolve matching problems between holed-regions belonging to different complex entity scenes. The multilevel chord length [30] is adopted to construct complex functions to describe the geometric shape from entity to part, and the shape descriptor is based on the fast Fourier transform of the complex function [31]. This paper is an extension of our precedent work [25]. The main idea of this model is not characterized merely by the iterative approach of our previous work; the whole spatial distribution information of the holed-regions in the complex holed-region entity scenes is represented in our model, which is ignored in many previous studies.

The rest of this paper is organized as follows. Section 2 introduces the hierarchical model for a holed-region entity scene and proposes the similarity measurement of the position graphs. We next introduce the matching of holed regions and holes and the similarity measurement of the micro-spatial-scene in Section 3. In Section 4, we propose the similarity measurement between complex holed-region entity scenes. Next, we apply the model to the Great Lakes by using data from different

years and then discuss and analyze the results in Section 5. Finally, the study's conclusions are presented in Section 6.

#### 2. Similarity Measurements of Two Complex Scenes

#### 2.1. Hierarchical Method for a Complex Holed-Region Entity Scene

The Open Geospatial Consortium (OGC) Abstract Specification describes complex geometry as "The complex space object is mainly composed of geometric combinations, such as merging and difference generation for basic spatial objects or their topological parts." A complex holed-region scene *S* is a set of n ( $n \ge 1$ ) multi-holed-regions  $R_i$ , i.e.,  $S = \{R_1 \cup R_2 \cup ... \cup R_n\}$ . Therefore, the similarity of a complex scene can be regarded as an overall generalization of the similarity of  $R_i$  and the distribution similarity of  $R_i$ .

The multi-holed-region R, which can be thought of as a micro-scene, is m ( $m \ge 0$ ) holes  $H_j$  merged, i.e.,  $R = \{H_1 \cup H_2 \cup \ldots \cup H_n\}$ . Likewise, the similarity of the micro-spatial-scene is considered to be a generalization of the geometric similarity of the hole  $H_j$  and the similarity of the distribution of  $H_j$  in R.

To simplify the overall similarity measurement of the complex scene, we introduce the concept of a hierarchy. The hole in the holed-region is the first level, designated as the "simple entity" layer; the holed-region entity is the second level, designated as the "micro-spatial-scene" layer; and the complex holed-region entity scene is the third level, designated as the "complex scene" layer. A hierarchical calculation is done for the similarity of the complex holed-region entity scene. The layered method is shown in Figure 1.



Figure 1. Hierarchical decomposition of a complex scene.

In the first step, the complex scene *S* is decomposed into several regions  $R_i$  (I = 1, 2, ..., n), which contain the exterior boundary and the corresponding micro-spatial-scene  $R'_i$  (I = 1, 2, ..., n), and the similarity of a complex scene layer is the aggregation of the similarity of all of the micro-spatial-scenes, all exterior shapes, and the matching and distribution of all regions.

In the second step, we focus on the similarity of the micro-spatial-scene. The hole in the micro-spatial-scene is regarded as a simple surface entity  $H_j$  (j = 1, 2, ..., m), and the simple entity  $H_j$  has a fixed position distribution in  $R_j$ . The micro-spatial-scene is a combination of simple entities;

therefore, the similarity of the micro-scene is a generalization of the shape similarity, the direction similarity, and the completeness of the matching of simple entities.

In the third step, the matching situation is not considered. The similarity of the simple entity layer focuses on the shape similarities and direction similarities of holes.

#### 2.2. Location Distribution and Similarity Measurement

## 2.2.1. Position Graphs

The hierarchical method of Section 2.1 implies that the similarity of a complex scene is closely related to the similarity of the distribution of the holed-regions. In this paper, we refer to the method of position graphs to describe the distribution of the holed entity in a complex scene. The boundary of a complex scene *S* is determined by the MBR method and the centroid *C* of scene *S* found by the diagonal of the rectangle. In the Position Graphs, the distance between the centroid of a holed entity and the scene center is defined as the relative distance between the holed entity and the scene. As shown in Figure 2, *C1*, *C2*, *C3* are the centroids of the holed entities *R1*, *R2*, *R3*, respectively. The point *O* is the scene center, and the points *NT1*, *NT2*, *NT3* are, respectively, the nearest tangent points from *O* to the outlines of *R1*, *R2*, *R3*. The center point position graph (CPPG) is formed by connecting the center point holed entities. The nearest tangent point position (NTPPG) is formed by connecting the nearest tangent points from scene centroid to the outlines of the holed-region.



**Figure 2.** (a) Selection of the center point and nearest tangent points; (b) center point position graph (CPPG); (c) nearest tangent point position (NTPPG).

As shown in Figure 3,  $N_1$ ,  $N_2$ ,  $N_3$  are, respectively, the nearest points from the centroid O to the outlines of  $R_1$ ,  $R_2$ ,  $R_3$ , and  $F_1$ ,  $F_2$ ,  $F_3$ , are, respectively, the farthest points from the centroid O to the exterior boundaries of  $R_1$ ,  $R_2$ ,  $R_3$ , so the nearest points position graph(NPPG) and farthest points position graph (FPPG) is formed by connecting O to  $N_i$  and  $F_i$  respectively.



**Figure 3.** (**a**) Selection of the nearest points and farthest points; (**b**) nearest points position graph (NPPG); (**c**) farthest points position graph (FPPG).

#### 2.2.2. Descriptive Measurement of the Position Graphs

Each edge of a shape is formed by two forces that have equal value but opposite direction [25]. As a result, a region can be described by a set of forces. A group of forces that form a region in a coordinate system is drawn and these forces represent a force diagram. According to the parallelogram law [32], the force can be decomposed along the x-axis and y-axis, and the sum in x-axis and y-axis always meet:

$$\sum_{i=1}^{N} f_{ix} = 0, \quad \sum_{i=1}^{N} f_{iy} = 0$$
<sup>(1)</sup>

The parameter  $f_{ix}$  is the decomposition part of force  $f_i$  on the x-axis,  $f_{iy}$  is the decomposition part of force  $f_i$  on the y-axis, and N is the force number. The positive or negative value of these forces will change periodically with the direction of the force, with a period of 180°.

Adding the absolute value of the forces in the positive and negative directions of the x-axis provides:

$$F_x(\alpha) = \sum_{i=1}^N |f_{ix}(\alpha)|$$
(2)

Similarly,  $F_{\nu}(\alpha)$  is available:

$$F_{y}(\alpha) = \sum_{i=1}^{N} \left| f_{iy}(\alpha) \right|$$
(3)

The projection interval function of the force diagram is defined as

$$F(\alpha) = F_x(\alpha) / F_y(\alpha) \tag{4}$$

where  $\alpha$  is the rotation angle of the force diagram, and the *F*( $\alpha$ ) is also a periodic function with a period of 180° and its function image is a waveform.

#### 2.2.3. Similarity Measurement of the Position Graph

Different position graphs can be described with different force diagrams, and different position graphs have different projection interval functions; therefore, the similarity between the distributions of the holed-regions can be measured by the degree of matching of the projection intervals. Because of the characteristics of the waveform, the degree of matching between the two projections ratios is calculated by the minimum mean error (MME) of the waveform.

When the sample points are distributed in a 180° range and the rotation angle  $\alpha$  is replaced by the discrete value *i*, the MME between the position graphs *P*<sub>1</sub> and *P*<sub>2</sub> is calculated as:

$$M_{P_1P_2}(l) = \frac{1}{N} \sum_{i=0}^{N-1} |F_1(i) - F_2(i-l)|$$
(5)

where  $F_1(i)$  and  $F_2(i - l)$  ( $0 \le l \le N - 1$ ) are the projection ratio functions of the position graphs  $P_1$  and  $P_2$ , l is the deviation of the two functions, and N is the number of discrete sampling points. The MME is more accurate when N is larger. Since  $F \in [0, 1]$ , the similarity of the two position graphs can be defined as:

$$S'_{p_g} = 1 - \frac{\min\{M_{P_1P_2}(l)\}}{\max\{F_1(i), F_2(i)\}}, \quad 0 \le l, \ i \le N - 1$$
(6)

#### 2.2.4. Geometric Transformation of the Position Graph

Climate change, ocean currents, crustal movements, natural disasters, artificial construction, and other factors may drive diverse phenomena for lakes that contain islands (i.e., holed-regions in a complex spatial scene), such as revolution, rotation, moving, and scaling. Obviously, it is necessary to measure the similarity of revolution in which the CPPG rotate around the centroid of the complex

scene, the rotation in which the CPPG rotates around its center point, and moving and scaling based on position graphs.

