A Voronoi-like model of spatial autocorrelation for characterizing spatial patterns in vector data

Xiang Zhang*†, Tinghua Ai†, Jantien Stoter*
*International Institute for Geo-Information Science and Earth Observation (ITC)
P.O. Box 6, 7500 AA Enschede, the Netherlands
Email: {xzhang, stoter}@itc.nl
†School of Resource and Environment Sciences, Wuhan University
430072 Wuhan, China
Email: tinghua_ai@tom.com

***Abstract*—The paper presents a computational model of spatial autocorrelation based on a Voronoi-like auxiliary structure. It shows that the Voronoi-like partition of map objects can be used to discern spatial patterns (e.g. clustered or dispersed) of geographic phenomena. In this paper, we transform the problem of characterizing the patterns for different geometry types (i.e. points, curves, and polygons) into a process of calculating spatial autocorrelation based on the auxiliary partition units. The method is shown to be successful for the designated tasks.**

***Keywords*-spatial patterns; spatial autocorrelation; voronoi-like structure; conforming delaunay triangulation;**

I. Introduction

In the geo-spatial information domain, spatial patterns in vector data (i.e. points, curves, and polygons) are of great importance. Here Spatial patterns are described in terms of how disperse spatial phenomena are distributed in a two-dimensional space; they are traditionally visualized and analyzed by maps [1], [2]. The Pattern analysis of map objects can provide insight into many processes (e.g. in geography, urbanization, soil, crime, and social sciences) where spatial component plays an essential role [1]–[3]. A well-known application of this analysis is in discovering the source of the London Cholera Epidemic of 1854 [4]. In addition, the spatial patterns of various map objects are crucial for the map generalization process, and some of the patterns (e.g. network patterns) should be preserved in the process (e.g. in an urban context) [5]. However, few methods are capable of describing the patterns of points, networks, and polygons in a quantitatively and objectively way.

Unlike the visual inspection of maps, spatial autocorrelation—a family of non-graphical indicators—is commonly used to characterize[1] spatial patterns [2]; it measures the degree to which the spatial objects are intercorrelated in space [6]. However, most of them can only be applicable to raster data or a few representations, i.e. the point patterns and the patterns of variables associated with adjacent geographic units (e.g. administrative boundaries), and the indicators have various limitations [3], [6]–[8]. To characterize spatial point patterns, for example, the Nearest Neighbor Index (NNI) and Ripley's K can be used to characterize the point patterns [1], [9], but the NNI index varies greatly depending on the boundary defined for the points [10]. Another statistic is the quadrat analysis of point features, which divides the study area into a regular grid; yet [7] shows that this method is not well suited for analyzing point patterns, the result of this method is unpredictable when the size of the grid is changing. This is also noted by the author of [10] that the quadrat method only yields promising results when an optimal grid is determined. For the spatial patterns of variables associated with adjacent polygonal units, commonly used statistics include Moran's I, Geary's C, and joint count analysis. Moran's I is, among these others, the most widely adopted statistic of spatial autocorrelation [7]. Although it can be used together with the quadrat analysis to summarize point patterns, the problem remains that an optimal grid has to be identified. While some other related methods can be found in the literature which can characterize the network features, such as fractal dimension on networks[2] and the Candy model [11], the former aims to describe the self-similarity in networks and the latter treats the line segments in a network as a result of a marked point process constrained by an interaction model [11]. Few applications of both fractal dimension and the Candy model are available for charactering whether a network is clustered or not. Above all, few of these above-mentioned methods are able to characterize spatial patterns for a set of curves (e.g. networks) or disjoint polygons, which not only are of particular interest for this paper but also have a wide range of applications in e.g. crime analysis, urbanization process, and automated map generalization.

To overcome the above deficiencies, especially the com-

[1]The meaning of 'characterize' in this paper is 'describe in a quantitative and objective way'.

[2]http://en.wikipedia.org/wiki/Fractal_dimension_on_networks

978-0-7695-3781-8/09 $25.00 © 2009 IEEE
DOI 10.1109/ISVD.2009.19

mon inability to characterize spatial patterns for curves and disjoint polygons, we propose a novel approach to calculate spatial autocorrelation for points, curves, and polygons. The underlying hypothesis is that the characterization of spatial patterns for various geometries can be achieved by a Voronoi-like partition of the geometries.

The key idea is as follows. After partitioning the space by Voronoi-like Diagram, we obtain a non-overlapping and exhaustive partition of the study area, and each partition unit contains a single feature (e.g. a point, curve, or polygon). The process of characterizing patterns for the underlying geometries is then transformed into a process of quantifying the spatial autocorrelation for the partition units. In this paper, we adopt Moran's I coefficient to calculate spatial autocorrelation based on the partition units. This way, this spatial autocorrelation could be extended in a way that enables the characterization of spatial patterns for e.g. point symbols, areal buildings, and river or transportation networks, in an objective and consistent manner.

In the reminder of this paper, we first introduce the statistical basis of Moran's I coefficient to make it clear what this coefficient is and how it can be used in characterizing spatial patterns (Section II). Then a computational model is presented in Section III, which integrates the idea of Moran's I coefficient with a Voronoi-like structure. This model can characterize spatial patterns for points, curves, and polygons in a consistent way. The experimental results in Section IV validate the proposed model. Finally, the paper ends with discussion (Section V) and conclusion (Section VI).

II. Fundamentals of Moran's I Coefficient

As mentioned above, Moran's I coefficient is widely used in quantifying spatial autocorrelation in the literature. We will now introduce how to calculate the I coefficient based on the above-mentioned materials [3], [6], [7], [9]. The Moran's I is defined as follows:

$$I = \frac{N \sum_i \sum_{j,\, i \neq j} W_{ij}(Y_i - \overline{Y})(Y_j - \overline{Y})}{(\sum_i \sum_{j,\, i \neq j} W_{ij}) \sum_i (Y_i - \overline{Y})^2} \qquad (1)$$

where N is the number of the variables Y_i associated with spatial units indexed by i in the study area; Y_i and Y_j are the values of Y associated with unit i and j respectively (see different colors in Figure 1); $\overline{Y}$ is the average of Y; W_{ij} is a weighting function measuring the spatial contiguity of locations i and j. Spatial units in Moran's I are usually represented by polygons that are topologically adjacent to each other (Figure 1). For example, administrative boundaries are commonly used spatial units.

The I coefficient presented above is a global index of spatial autocorrelation, which has several important properties. Firstly, the index evaluates whether the pattern expressed is clustered, dispersed, or random. A Moran's I value ($|I| \leq 1$) near $+1.0$ indicates clustering (Figure 1, right end) while a

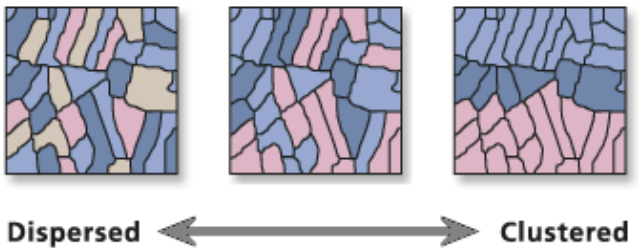


Figure 1. An illustration of Moran's I coefficient (Modified from the source: ArcGIS 9.2, Moran's I, http://edndoc.esri.com)

value near -1.0 indicates dispersion (Figure 1, left end). When the value is not significantly different from 0.0, there is no spatial autocorrelation: the variables associated with spatial units is completely random [6], [7]. Secondly, the polygonal units based on which variables are collected must be non-overlapping and exhausting the study area [7].

Another important issue is the determination of the weighting function (W_{ij}). In the simplest case, topological contiguity is a popular way to calculate W_{ij}. The function is defined as:

$$W_{ij} = \begin{cases} 1 & \text{if i, j are 1st-order neighbors} \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

Note that the first-order neighbor is a relationship that holds for those spatial units who are topologically adjacent to each other; W_{ij} can be either symmetric or asymmetric. Other options to define W_{ij} include area, common boundary length, distance (e.g. $\frac{1}{d^2}$), higher-order neighbor, and their combinations [3], [7]. Most of them need the support of geometric algorithms.