**Revolution Similarity** 

As shown in Figure 4, *O* is the centroid of a complex scene, and  $O_1$  is the center of the CPPG. The vector from the point *O* to the nearest vertex of the MBR of the complex scene is  $u_2$ , and  $u_1$  is the vector from the point *O* to the point  $O_1$ . The angle between  $u_1$  and  $u_2$  is  $\beta$ , which is defined as the central angle.  $\beta$  is negative when  $u_1$  is on the left side of  $u_2$ . The change value  $\theta$  of  $\beta$  is the revolution angle.



Figure 4. Revolution of CPPG.

The revolution similarity of two CPPGs can be defined as:

$$S'_{r1} = 1 - \frac{|\beta_1 - \beta_2|}{360^{\circ}} \tag{7}$$

**Rotation Similarity** 

Draw the CPPG and the complex scene of the MBRs and move the MBR of the scene to vertex of the MBR of CPPG, where  $v_1$  is the direction vector of the scene and  $v_2$  is the direction vector of CPPG. The angle  $\alpha$  between  $v_1$  and  $v_2$  is defined as the relative direction angle, and is negative when  $v_1$  is on the left side of  $v_2$ . (Figure 5).



Figure 5. Rotation of CPPG.

The rotation similarity of the two CPPGs with rotation can be defined as:

$$S'_{r2} = 1 - \frac{|\partial_1 - \partial_2|}{360^\circ}$$
 (8)

#### Moving Similarity

The average radius of a complex scene is *R*, and the relative center distance (RCD) *l* is defined as:

$$l = \frac{\sqrt{[x(o) - x(o_1)]^2 + [y(o) - y(o_1)]^2}}{R}$$
(9)

The center coordinate of the complex scene is (X(O), Y(O)), and at the same time  $(X(O_1), Y(O_1))$  is the center coordinate of the CPPG. The moving translation similarity [33] can be defined as:

$$S'_{m} = 1 - \frac{|l_{1} - l_{2}|}{\max(l_{1}, l_{2})}$$
(10)

Scaling Similarity

The Hausdorff distance [28] is introduced to describe the zoom of a complex scene with a holed-region. The directional Hausdorff distance formula is:

$$h^{f_1}(A, B) = \min\left\{\varepsilon_i : f_1 = \frac{\vartheta((B \oplus S(\varepsilon_i)) \cap A)}{\vartheta(A)}\right\}$$
$$h^{f_2}(B, A) = \min\left\{\varepsilon_j : f_2 = \frac{\vartheta((A \oplus S(\varepsilon_j)) \cap B)}{\vartheta(B)}\right\}$$
(11)

The extended Hausdorff distance formula is:

$$H^{f_1f_2}(A,B) = \max\left\{\min\left\{\varepsilon_i : f_1 = \frac{\vartheta((B \oplus S(\varepsilon_i)) \cap A)}{\vartheta(A)}\right\}, \min\left\{\varepsilon_j : f_2 = \frac{\vartheta((A \oplus S(\varepsilon_j)) \cap B)}{\vartheta(B)}\right\}\right\}$$
(12)

In Formula (12),  $\vartheta(\bullet)$  is the area or line length measure function of region. When *A* or *B* is a point, the value of  $\vartheta(\bullet)$  is 0 ("•" is the empty set).  $\oplus$  is the mathematical morphological expansion operator that represents the Minkowski value, and  $S(\varepsilon)$  is a disc with a radius of  $\varepsilon$  [34]. The position graph scaling can be derived from the ratio of the extended Hausdorff distance from the center of a scene to its contour to the average radius of the scene. This ratio is defined as the relative expansion of the Hausdorff distance:

$$Dis = \frac{1}{NR} \sum_{i=0}^{N} H_i \tag{13}$$

where *H* is the extended Hausdorff distance of the center of a scene to a holed-region, *N* is number of a holed-region, and *R* is the average radius of a region. The similarity of the CPPG of the holed-region can be calculated by the following formula:

$$S'_{s} = 1 - \frac{\left| Dis_{p} - Dis_{q} \right|}{\max(Dis_{p}, Dis_{q})}$$
(14)

## 3. Similarity Measurement of Micro-Spatial-Scene

#### 3.1. Pairwise Matching of a Holed-Region Entity

It is necessary to match pairs of holed-region entities to measure the similarity of a holed-region in different complex scenes. Figures 6 and 7, respectively, show three and four holed-regions in the complex scenes  $S_1$  and  $S_2$ , which may produce different matching results: ( $R_1$ ,  $R_1'$ ), (R1, R2'), (R1, R3'), (R1, R4'), (R2, R1'), (R2, R2'), (R2, R3'), (R2, R4'), (R3, R1'), (R3, R3'), (R3, R4'). Different matching results will yield different similarity measurements.



**Figure 6.** Complex holed-region entity scene  $S_1$ .



**Figure 7.** Complex holed-region entity scene *S*<sub>2</sub>.

The scenes  $S_1$  and  $S_2$  are, respectively, the reference scene and the matching scene. Any holed-region  $R_i$  in  $S_1$  will match with a corresponding set of holed-regions  $\{R_i\}$  in  $S_2$ . To obtain the matching set of the holed-region  $R_i$  in  $S_1$ , we first determine the vertex P', which is the closest point on the MBR of the CPPG to the center of the CPPG; then, we obtain the vector  $\overrightarrow{P'd_i}$  by connecting  $d_i$  and P', whose length is  $d_{1\_i}$ , as shown in Figure 8a. The angle between the vector  $\overrightarrow{P'd_i}$  and the direction vector of the MBR is  $\delta$ . Next, a point Q is determined in  $S_2$ . A circular buffer with the center of Q is drawn and the average radius of the holed-region  $R_i$  is  $r_i$ ; furthermore, the regions in the buffer or that intersect with the buffer are regarded as the matching region candidates of  $R_i$ . The angle between the vector  $\overrightarrow{PQ}$  and the vector of the MBR direction vector is  $\delta$ , as shown in Figure 8b. The length of  $\overrightarrow{PQ}$  is defined as  $d_{2\_i}$ :

$$d_{2_i} = d_{1_i} \times \frac{R_2}{R_1} \tag{15}$$

where  $R_2$  is the average radius of scene  $S_2$  and  $R_1$  is the average radius of scene  $S_1$ .



Figure 8. Buffer analysis determines the matching set.

After obtaining the matching candidate set  $\{R_i\}$  of the reference entity  $R_i$ , our next step is to determine the matching entity with the highest degree of matching in the matching candidate set. For this case, the Hungarian algorithm [35] is introduced. It is supposed that the similarity between  $R_i$  and the regions belonging to  $S_2$  but not included in  $\{R_i\}$  is 0, and the distance is infinite. A matrix *D* of *m* rows and *n* columns is constructed, where *m* is the number of regions in scene  $S_1$ , *n* is the number of regions in  $S_2$ , and the element  $d_{ij}$  in *D* is defined as  $d_{ij} = \text{dis}(R_i, R'_j)$ , which expresses the Euclidean distance between centroid  $R_i$  and  $R'_j$ .

$$dis(R_i, R_j') = \sqrt{\left(x(R_i) - x(R_j')^2\right) + \left(y(R_i) - y(R_j')^2\right)}$$
(16)

where  $(x(R_i), y(R_i))$  and  $(x(R'_i), y(R'_j))$  are the centroids of  $R_i$  and  $R'_j$ , respectively.

The decision variable  $x_{i,j}$  ranges from

$$x_{x,j} = \begin{cases} 0, & R_i \text{ in } S_1 \text{ not matching with } R_j' \text{ in } S_2 \\ 1, & R_i \text{ in } S_1 \text{ match with } R_j' \text{ in } S_2 \end{cases}$$
(17)

The decision variable matrix X is

$$D = \begin{bmatrix} d_{11} & d_{12} & d_{1...} & d_{1n} \\ d_{21} & d_{22} & d_{2...} & d_{2n} \\ d_{...1} & d_{...2} & d_{...} & d_{...n} \\ d_{m1} & d_{m2} & d_{m...} & d_{mn} \end{bmatrix} \quad X = \begin{bmatrix} x_{11} & x_{12} & x_{1...} & x_{1n} \\ x_{21} & x_{22} & x_{2...} & x_{2n} \\ x_{...1} & x_{...2} & x_{...} & x_{...n} \\ x_{m1} & x_{m2} & x_{m...} & x_{mn} \end{bmatrix}$$
(18)

So the matching results can be calculated by

$$\min Z = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} x_{ij} \quad s.t. \begin{cases} \sum_{\substack{i=1 \ j=1}^{m}}^{m} x_{ij} = 1, & j = 1, 2, \dots, n \\ \sum_{\substack{j=1 \ x_{ij} = 1 \text{ or } 0, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, n \\ x_{ij} = 1 \text{ or } 0, \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n \end{cases}$$
(19)

In this process, the value of  $x_{ij}$  that minimizes the sum of each column of Z for each row is vital, and this calculation can be derived from the Hungarian algorithm. The decision variable matrix X obtained can determine the result between the regions  $R_i$  in  $S_1$  and  $R'_j$  in  $S_2$ . Furthermore, the matching similarity between  $S_1$  and  $S_2$  can be obtained by:

$$S'_{R\_S} = \frac{1}{n} \sum_{i=0}^{n} (1 - d_i)$$
<sup>(20)</sup>

where *n* is the number of matching regions, and  $d_i$  is the Euclidean distance mentioned above.

#### 3.2. Pairwise Matching of Holes

In this part, we consider a holed-region entity as a micro-spatial-scene, and the "hole" in the micro-spatial-scene is an independent simple surface entity. Therefore, in the measurement of the similarity of a micro-spatial-scene, we not only must consider the scene of the internal and external boundaries, but we also must consider the direction similarity of the holes. From a human perception viewpoint, a scene comparison is a process of associating similar objects across a scene, and the relationships between objects must correspond as well [36]. Therefore, the first task in the common models of spatial scene similarity measurement is the identification of corresponding objects [37,38]. As in the similarity measurement process of the micro-spatial-scene, it is most important to find the matching relationship of the hole. This matching problem is treated as a constraint satisfaction problem

(CSP), in which the contour and direction relationships of holes in reference to the multi-holed-region act as constraints.

The CSP has the following conditions: (1) the variables  $eb_n$  and  $eb_m$  represent the edges of the region  $Rh_n$  and  $Rh_m$ ; (2) the variable set  $H_1, \ldots, H_n$  represents the holes in  $Rh_n$ ; (3) for each  $H_i$ , sets of possible matching variables  $H'_1 = \{H'_1, \ldots, H'_l\}$  corresponding to holes in  $Rh_m$  exist; (4) each variable  $H_i$  has a unit variable  $S_i$  that describes its shape; and (5) for each variable set  $(H_i, H'_j)$ , a binary variable  $d_{ii}$  exists that describes the relationship between the variables.

In this study, the matching association graph is utilized to describe the constraints in  $Rh_n$ , in which the value of each node represents the shape of the contour constraints, and the values attaching to edges indicate the directional constraint between the holes, as shown in Figure 9.



**Figure 9.** Expression graph of the constraints of *Rh<sub>n</sub>*.

Since the number of holes in  $Rh_n$  and  $Rh_m$  can be different, the hole-matching in this study does not compare the entire graph, just the subgraphs. There are three matching solutions as follows (Figure 10):

- Complete solution. A subgraph  $G'_m$  of the graph  $G_m$  of  $Rh_m$  matched with the graph  $G_n$  of  $Rh_n$ , in which all holes and directions of  $Rh_n$  have a correspondence in  $Rh_m$ .
- Incomplete solution. A subgraph  $G'_m$  of  $G_m$  matches the subgraph  $G'_n$  of  $G_n$ , in which a subset of holes and directions of  $Rh_n$  have a correspondence in  $Rh_m$ .
- No solution.



Figure 10. (a) Complete solution; (b) incomplete solution; (c) no solution.

The formation of the association graph for the holed-region entities  $Rh_n$  and  $Rh_m$ , for which  $Rh_n$  has the graph *G* with the node set  $(g_1, g_2, ..., g_n)$  and  $Rh_m$  has the graph  $G'(g'_1, g'_2, ..., g'_n)$ , contains

two important steps. First, a node  $a_{ij}(g_i, g'_j)$  is created if  $g'_j$  is in G' that satisfies the univariates (shape matching) of  $g_i$  in G. Second, the nodes are connected with their edges by inserting an edge to connect the node  $a_{ij}(g_i, g'_j)$  with  $a_{lk}(g_l, g'_k)$ . The value attached to the node is the geometric similarity between the two holes, and the value of the edge is the similarity of the direction relationship. As a result, the associated graph with weight is formed.

## 3.3. Similarity Measurement of Holes

The hierarchical method of Section 2.1 shows that the geometric similarity of the inner and outer contours of a holed-region entity belongs to different similarity measurement levels, and the geometric similarity of the inner contours belongs to the measuring dimension of the micro-spatial-scene level, while the geometric similarity of the outer contours is the measuring dimension of a complex scene level. Therefore, we do not account for the outer contour of the holed-region entities when we compute the similarity of Holes.

(1) Assuming that the holed-region entities  $Rh_n$  and  $Rh_m$  have t pairs of matching holes,  $S_{Hi}$  is the shape similarity of the  $i^{\text{th}}$  pair of holes and  $W_{Hi}$  is the weight of the  $i^{\text{th}}$  pair of holes. Next, the shape similarity between  $Rh_n$  and  $Rh_m$  is defined by  $S'_{shape}$  for

$$S'_{inshape} = \frac{\sum_{i=1}^{t} W_{Hi} \bullet S_{Hi}}{\sum_{i=1}^{t} W_{Hi}}$$
(21)

(2) Supposing that the holed-region entities  $Rh_n$  and  $Rh_m$  have t pairs of matching holes with t(t-1)/2 pairs of matching directions,  $S_{Di}$  is the similarity of the i<sup>th</sup> pair of direction relationships, which have a weight  $W_{Di}$ . Thus, the direction relationship similarity  $S'_{direction}$  between  $Rh_n$  and  $Rh_m$  can be defined as

$$S'_{direction} = \frac{\sum_{i=1}^{t(t-1)/2} W_{Di} \bullet S_{Di}}{\sum_{i=1}^{t(t-1)/2} W_{Di}}$$
(22)

(3) The weight of the shape similarity and the direction relationship similarity is set to  $W_{shp}$  and  $W_{dir}$ , respectively; then, the similarity  $S'_{Rhn,Rhm}$  between micro-spatial-scene  $Rh_n$  and  $Rh_m$  can be calculated:

$$S'_{Rhn,Rhm} = \frac{(W_{shp} \times S'_{inshape}) + (W_{dir} \times S'_{direction})}{W_{shp} + W_{dir}}$$
(23)

(4) Since  $Rh_n$  and  $Rh_m$  may not be able to match all holes successfully, the unmatched holes should be considered as penalties in the similarity calculation of  $Rh_n$  and  $Rh_m$ . This situation is termed matching completeness. Assuming that  $Rh_n$  and  $Rh_m$ , respectively, have n and m holes, t pairs of matching holes, and the unmatched hole weights of  $Rh_n$  and  $Rh_m$  are  $\alpha$  and  $\beta$ , the matching completeness can be defined as:

$$S'_{Comp(Rhn,Rhm)} = \frac{t}{t + \alpha \times (n-t) + \beta \times (m-t)}$$
(24)

Three adjustment schemes for the weights  $\alpha$  and  $\beta$  in the completeness computation for spatial scene similarity [39] can serve different retrieval purposes:

• If  $\alpha = \beta = 1$ , two scenes in the matching hole are equally important (the desired outcome of this experiment).