Different from common applications of Moran's I which analyzes the patterns of variables associated with topologically adjacent spatial units, we aim to explore a pragmatic way to characterize how disjoint or connected geometries (sets of points, curves and polgons) are intercorrelated in space. The spatial units of these geometries are later on specified by the Voronoi-like structure.

III. A Voronoi-like Computational Model

To achieve our goals defined in the Introduction, a Voronoi-like computational model integrating the idea of Moran's I statistic (Section II) is presented in this section. We first define the basic geometries and topology required for the proposed model in Section III-A. After partitioning the space by skeletonizing the conforming Delaunay triangulation, the Voronoi-like partition of the underlying geometries are formed (Section III-B), based on which the elements of Moran's I statistic (i.e. the variable Y and weighting function W_{ij}) can be defined (Section III-C).

A. Basic geometries and topology

Let $\mathcal{S}$ be a finite set of points; $\mathcal{L}$ be a finite set of curves; $\mathcal{A}$ be a finite set of polygons ($\mathcal{S}, \mathcal{L}, \mathcal{A}$ is in $\mathbb{R}^2$). Note that a curve $l_i \in \mathcal{L}$ commonly defined in vector data consists of a set of connected closed linear segments (see the definition of linear segments in [12]); a curve can either have at most one common point (i.e. two endpoints coincide) or have no common point at all. We use $\mathcal{CDT}$ in $\mathbb{R}^2$ to denote the conforming Delaunay triangulation [13]. $\mathcal{CDT}$ is constructed on these primitive geometries respectively. In the case of points, $\mathcal{CDT}$ is nothing but the usual Delaunay triangulation for a finit point pattern

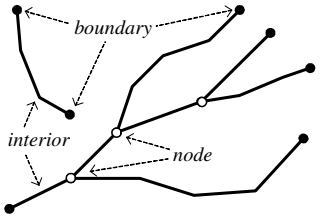
In [12], the Voronoi diagram for a set of segments is defined as the set of points $\in \mathbb{R}^2$ that are closer (or at equal distance) to $segment_i$ than any other segment in the set, where the Voronoi cells for segments and their endpoints are treated separately. We adapt this definition to the case of a set of curves and polygons, with a different treatment to the Voronoi cells of endpoints. We define the Voronoi-like structure as an approximate realization of the Voronoi diagram for curves and polygons. Its construction is outlined in Section III-B. In our definition, all points in a curve or edges of a polygon are treated as a whole when constructing the Voronoi-like structure.

Topological relationships of the sets (relationships that should be preserved for elements in a set) are crucial for the consistency of this model. The definitions and notions of *interior* (x_i°) and *boundary* (∂x_i) of a geometry x_i are adopted from the Point-Set Topology [14] to define the required topology, where $x_i \in \{\mathcal{S} \cup \mathcal{L} \cup \mathcal{A}\}$ is an instance of any geometry type. Illustrations are shown in Figure 2.

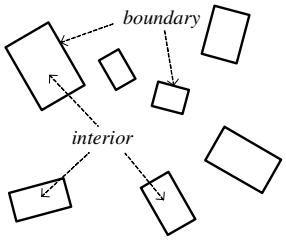
1) Topology of $\mathcal{L}$: Basic topology for $\mathcal{L}$ where its containing curves are disjoint is: $\forall\, l_i, l_j \in \mathcal{L}, i \neq j : l_i^\circ \cap l_j^\circ = \emptyset$, where $1 \leq i, j \leq |\mathcal{L}|$(the number of $\mathcal{L}$); in the case of network features, an additional rule must be added: $\forall\, l_i \in \mathcal{L}, \exists l_j \in \mathcal{L}, i \neq j : \partial l_i \cap \partial l_j \neq \emptyset$. These topological rules make it impossible to intersect any two curves in $\mathcal{L}$ without adding *nodes* to intersections. A *node* ($n_i \in \mathcal{N}$) can thus be defined as a point that coincides with boundaries of at least two curves: $\exists\, l_i, l_j \in \mathcal{L}, i \neq j : (n_i \cap \partial l_i \neq \emptyset) \wedge (n_i \cap \partial l_j \neq \emptyset)$. The concept of *interior*, *boundary*, and *node* in the occurrence of curves are illustrated in Figure 2a. Note that a curve has two boundaries, i.e. the two end vertices, and that *nodes* only occur in network features.