- If  $\alpha = \beta = 0$ , the similarity measurement process ignores unmatched holes.
- If  $\alpha = 1$ ,  $\beta = 0$ , only the unmatched holes in the reference micro-spatial-scene have an impact on the matching completeness.
- (5)  $W_{comp}$  is set as the weight value of the matching completeness, and the similarity of the micro-spatial-scene  $Rh_n$  and  $Rh_m$  can be calculated by:

$$S_{MSCi} = S_{Rhn,Rhm} = S'_{Rhn,Rhm} \times (W_{Comp} \times (S'_{Comp} - 1) + 1)$$
(25)

# 4. Similarity of a Complex Holed-Region Entity Scene

The above discussion indicates that the similarity between two complex scenes is the similarity measurement of the complex layer. The Pseudocode for the Measurement of Complex Holed-region Entity Scene's Similarity is shown in Table 1.

1	Input: The two complex holed-regions entity scenes Scene <i>A</i> and Scene <i>B</i> Output: The similarity of the two scenes double ComputeSimofScene (Scene <i>A</i> , Scene <i>B</i> )
2	BuildPositionGraph (A); //CPPG_A, NTPPG_A, NPPG_A, FPPG_A
3	BuildPositionGraph ( <i>B</i> ); //CPPG_B, NTPPG_B, NPPG_B, FPPG_B
4	//the similarity of the two position graphs can be computed by Formula (6)
5	SimofDistribution (CPPG_A, CPPG_B); //S'_pg1 Similarity of CPPG
6	SimofDistribution (NTPPG_A, NTPPG_B); //S'_pg2 Similarity of NTPPG
7	SimofDistribution (NPPG_A, NPPG_B); //S'_pg3 Similarity of NPPG
8	SimofDistribution (FPPG_A, FPPG_B); //S'_pg4 Similarity of FPPG
9	SimofDis (S'_pg1, S'_pg2, S'_pg3, S'_pg4); //(S'_pg1+S'_pg2,+S'_pg3,+S'_pg4)/4
10	//the revolution, rotation, moving and scaling similarities. Formulas (7), (8), (10) and (14)
11	SimofGeometricTrans (CPPG_A, CPPG_B)
12	PairwiseofMicroScene (A, B)
13	SimofShape (Ri, R'i); //by the method of multilevel length chord description
14	ComputeSimofMicroScene ( $Ri, R'_i$ )
15	return SimofScene
16	end procedure
17	double ComputeSimofMicroScene (Ri, R'i)
18	SimofShape ( $Hi$ , $H'_i$ );
19	SimofDirction ( $Hi$ , $Hj$ , $H'_i$ , $H'_j$ )
20	PairwiseofHole $(H_i, H'_i)$
21	return SimofMicroScene
22	end procedure

Table 1. Pseudocode for the measurement of	complex	holed-region	entity sce	ne's similarity.
	1	0	2	

(1) Suppose that the complex scenes  $S_1$  and  $S_2$  have *t* pairs of matching micro-spatial-scenes, such that  $S_{MSCi}$  is the similarity of the *t*<sup>th</sup> pair of matching micro-spatial-scenes, and  $W_{Ri}$  is the weight of the corresponding holed-region entity. Thus, the similarity of *S* and *S'* can be defined as

$$S'_{MSC} = \frac{\sum_{i=1}^{t} W_{Ri} \bullet S_{MSCi}}{\sum_{i=1}^{t} W_{Ri}}$$
(26)

(2) Suppose that the complex scenes  $S_1$  and  $S_2$  have *t* pairs of matching micro-spatial-scenes, such that  $S_{Ri}$  is the shape similarity of the outer contour of the *i*<sup>th</sup> matching holed-region entities, and  $W_{Ri}$  is the weight of the corresponding holed entities. Thus, the outline similarity  $S'_{exshape}$  between  $S_1$  and  $S_2$  can be defined as

$$S'_{exshape} = \frac{\sum_{i=1}^{t} W_{Ri} \bullet S_{Ri}}{\sum_{i=1}^{t} W_{Ri}}$$
(27)

(3) Assume that the weight of the position graph similarity is  $W_{p_g}$ , the similarity weights of the rotation, rotation, translation, and scaling of the position graph are respectively  $W_{r1}$ ,  $W_{r2}$ ,  $W_m$ ,  $W_s$ , the matching similarity of the holed-region entity is  $W_{R_s}$ , the weight of the matching micro-spatial-scene set is  $W_{MSC}$ , and the similarity degree of the outer contour of the holed-region entity is  $W_{exshape}$ . Thus, to calculate the similarity of the complex scene,

$$Sim = W_{p_g} \times S'_{p_g} + W_{r1} \times S'_{r1} + W_{r2} \times S'_{r2} + W_m \times S'_m + W_s \times S'_s + W_{R_s} \times S'_{R_s} + W_{R_s} \times S'_{R_s} + W_{exshape} \times S'_{exshape} + W_{MSC} \times S'_{MSC}$$
(28)

#### 5. Experiments and Discussion

#### 5.1. Experiment Data

We chose the Great Lakes in North America in 1986 and 2015 as the study area to calculate the similarity of the five lakes during different periods to validate the hierarchical model for the similarity measurement of the complex holed-region entity scene. The study area is composed of five large freshwater lakes at the junction of Canada and the United States of America, which are Lake Superior, Lake Huron, Lake Michigan, Lake Erie, and Lake Ontario, as shown in Figure 11. The outer contours of the five Great Lakes and the area of the lakes have changed during different periods due to global warming, human activity, and other factors.

The fundamental method for the comparison of the variation of the Great Lakes is to calculate the similarity of Great Lakes in different years. We chose the Great Lakes' data in 1986 and 2015 to validate the hierarchical similarity model proposed in this paper. We treat the Great Lakes as regions that contain holes, and the islands in the lakes are regarded as holes in regional entities. The collection of the five lakes is a complex scene. The Great Lakes scene in 1986 and 2015 are, respectively, designated as Scenario *A* and Scenario *B*, and the number of holes and entities included in Scenario *A* and Scenario *B* are as shown in Table 2:



Figure 11. Great lakes study area.

Table 2. The number of holes and entities in Scenario A and B.

	Lake Superior	Lake Huron	Lake Michigan	Lake Erie	Lake Ontario	Sum
1986	3	6	2	2	3	16
2015	4	6	2	2	2	16

During the similarity measurement process of the two complex holed-region entity scenes, the similarities of the position graphs and geometric transformation similarities of CPPG should be first taken into account. Then, the geometry of the multi-holed-region and the holes is described by the multilevel chord length. The shape similarity is derived from the distance of the FFT description. Next, the directional consistency of the holes in the two matching regions should also be measured. To obtain the geometric similarity of the region and its holes, and the direction similarity between the holes, the top-down stratification method is used to find the correspondence between the regions and holes in this experiment. Finally, the weights of different dimensions are assigned to different levels, and the quantitative similarity between Scene *A* and Scene *B* is calculated.

# 5.2. Experimental Results and Analysis

# 5.2.1. Building and Measuring the Position Graphs

To build the complete distribution of the Great Lakes using position graphs, we ignore the holes in the multi-holed entities in Scene *A* and Scene *B*, considering them to be simple entities, and construct the CPPG, NTPPG, NPPG, and FPPG based the relationships noted in Section 2.2.1 of the two scenes, as shown in Figures 12 and 13.



**Figure 12.** Position graphs of the Great Lakes in 1986. (**a**) CPPG in 1986; (**b**) NTPPG in 1986; (**c**) NPPG in 1986; and (**d**) FPPG in 1986.



**Figure 13.** Position graphs of the Great Lakes in 2015. (a) CPPG in 2015; (b) NTPPG in 2015; (c) NPPG in 2015; and (d) FPPG in 2015.

Section 2.2.2 indicates that each position graph can be described by a set of tension, and the projection interval function  $F(\alpha)$  is a periodic change function with a variation period of  $180^{\circ}$ . In this experiment, we selected 180 sample points in the range of  $180^{\circ}$ . Each waveform corresponds to a single projection interval function, and each projection interval function corresponds to a unique position graph. The degree of matching of the two waveforms represents the similarity of the two position graphs. Using Formula (5), we obtain the MME of the two waveforms, and the similarity of the position graphs between Scenes *A* and *B* can be calculated by Formula (6). The CPPG, NTPPG, NPPG, and FPPG of the two scenarios are calculated and the results are shown in Table 3.

	CPPG	NPPG	FPPG	NTPPG
MME	0.101415	0.039020	0.030832	0.041860
MaxV	2.962511	1.365702	1.364719	1.420649
Sim	0.965755	0.971428	0.977408	0.970535

Table 3. The similarity results for the position graphs.

The data in Table 3 can be used to calculate the similarity between Scene A and Scene B as:

$$S'_{p_g} = (S'_{p_g1} + S'_{p_g2} + S'_{p_g3} + S'_{p_g4})/4$$
  
= (0.965755 + 0.971428 + 0.977408 + 0.970535)/4 = 0.9712815 = 97.12815\%

## 5.2.2. Great Lakes' Distribution Transformation Measurement

In reality, the surface entities of two scenes may shift position, or change in size. Thus, the similarity measurement process must consider the geometric transformation similarity of the position graph. We defined the center angle of the CPPG above and used it to expound the rotation of the position graph. One side of the center angle is the connection between the centroid of the scene and the centroid of the position graph, and the other side is the edge that connects the start point on the MBR of the scene with the centroid of the scene.

Figures 14 and 15 show that in this case, vector OOA runs from the center of Scene *A* to the center of the CPPG, vector OMA runs from the scene center to the start point of the MBR, and, in a similar manner, the vector OOB runs from the center of Scene *B* to the center of the CPPG, and vector OMB runs from the scene center to the start point of the MBR of Scene *B*.



Figure 14. The center angle of Scene *A*.



Figure 15. The center angle of Scene *B*.

The center angle  $\beta_A$  of Scene *A* is calculated to be 72.379°, the center angle  $\beta_B$  of Scene *B* is 70.935°, and the rotation transformation similarity (*Simr1* = 0.99599) between Scenes *A* and *B* can be obtained by Formula (7).

The degree of change in the relative direction angle is expounded when measuring the rotation similarity. As shown in Figure 16, the direction angles of the MBRs of Scene *A* and its CPPG are, respectively,  $43.677^{\circ}$  &  $45.098^{\circ}$ , and the direction angles of the MBRs of Scene *B* and its CPPG are, respectively,  $46.142^{\circ}$  &  $46.115^{\circ}$ .



Figure 16. Relative direction angle of Scene *A* and Scene *B*.

Therefore, the relative direction angles of Scenes *A* and *B* are, respectively,  $\theta_A - \lambda_A = 1.421^{\circ}$  and  $\theta_B - \lambda_B = -0.027^{\circ}$ . The rotation similarity (*Simr2* = 0.99598) between Scenes *A* and *B* can be calculated by Formula (8).

To determine the movement of the holed entities inside the scene, we measure the degree of the relative change of the center distance. According to Formula (9), the relative center distance of Scenes *A* and *B* can be computed as  $l_A = 0.24599$ ,  $l_B = 0.22117$ . Using the results in Formula (10), the translation similarity of the Scenes *A* and *B* is obtained (*Sim<sub>m</sub>* = 0.8991).

Because the size of the entities in the scene may change, we must measure the scaling similarity of the two scenes. After determining the scene center, it is easy to derive the extended Hausdorff distance of Scenes *A* and *B*, respectively, from Formula (11)–(13). The scaling similarity of two scenes can be obtained by taking the expansion Hausdorff distance into Formula (14) as  $Sim_s = 0.9944$ . The calculation results are shown in Table 4.

Hausdorff Dis	$R_1/R'_1$	$R_2/R'_2$	$R_3/R'_3$	$R_4/R'_4$	$R_5/R'_5$
$H^{f1}(O_A, R_i)/$	1647.6142/	1082.5446/	965.5698/	1355.1922/	1852.9394/
$H^{f1}(O_B, R'_i)$	1679.754	1063.1474	991.2311	1346.6466	1861.6845
$\frac{H^{f2}(R_i, O_A)}{H^{f2}(R'_i, O_B)}$	595.8135/	235.1493/	184.2397/	737.9486/	1099.1535/
	609.6859	201.5952	210.6782	741.5354	1092.4108
$\begin{array}{c} H^{f1f2} \left( O_{A}, R_{i} \right) / \\ H^{f1f2} \left( O_{B}, R'_{i} \right) \end{array}$	1647.6142/	1082.5446/	965.5698/	1355.1922/	1852.9394/
	1679.754	1063.1474	991.2311	1346.6466	1861.6845

Table 4. Hausdorff distance calculation results for Scenes *A* and *B*.

According to the above, transformation similarity results for the position graphs are shown in Table 5.

	Sim <sub>s</sub>	Sim <sub>m</sub>	Sim <sub>r1</sub>	Sim <sub>r2</sub>
Transformation Similarity (%)	0.9944	0.8991	0.99599	0.99598

5.2.3. Pairwise Matching and Shape Similarity of the Great Lakes

In this experiment, Scene *B* is the reference scenario, and Scene *A* is the matching scenario. The matching process of the holed-region entities comprises two steps: (1) filtering multi-holed-regions by a buffer analysis; (2) using the decision matrix to calculate the minimum Euclidean distance to determine the holed-region entity matching results.

The center and radius of the circular buffer must be determined in the first filtering step. The average radius of Scenes *A* and *B* is calculated first ( $R_A = 1382.3$ ,  $R_B = 1408.8$ ). The scene center  $d_{i_B}$  in Scene *B* is connected to the start point *P'* of the MBR of the CPPG to obtain a set of distances  $d_{Bi}$ , and the connection is vector  $\overrightarrow{P'd_i}$ ,  $d_{Ai} = d_{Bi}*(R_A/R_B)$ . Next, the angle between the direction vector of the MBR of the CPPG and  $\overrightarrow{P'd_i}$  is computed as follows: {62.1471°, 70.3416°, 90°, 86.3311°, 70.3416°}. Eventually, the corresponding point  $Q_i$  of the center point of each holed-region entity in Scene *B* can be found in Scene *A*.

As shown in Figure 17, with  $Q_i$  as the center, half of the maximum chord length of Hi is the radius ri, and a circular buffer is built on scene A. The matching holed-region entities are those contained in or that intersect  $\bigcirc Q_i$ .



Figure 17. Buffer analysis of Scenes A and B.

The matching results of the holed-region entities in Scene *B* are: the matching set of  $R'_1$  is  $\{R_1, R_3\}$ , the matching set of *R*' is  $\{R_1, R_2\}$ , the matching set of  $R'_3$  is  $\{R_2, R_3\}$ , and the matching set of  $R'_4$  is  $\{R_2, R_4, R_5\}$ ,  $R'_5$  is  $\{R_4, R_5\}$ .

Using the buffer filter, we obtain the matching set of each holed-region entity in Scene *B*. Next, we perform the Fast Fourier Transform (FFT) [31] to obtain the contour shape similarity of each pair of possible matching results. The results are shown in Table 6.