2) Topology of $\mathcal{A}$: The boundaries of a polygon (∂a_i) can be seen as a closed curve, where one end vertex coincides with the other. The internal topology of polygons in the paper is relatively simple: $\forall\, a_i, a_j \in \mathcal{A}, i \neq j : a_i \cap a_j = \emptyset$, meaning that all polygons can neither touch nor overlap each other, i.e. they are disjoint. The *interior* and *boundary* are displayed in Figure 2b.

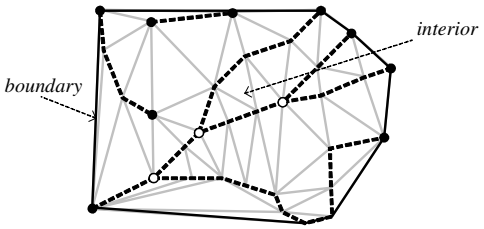
3) Topology of $\mathcal{CDT}$: The topology of $\mathcal{CDT}$ is defined with respect to the primitive geometries. An important rule that has to be kept is that: $\forall\, t_i \in \mathcal{CDT}, s_i \in \mathcal{S}, l_i \in \mathcal{L}, a_i \in \mathcal{A} : (t_i^\circ \cap s_i = \emptyset) \wedge (t_i^\circ \cap l_i = \emptyset) \wedge (t_i^\circ \cap \partial a_i = \emptyset)$. It means that no *interior* of any triangle intersects points, curves, or the edge of polygons. The extent of $\mathcal{CDT}$ is the union of all its triangles which forms a polygon, and the *interior* and *boundary* of $\mathcal{CDT}$ refer to those of this polygon respectively. These two concepts and the relationship between a $\mathcal{CDT}$ and a set of curves are demonstrated in Figure 2c.



(a) Topology in a curve case (interior, boundary, and node)



(b) Topology in a polygon case



(c) Conforming Delaunay Triangulation

Figure 2. Illustrations of primitive geometries and topology

$\mathcal{CDT}$ is constructed with Steiner Points [13] added to curves or edges of polygons, aiming at keeping the shape of triangles as right as possible. According to [15], the additional points on the geometries will improve the quality of the partition units derived in Section III-B.

B. Partitioning the geometric space with a Voroni-like structure

Now we will describe later in this section the first step of the proposed model, i.e. to partition the geometry space covered by a passage of triangles in the conforming Delaunay triangulation, resulting in the Voronoi-like partition. The partition for points is specialized into calculating the Voronoi diagram for them, thus it will not be described in detail. Here we mainly focus on the procedure of partition for curves. Later in this section, we will show that this procedure can be generalized to cover

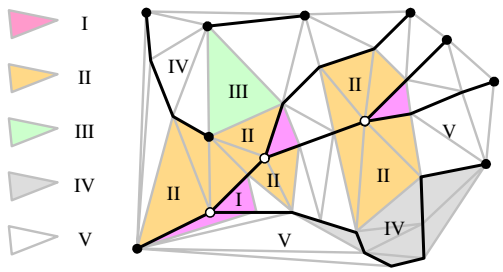


Figure 3. Classification of triangles in the Delaunay triangulation

the case of polygons.

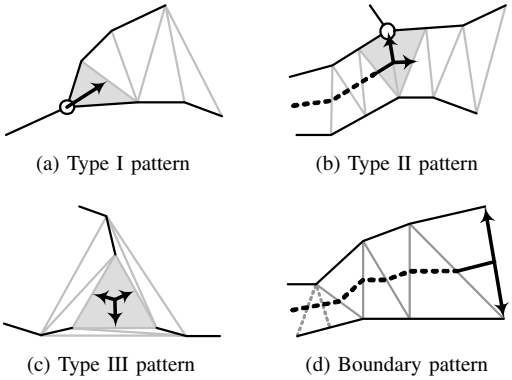
Based on the primitive geometries defined in Section III-A, five types of triangles that are crucial for the consequent partitioning procedure are identified and illustrated in Figure 3. The triangle types are defined as follows:

- T-I This is a $t_i \in \mathcal{CDT}$ that satisfies: $\partial t_i \cap n_j \neq \emptyset$ and $|\{l_k \in \mathcal{L} \,|\, dim(\partial t_i \cap l_k) = 1\}| = 2$, meaning that at least one *node* coincides with a vertex of the triangle, and edges from exactly two curves coincide with the edges of the triangle.
- T-II triangle satisfies: $\partial t_i \cap n_j \neq \emptyset$ and there are always two vertices of t_i which coincide with two vertices of a same curve.
- T-III For each triangle, it is possible the case that each of its vertices coincides with more than one curve. For each vertex in a T-III triangle, there is at least one coincident curve differs from the curves coinciding with the other two vertices of the triangle.
- T-IV triangle has all its vertices coincided with some curve, though some other curves may coincide with them.
- T-V is the type of all the remaining triangles, and is the white color triangles in Figure 3.

Note that if the $\mathcal{CDT}$ is constructed upon polygonal objects as defined in Section III-A, only T-III, -IV, and -V are possible. This is due to the absence of *nodes*, or of intersections between polygons.

1) Skeletonization patterns: Partitioning is based on the skeletonization of all types of triangles except the T-IV triangles, because this type is always inside polygons or the bends of curves, and is therefore regarded to be the private space of these geometries. Four skeletonization patterns are described as follows (see also Figure 4).

First of all, new skeleton points are usually determined at the middle of triangle edges (notated as *MP*). The skeletonization of a T-I triangle is shown in Figure 4a. The skeleton starts from a *node* coinciding with the triangle, to the opposite edge on the triangle. A T-II triangle is a place where a skeleton has to be divided into two branches, as the skeleton is entering a place occupied by two objects (Figure 4b). Likewise, a skeleton starting from the *barycenter* of



(a) Type I pattern (b) Type II pattern

(c) Type III pattern (d) Boundary pattern

Figure 4. Four skeletonization patterns

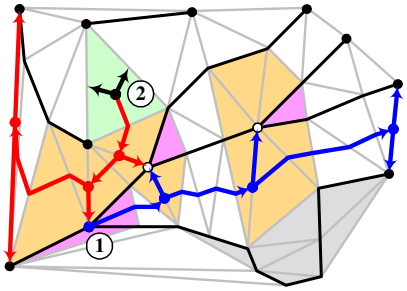


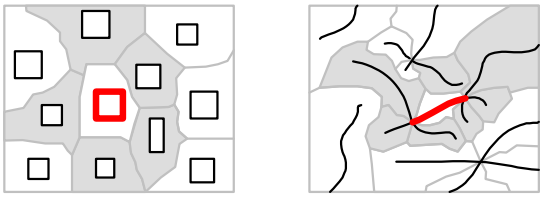
Figure 5. Examples of partitioning scheme for a network feature

a T-III triangle always have three branches (Figure 4c), since the triangle is occupied by three objects. When a skeleton is approaching to the boundary of the $\mathcal{CDT}$, as is depicted in Figure 4d, it must be divided into two branches pointing toward opposite directions. The skeletons of T-V triangles are formed by consecutively connecting the *MP* pairs on two edges of each T-V triangle.

2) Partitioning scheme: The next step is to put all the skeletonization patterns into a partitioning scheme described as follows and demonstrated in Figure 5.

1) The partitioning starts from a T-I triangle and ends with a T-III, another Type I triangle, or the boundary of $\mathcal{CDT}$ (No. 1 in Figuer 5);
2) Alternately, it starts from a T-III triangle and ends with a T-I, another T-III triangle, or the boundary of $\mathcal{CDT}$ (No. 2 in Figuer 5);
3) During step 1, and 2, the skeletonization pattern described in Figure 4b must be deployed in case of entering those T-II triangles.
4) The procedure repeats the first three steps until all the T-I and -III triangles have been visited. Using the information collected during the skeletonization process, e.g. the left and right objects associated with each skeleton, the partition units are then formed by connecting adjacent skeletons around each object.

Note that for the partition of polygons, step 1 and 3 are ignored, as there is no triangle of T-I and -II in such a case.