Region	<i>R′</i> <sub>1</sub>	<i>R</i> ′ <sub>2</sub>	<i>R</i> ′ <sub>3</sub>	$R'_4$	<i>R</i> ′ <sub>5</sub>		
	0.9007/0.9053	0.7223/0.7660	-	-	-		
$R_2$	-	0.8991/0.9201	0.8014/0.8392	0.7879/0.8185	-		
$R_3$	0.8009/0.8474	-	0.9528/0.9791	-	-		
$R_4$	-	-	-	0.9042/0.9335	0.7505/0.7860		
$R_5$	-	-	-	0.7716/0.7964	0.9832/0.9855		
Comment	A/B: $A$ is the results computed by the method in literature [25]; $B$ is the results computed by the method of multilevel chord description.						

Table 6. The contour shape similarity of the matching holed-region entities after filtering.

After filtering, we must obtain a definite set of matches so that each holed-region entity of the matching set in Scene *B* has the only corresponding surface entities within scene *A*. From Formula (16), we calculate the Euclidean distance between each holed-region entity of Scene *B* and the entities in its matching set, and bring the result into Formula (17) to obtain the decision matrix *D*.

	0.0114	0.1708	0	0	0		1	0	0	0	0	
	0	0.0417	0.2049	0.0884	0		0	1	0	0	0	
D =	0.1690	0	0.0338	0	0	X =	0	0	1	0	0	(29)
	0	0	0	0.0219	0.0986		0	0	0	1	0	
	0	0	0	0.1479	0.0115		0	0	0	0	1	

According to Formula (19), the minimum value of *Z* is 0.1203. Combined with the contour shape similarity measurement from Table 6, the final matching result can be determined by the Hungarian algorithm:  $\{(R'_1, R_1), (R'_2, R_2), (R'_3, R_3), (R'_4, R_4), (R'_5, R_5)\}$ .

On the basis of Formula (20), the similarity degree between scene *A* and Scene *B* is calculated as  $S'_{R S} = 0.9759$ .

Formula (27) is used to calculate the outer contour shape similarity of all pairwise holed-region entities of scene *A* to scene *B* ( $S'_{exshape} = 0.9447$ ).

#### 5.2.4. Inner Contour Similarity Measurement of the Matching Lakes

According to Section 2.1, a multi-holed-region can be treated as a micro-spatial-scene, and the hole within the region can be regarded as an ordinary surface entity. To obtain a matching relationship between two holes in corresponding micro-spatial-scenes, the association graphs  $G_n$  and  $G_m$  should be created to describe the constraints of the two micro-spatial-scenes. The node value is the shape similarity of two holes and the edge value is the directional similarity of two pairs of holes.

The shape similarity of holes in each pair of matching micro-spatial-scenes can be computed by fast Fourier descriptor. In this experiment,  $R_2$  (Lake Huron) is used as an example to show the matching analysis process, and the results are shown in Figure 18.

1.2 1 0.8 0.6 0.2 0.2 0 0 0	H4	H5	H6	H7	H8	H9
⊠ H'5	0.9486	0.7992	0.8722	0.7261	0.8265	0.8562
<b>⊡</b> H'6	0.8836	0.97	0.9105	0.8599	0.8956	0.8889
🖾 H'7	0.8551	0.9144	0.9502	0.8423	0.8668	0.902
🖸 H'8	0.8829	0.8652	0.8195	0.9652	0.7111	0.876
🖾 H'9	0.9049	0.8319	0.8922	0.8818	0.8212	0.9008
⊠H'10	0.8125	0.9121	0.8728	0.8838	0.7238	0.9253

Figure 18. Shape similarities of the inner contours of Lake Huron.

The association graph of the micro-spatial-scene  $R'_2$  can be arbitrarily constructed by a hole  $H'_{j}$ , and a corresponding hole of  $H'_j$  should be found in  $R_2$ . The association graph development of  $R'_2$  begins with  $H'_5$ , and the subsequent order of priority is set as  $H'_6$ ,  $H'_7$ ,  $H'_8$ ,  $H'_9$ . The construction process is:

- (1) According to the results in Figure 18, the shape similarity between  $H'_5$  and  $H_4$  is the highest. So,  $H'_5$  is paired with  $H_4$  and  $(H'_5, H_4)$  is designated as a node in the association graph.
- (2) Because  $H_4$  has been matched with  $H'_5$ ,  $H_4$  should be excluded from the matching analysis of  $H'_6$ .  $H_5$  has the highest similarity score with  $H'_6$  in the remaining holes, and consequently  $(H'_6, H_5)$  id designated as an associated node.
- (3)  $H'_{7}, H'_{8}, H'_{9}, H'_{10}$  are matched with  $H_{6}, H_{7}, H_{9}$  and  $H_{8}$  according to the same rules in step 1 and step 2, forming the nodes  $(H'_{7}, H_{6}), (H'_{8}, H_{7}), (H'_{9}, H_{9})$ , and  $(H'_{10}, H_{8})$ .

The' matching result for the inner holes of the micro-spatial-scenes  $R'_2$  and  $R_2$  generated by the above steps is { $(H'_5, H_4)$  ( $H'_6, H_5$ ) ( $H'_7, H_6$ ) ( $H'_8, H_7$ ) ( $H'_9, H_9$ ) ( $H'_{10}, H_8$ )}. The weights of all holes are set to the same value in this experiment. Given that different priority matching orders will lead to different matching results, we must consider the entire matching order of holes in the test.

In addition to the shape similarity, the similarity measurement of the matching holes also must consider the consistency of the direction relationship. When evaluating the consistency of the direction relationship between two holes in the micro-spatial-scenes  $R_2$  and  $R'_2$ , which are multi-hole regions, we should give the characteristic matrix of any pairs of holes in the scenes. A conceptual space of the F-Matrix is represented by the 4-D lattice [25]. The distance between two F-Matrixes on the 4-D lattice stands for the dissimilarity of the directions. After eliminating the redundancy, we obtain the association graphs in Figure 19.



Figure 19. Cont.



(b) Association graphs of solution2

**Figure 19.** Association graphs of possible matching solutions between  $R_2$  and  $R'_2$ .

Using Formulas (21) and (22), we can obtain the shape similarity and direction relationship similarity of each matching solution. The results are shown in Table 7.

Table 7. Shape similarity and direction relationship similarity of each matching solution.

Solution	1	2
Shape	0.9093	0.9303
Direction	0.8583	0.9083

It is easily to see that the shape similarity and direction relationship similarity of solution 1 are all higher than those in solution 2 from Table 7. We can compute the matching completeness of the two solutions with Formula (24) and then substitute the results of matching completeness, shape similarity, and direction similarity into Formulas (23) and (25) to obtain the similarities of the two solutions, as shown in Table 8.

Table 8. Integral similarities of the two matching solutions.

Solution	1	2		
Similarity	0.8838	0.9193		

In this experiment, the importance of the holes inside the Great Lakes are treated equally. It follows that all holes have the same weight, that is,  $\sum_{i=1}^{n} W_{Hi} = 1$  and  $\sum_{i=1}^{n(n-1)} W_{Di} = 1$ . We assume that the same weight is attached to the shape similarity, the direction relationship similarity, and the matching completeness; that is,  $W_{shp} + W_{dir} + W_{comp} = 1$ . Repeating the above processes yields all of the integral similarities of matching lakes, and the results are shown in Table 9.

Table 9. The similarities of all matching micro-spatial-scenes in Scenes A and B.

Region	$(R_1, R'_1)$	$(R_2, R'_2)$	$(R_3, R'_3)$	$(R_4, R'_4)$	$(R_5, R'_5)$
S' shape	0.9884	0.9608	0.9783	0.9936	0.9595
S' <sub>direction</sub>	0.9861	0.7833	1	0.875	0.9375
S' <sub>Comp</sub>	3/4	1	1	1	2/3
Similarity	0.9050	0.8721	0.9892	0.9343	0.8431

## 5.2.5. Similarity of the Great Lakes Scene

Based on the hierarchical model in Section 2.1, the similarity of the two complex scenes is the generalization of the similarity of its component micro-spatial-scenes  $R_i$  and the distribution similarity of  $R_i$ . The micro-spatial-scene similarity is considered to be an overall generalization of the geometric similarity of the hole  $H_i$  and the similarity of the distribution of  $H_i$  in R. Consequently, the similarity of

two complex holed-region entity scenes is regarded as the unity of the micro-spatial-scene similarities, the outer contour shape similarities, and distribution similarities of the holed-region entities, along with the transformation similarities of the CPPG.

In Section 5.2.3, we calculated the similarities of the outer contour shape of the multi-holed-regions as simple entities. In this experiment, we consider each holed-region equally important, that is, each holed-region has the same weight. Additionally,  $\sum_{i=1}^{n} W_{Ri} = 1$ . The overall similarity of all matching micro-spatial scenes can be obtained by Formula (26) ( $Sim_{MSC} = 0.9087$ ). At the end of experiment, according to the Analytic Hierarchy Process (AHP), the first-order weights of the similarity measurement is  $W_{MSC} = 0.3822$ ,  $W_{exshape} = 0.1387$ ,  $W_{p_{-}g} = 0.1991$ ,  $W_{r1} = 0.0471$ ,  $W_{r2} = 0.0471$ ,  $W_m = 0.0471$ ,  $W_s = 0.0471$ , and  $W_{R_{-}S} = 0.0916$ . As a result, the total similarity between Scenes *A* and *B* is

$$\begin{split} Sim &= W_{p\_g} \times S'_{p\_g} + W_{r1} \times S'_{r1} + W_{r2} \times S'_{r2} + W_m \times S'_m + W_s \times S'_s \\ &+ W_{R\_S} \times S'_{R\_S} + W_{exshape} \times S'_{exshape} + W_{MSC} \times S'_{MSC} \\ &= 0.9713 \times 19.91\% + 0.9960 \times 4.71\% + 0.9960 \times 4.71\% + 0.8991 \times 4.71\% + 0.9944 \times 4.71\% \\ &+ 0.9759 \times 9.16\% + 0.9447 \times 13.87\% + 0.9087 \times 38.22\% \\ &= 93.48\% \end{split}$$

## 5.3. Experimental Results and Analysis

The experiment measuring the similarity of the two Great Lakes scenes in North America between 1986 and 2015 does not involve other scenarios that might match the map of the Great Lakes; therefore, we do not need to consider that little difference exists in the index that may exist in the similarity calculation process of the multiple matching scenarios. Only when the comparison of degree of importance between two similarity metrics can be determined, can the Analytic Hierarchy Process be used to obtain the index weights of a complex scene with a multi-holed entity similarity measurement system.

It is known that the similarity measurement of a complex holed-region entity scene has eight first-order indexes: the similarity of the matching micro-spatial-scene collection  $S'_{MSC}$ , the outer contour shape similarity of the matching holed-region entities  $S'_{exshape}$ , the similarity of the position graphs  $S'_{p_g}$ , the transformation similarity of the CPPG (revolution similarity  $S'_{r1}$ , rotation similarity  $S'_{r2}$ , moving similarity  $S'_m$ , scaling similarity  $S'_s$ ), and the matching similarity of the holed-regions  $S'_{R_s}$ . In accordance with the AHP, an 8th order matrix A with respect to the relative weight of the similarity index is obtained, as shown in Table 10.

W(A)	W <sub>MSC</sub>	W <sub>exshape</sub>	$W_{p_g}$	<i>W</i> <sub><i>r</i>1</sub>	$W_{r2}$	$W_m$	$W_s$	$W_{R\_S}$
W <sub>MSC</sub>	1	3	2	8	8	8	8	4
Wexshape	1/3	1	1/2	3	3	3	3	2
$W_{p_g}$	1/2	2	1	4	4	4	4	2
$W_{r1}$	1/8	1/3	1/4	1	1	1	1	1/2
$W_{r2}$	1/8	1/3	1/4	1	1	1	1	1/2
$W_m$	1/8	1/3	1/4	1	1	1	1	1/2
$W_s$	1/8	1/3	1/4	1	1	1	1	1/2
$W_{R\_S}$	1/4	1/2	1/2	2	2	2	2	1

Table 10. Similarity index of the relative weight of the 8-order matrix.

The 8-order matrix of similarity index relative weight is normalized where the normalization formula is  $b_{ij} = \frac{a_{ij}}{\sum\limits_{i=1}^{m} a_{ij}}$ . The normalized matrix *B* of the similarity index is:

	0.3871	0.3830	0.4000	0.3810	0.3810	0.3810	0.3810	0.3636
D	0.1290	0.1277	0.1000	0.1429	0.1429	0.1429	0.1429	0.1818
	0.1935	0.2553	0.2000	0.1905	0.1905	0.1905	0.1905	0.1818
	0.0484	0.0426	0.0500	0.0476	0.0476	0.0476	0.0476	0.0455
D =	0.0484	0.0426	0.0500	0.0476	0.0476	0.0476	0.0476	0.0455
	0.0484	0.0426	0.0500	0.0476	0.0476	0.0476	0.0476	0.0455
	0.0484	0.0426	0.0500	0.0476	0.0476	0.0476	0.0476	0.0455
	0.0968	0.0638	0.1000	0.0952	0.0952	0.0952	0.0952	0.0909

 $B_i$  is obtained by summing each row of matrix *B* and is the eigenvector  $\lambda$  of matrix *A*.

$$\lambda^{-1} = \begin{bmatrix} 3.0575 & 1.1099 & 1.5926 & 0.3769 & 0.3769 & 0.3769 & 0.3769 & 0.7325 \end{bmatrix}$$
(30)

According to the AHP, the weights of the eight first-level indicators of the similarity measurement of the complex holed-region entity scene must satisfy:

$$W_i = \frac{\lambda_i}{\sum\limits_{i=1}^n \lambda_i}$$
(31)

The weights of the eight similarity index are computed as  $W_{MSC} = 0.3822$ ,  $W_{exshape} = 0.1387$ ,  $W_{p\_S} = 0.1991$ ,  $W_{r1} = 0.0471$ ,  $W_{r2} = 0.0471$ ,  $W_m = 0.0471$ ,  $W_s = 0.0471$ , and  $W_{R\_S} = 0.0916$ .

To test the consistency of the contrast matrix A, the maximum eigenvalue  $\lambda_{max} = 8.028$  of the matrix is calculated, and the consistency index is CI = 0.004. The average random randomness index RI is 1.41, CR = CI/RI = 0.004/1.41 = 0.0028 < 0.1, meaning that the contrast matrix A remains consistent.

## 6. Conclusions

To solve the matching problem of multiple lakes with multiple islands, several land areas with lakes, complex holed buildings, and other geographic entities with inner boundaries in GIS, we propose a new hierarchy similarity measurement method in this paper. We believe that the hierarchy similarity measurement method presented in this paper may be useful and is readily applicable to actual scenes for performing complicated spatial scene matching and querying by various users. During the process of similarity measurement, we dismantle the relationship between layers, and degrade a holed-region entity to a simple entity, initially. Afterwards, the similarity indexes of the complex scene's similarity measurement are analyzed. The holed-region entity is subsequently regarded as a non-boundary micro-spatial-scene, and the similarity of the micro-spatial-scene is taken into a complex scene to present the similarity measurement results of the complex holed-region entity scene.

At the complex scene level, we take the holed-region as a micro-spatial-scene, and a position graph method is proposed to generate the distribution similarity of the holed-regions. The relationship between the shape and the direction of the micro-spatial-scene is described by the association graph. Additionally, the similarity of the two micro-scenes is also calculated by the matching association graphs.

Although the eight first-order index weights of a complex holed-region entity scene are discussed in Section 5.3, further improvement is still needed. In this paper, we did not study the weight ratio of each measurement index, and we treated all holes and all regions that contain holes equally. Whereas, in reality, the areas, the hole positions, multi-holed-regions, and their associated factors all will impact the weight. In addition, the matching of multiple complex holed-region entity scenes has not been studied yet. The matching efficiency of a complex scene similarity measurement and the improvement of the weight adjustment scheme should be studied in future research.