(a) A set of polygons (b) A set of curves

Figure 6. 1st-order neighbor in the Voronoi-like structure (the red bold shapes are target objects; the gray partition units are the 1st-order neighbors of the target objects)

The resulting partition is termed Voronoi-like structure as it is derived from a Delaunay triangulation structure in a similar way in which the Voronoi diagram is generated. The Voronoi-like structure for various geometry types is demonstrated in Section IV (Figure 7). An observation of the partitions indicates that there is an intrinsic relationship between the distribution of the partition units and the pattern of the original geometries. Hence they can be used as an indirectly indicator to characterize the spatial patterns for these geometries.

C. Defining the variable and weighting function

1) The Variable in Moran's I: We introduce two measures based on the partition units for defining the variable Y in Moran's I statistic. They are the *area* and the *individual density* of partition units. The individual density is defined as follows:

$$Density_S = 1/Area_{PU} \quad (3)$$

$$Density_L = Length_L/Area_{PU} \quad (4)$$

$$Density_A = Area_A/Area_{PU} \quad (5)$$

where Equation 3, 4, and 5 measures the density of a point, curve, and polygon in $\mathcal{S}$, $\mathcal{L}$, and $\mathcal{A}$, respectively; $Area_{PU}$ denotes the area of the partition unit of the geometry; $Length_L$ and $Area_A$ stand for the length of the curve and the area of the polygon in question, respectively.

Note that both the area of partition units and the individual density can be used to calculate Moran's I coefficient for characterizing the patterns of these geometries, as shown in Section IV.

2) The weighting function: For simplicity and clarity reasons, we only examplify the use of the first-order neighbor (Equation 2) as the weighting function. The first-order neighbor relationship can be clearly defined based on the proposed Voronoi-like structure. That is, the first-order neighbor for points are represented by adjacent Voronoi cells, and curves and disjoint polygons by adjacent Voronoi-like cells. For curves and polygons, more specifically, those two objects that are connected directly by at least a Delaunay edge or by a *node* (in the case of curves) are first-order neighbors, while those otherwise are not (see Figure 6 for example). Information on the first-order neighbors of geometries are stored in the Voronoi-like structure.

Until now we can say that the basic computational model has been complete. The experiment carried out on a set of datasets in Section IV verifies the usefulness of this model for various geometry types. Also the interpretation of results is given with respect to the idea of spatial autocorrelation.

IV. Experimental Results

The major aim of the section is to validate the proposed extended spatial autocorrelation method, and to explore its descriptive power in characterizing the spatial patterns for various features. In the experiment, the computational model was implemented in a interactive map generalization system, called DOMAP, written in OBJECT-ORIENTED C++.

The results of the partitioning algorithm for networks and polygons are illustrated in Figure 7. Here, dataset No. 1 is a river network data and dataset No. 2 is a set of polygons at map scale 10k (1:10000).

A. Evaluation of spatial patterns

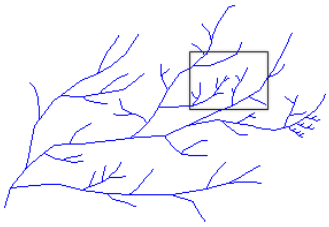
1) Patterns of points: Four point datasets exhibiting three hypothetical spatial patterns, i.e. regular, complete spatial randomness, and clustered, are shown in Figure 8.

Moran's I coefficient was calculated using the proposed model; individual density was selected for the variable Y. Note that the use of area measure has a similar effect on the I values in the case of point patterns. The values of the I is listed in the Table I.

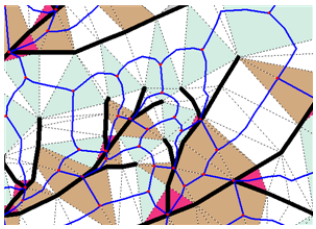
Table I
THE CALCULATED I VALUES FOR THE HYPOTHETICAL POINT PATTERNS

Point patterns					
Dataset No.		3	4	5	6
I	density	0.32	0.01	0.74	0.74

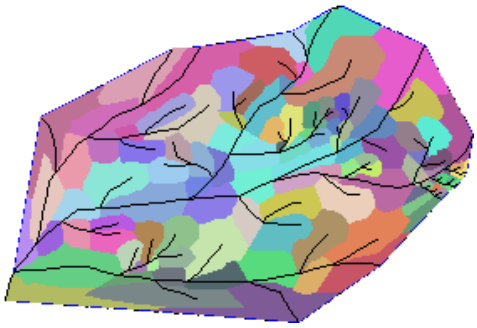
The computed I clearly distinguishes clustered point patterns from dispersed ones. The higher the value, the more clustered the pattern is. An exception is the value for the dataset No. 3. It seems plausible that the value (0.32) indicate a clustered distribution, while the dataset exhibit a perfect regular pattern visually. A possible account for this is that the Voronoi units laying on the boundary are clipped in our implementation, which thus makes the boundary units smaller or much denser than others (Figure 9, a darker color indicates a denser unit). This effect makes the distribution a bit clustered for these boundary units. In the experiment, we also calculated the coefficient for a similar point pattern (i.e. 20×20 points aligned on a quadrat grid), the result is 0.25 indicating also a bit clustered. Note that the boundary effect



(a) Dataset No. 1: A branching river network

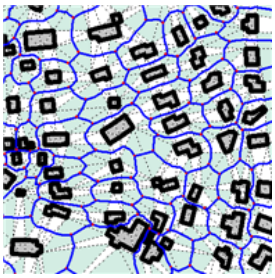


(b) Skeletonization: A snapshot for the portion in the rectangle of the left figure

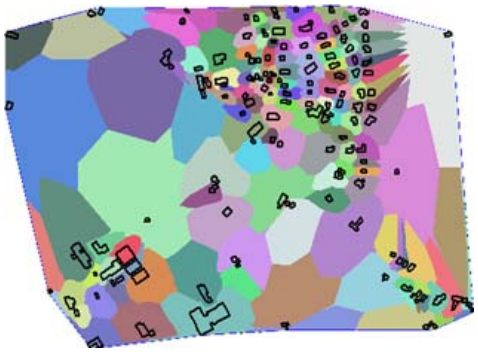


(c) The partition of the whole network

(d) Dataset No. 2: A set of buildings at 10k (Source: Kadaster, the Netherlands)

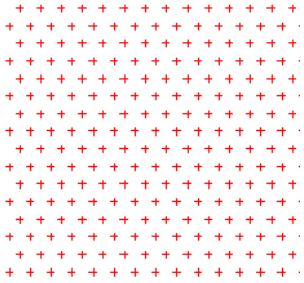


(e) Skeletonization: A snapshot (the rectangle portion in the left figure)

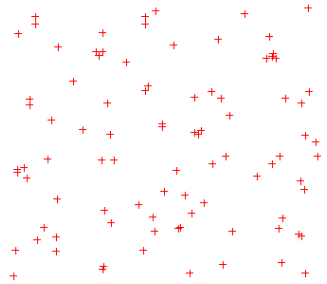


(f) The partition of the set of polygons

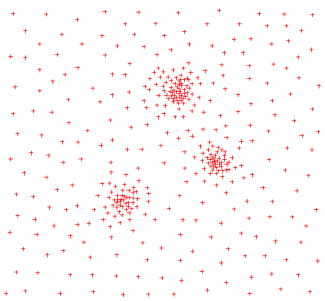
Figure 7. Voronoi-like partitions



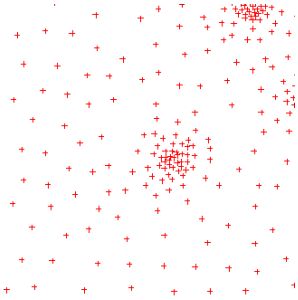
(a) Dataset No. 3: Regular



(b) Dataset No. 4: Completely Random



(c) Dataset No. 5: Clustered



(d) Dataset No. 6: Clustered

Figure 8. Hypothetical point patterns

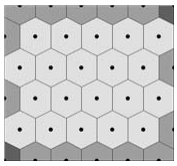
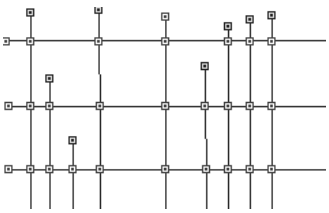


Figure 9. Boundary effect

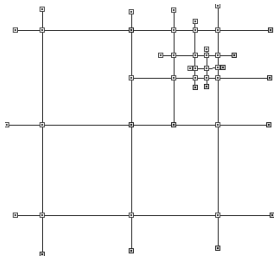
influences the estimated I values for all datasets, but the influence is even greater in the perfect regular point patterns.