Acknowledgments: This work is supported by the National key R & D program of China (2017YFC0602204), the National Natural Science Foundation of China (No. 41401443), the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (No. CUG160226), and the Open Research Fund of Teaching Laboratory in China University of Geosciences (No. skj2014168).

**Author Contributions:** Zhanlong Chen conceived and designed the experiments and performed the modeling; Rongrong Zhu analyzed the data and reviewed and edited the draft. All authors discussed the basic structure of the manuscript and read and approved the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

## References

- 1. Liu, Z. The Research on Areal Feature Matching among the Conflation of Urban Geographic Databases. Master's Thesis, Wuhan University, Wuhan, China, June 2006.
- 2. Walter, V.; Fritsch, D. Matching spatial data sets: A statistical approach. *Int. J. Geogr. Inf. Sci.* **1999**, *13*, 445–473. [CrossRef]
- 3. Cobb, M.A.; Chung, M.J.; Foley, H., III; Petry, F.E.; Shaw, K.B.; Miller, H.V. A rule-based approach for the conflation of attributed vector data. *GeoInformatica* **1998**, *2*, 7–35. [CrossRef]
- 4. Saalfeld, A. Conflation automated map compilation. Int. J. Geogr. Inf. Syst. 1988, 2, 217–228. [CrossRef]
- 5. Tang, W.; Hao, Y.; Zhao, Y.; Li, N. Feature matching algorithm based on spatial similarity. *Int. Soc. Opt. Photonics* **2008**, *7*147, 714704.
- Zhang, X.H.; Zhu, H.H.; Zheng, Q.Z. Hungary algorithm concerning best allotting and c++. *Logist. Technol.* 2004, *5*, 25.
- 7. Van der Werff, H.M.A.; van der Meer, F.D. Shape-based classification of spectrally identical objects. *ISPRS J. Photogr. Remote Sens.* **2008**, *63*, 251–258. [CrossRef]
- 8. Ai, T.; Cheng, X.; Liu, P.; Yang, M. A shape analysis and template matching of building features by the fourier transform method. *Comput. Environ. Urban Syst.* **2013**, *41*, 219–233. [CrossRef]
- 9. Xie, P.; Xiaoyong, M.A.; Zhang, X.; Lin, M. A new fast algorithm for complex polygons matching. *Comput. Eng.* **2003**, *29*, 177–181.
- Mount, D.M.; Netanyahu, N.S.; Moigne, J.L. Efficient algorithms for robust feature matching. *Pattern Recognit*. 1999, 32, 17–38. [CrossRef]
- 11. Masuyama, A. Methods for detecting apparent differences between spatial tessellations at different time points. *Int. J. Geogr. Inf. Sci.* 2006, 20, 633–648. [CrossRef]
- 12. Yan, H.W.; Wang, J.Y. A generic algorithm for point cluster generalization based on voronoi diagrams. *J. Image Graph.* **2005**, *17*, 633–636.
- 13. Du, Q.; Yang, P.; Tan, R. Classification of river networks structure based on spatial statistical character. *Geomat. Inf. Sci. Wuhan Univ.* **2006**, *31*, 419–422.
- 14. Lee, D.J.; Antani, S.; Long, L.R. Similarity measurement using polygon curve representation and fourier descriptors for shape-based vertebral image retrieval. *Proc. SPIE* **2003**, *5032*, 1283–1291.
- 15. Al, T.; Shuai, Y.; Li, J. A spatial query based on shape similarity cognition. *Acta Geod. Cartogr. Sin.* **2009**, *38*, 356–362.
- Schneider, M.; Behr, T. Topological relationships between complex spatial objects. ACM Trans. Database Syst. 2006, 31, 39–81. [CrossRef]
- 17. Disser, Y.; Ghosh, S.K.; Mihalák, M.; Widmayer, P. Mapping a polygon with holes using a compass. *Theor. Comput. Sci.* **2014**, *553*, 106–113. [CrossRef]
- Vasardani, M.; Egenhofer, M.J. In Comparing Relations with a Multi-Holed Region, Spatial Information Theory. In Proceedings of the International Conference, COSIT 2009, Aber Wrac'h, France, 21–25 September 2009; pp. 159–176.
- 19. OpenGIS Consortium Inc. Opengis simple features specification for SQL. In *Technical Report*; Revision 1.0; OpenGIS Consortium Inc.: Wayland, MA, USA, 1999.
- 20. Egenhofer, M. Topological relations between regions with holes. *Int. J. Geogr. Inf. Sci.* **1994**, *8*, 129–142. [CrossRef]

- 21. Clementini, E.; Felice, P.D.; Califano, G. Composite regions in topological queries. *Inf. Syst.* **1995**, *20*, 579–594. [CrossRef]
- 22. Egenhofer, M.J.; Vasardani, M. Spatial Reasoning with a Hole, Spatial Information Theory. In Proceedings of the International Conference, COSIT 2007, Melbourne, Australia, 19–23 September 2007; pp. 303–320.
- 23. Vasardani, M.; Egenhofer, M.J. Single-Holed Regions: Their Relations and Inferences. In Proceedings of the International Conference on Geographic Information Science, Park City, UT, USA, 23–26 September 2008; pp. 337–353.
- 24. Yokoo, M. Constraint Satisfaction Problem; Springer: Berlin/Heidelberg, Germany, 2001; pp. 21-70.
- 25. Xu, Y.; Xie, Z.; Chen, Z.; Wu, L. Shape similarity measurement model for holed polygons based on position graphs and fourier descriptors. *Int. J. Geogr. Inf. Sci.* **2016**, *31*, 253–279. [CrossRef]
- 26. Heijmans, H.J.A.M.; Tuzikov, A.V. Similarity and symmetry measures for convex shapes using minkowski addition. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 980–993. [CrossRef]
- 27. Fan, L.T.; Wu, S.Y.; Chen, J. An approach to similarity measures for polygonal shapes based on mechanics. *J. Shanghai Jiaotong Univ.* **2003**, *37*, 874–877.
- Deng, M.; Li, Z. L.; Chen, X. Y. Extended hausdorff distance for spatial objects in GIS. *Int. J. Geogr. Inf. Sci.* 2007, 21, 459–475.
- 29. Feng, X.U.; Deng, M.; Zhao, B.; Chen, J. A detailed investigation on the methods of object matching: A detailed investigation on the methods of object matching. *J. Geo-Inf. Sci.* **2009**, *11*, 657–663.
- 30. Xiaoya, A.N.; Sun, Q.; Qiang, X.; Wei, Y. A shape multilevel description method and application in measuring geometry similarity of multi-scale spatial data. *Acta Geod. Cartogr. Sin.* **2011**, *40*, 495–508.
- Chen, Z.; Qin, M.; Liang, W.U.; Xie, Z. Establishment of the comprehensive shape similarity model for complex polygon entity by using bending mutilevel chord complex function. *Acta Geod. Cartogr. Sin.* 2016, 45, 224–232.
- 32. Kaplansky, I. Rings of Operators; W. A. Benjamin, Inc.: Amsterdam, NY, USA, 1968.
- 33. Liu, T. Similarity of Spatial Group Objects; Wuhan University: Wuhan, China, 2011.
- 34. Serra, J. Image Analysis and Mathematical Morphology; Academic Press: Cambridge, MA, USA, 1982; p. 536.
- 35. Kuhn, H.W. The hungarian method for the assignment problem. Nav. Res. Logist. 2005, 52, 7–21. [CrossRef]
- 36. Markman, A.B.; Gentner, D. Structural alignment during similarity comparisons. *Cogn. Psychol.* **1993**, 25, 431–467. [CrossRef]
- 37. Blaser, A.D. Sketching Spatial Queries. Ph.D. Thesis, University of Maine, Orono, ME, USA, 2000.
- 38. Egenhofer, M.J. Query processing in spatial-query-by-sketch. J. Vis. Lang. Comput. 1998, 8, 403–424. [CrossRef]
- 39. Nedas, K.A.; Egenhofer, M.J. Spatial-scene similarity queries. Trans. GIS 2008, 12, 661-681. [CrossRef]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).