2) Patterns of curves: Since to our knowledge there is no comprehensive studies on characterizing the spatial patterns of curves and disjoint polygons in terms of dispersion, there seems to be no way to give standard hypothetical patterns for such features. As a result, we validate the calculated values based on the common human perception.

The datasets used here is dataset No. 1, 7, and 8. The latter



(a) Dataset No. 7



(b) Dataset No. 8

Figure 10. Network features

two are shown in Figure 10. Both area and density measures are used to calculate the variable Y, and the resulting I values are listed in Table II.

Table II
THE I VALUES FOR DIFFERENT NETWORKS

Network patterns				
Dataset No.		1	7	8
I	density	0.25	-0.03	0.49
	area	0.36	0.28	0.74

The above results reflect, regardless of the choice of measures, the ordinal relationship between the three datasets in terms of dispersion: dataset No. 7 $\Rightarrow$ No. 1 $\Rightarrow$ No. 8 (from lower to higher clustering). They are consistent with our perception. For instance, the north-east part of dataset No. 8 is a clustering center, while the branching river (dataset No. 1) is not as clustered.

It is worth noting that the use of different measures for calculating the I for curves shows deviations between the I values. The I value of dataset No. 8 based on area measure indicate a significantly high level of clustering, it however cannot be validated with quantitative methods. A general observation is that, the I value based on the area of partition units for each dataset (1, 7, and 8) is larger than the value based on the individual density, indicating a higher level of clustering for these datasets. It can be explained according to Equation 4 that, considering the length of curves reduces the differences of Y_i between neighboring units, and the Moran's I coefficient may reduce accordingly.

3) Patterns of polygons: The example datasets tested here are two building datasets, produced by TD Kadaster, the Netherlands. One is building features represented at scale 10K (dataset No. 2 shown in Figure 7d); the other is building features represented at 50k (dataset No. 9), interactively generalized from the dataset No. 2. The dataset No. 9 is shown in Figure 11. We assume here that the dataset at 50k is a good abstraction of the dataset at 10k, and hence the former is similar to the latter in terms of e.g. spatial distribution.

The proposed spatial autocorrelation model was computed on the two datasets, and the results are listed in Table III.

The results in the table indicate that, regardless of the

Figure 11. Dataset No. 9: Buildings at 50k generalized from dataset No. 2 (Source: Kadaster, the Netherlands)

Table III
THE I VALUES FOR THE SAME SETTLEMENT AREA (POLYGONS) REPRESENTED AT TWO SCALES

Polygon patterns			
Dataset No. (map scale)		2 (10k)	9 (50k)
I	density	0.33	0.55
	area	0.52	0.51

different measures used for computing the variable Y, both datasets are to some extent clustered. This is coherent to our perception. The clustering centers for the two datasets are round the north-east part.

The I coefficient values based on the area of partition units show that the degrees to which the buildings are clustered for both datasets are almost the same. The values based on individual density on the other hand, show that the generalized dataset (No. 9) is much more clustered than the base dataset (No. 2). This is mainly due to the fact that during generalization, smaller buildings in the north-east were enlarged to be more readable, and thus the densities increase considerably. Meanwhile, the bigger buildings in the south-west were left unchanged, so were the density values. This makes the cluster center in the dataset after generalization a much denser area, and hence increase the Moran'I coefficient.

V. DISCUSSION

First and foremost, the experimental results show that the proposed model extends the original capability of a well-studied spatial autocorrelation method (i.e. Moran's I), so that it is able to characterize the spatial patterns for points, curves (e.g. transportation networks) or disjoint polygons (e.g. buildings) in a uniform manner. Most of the existing quantitative spatial autocorrelation indicators are not able to handle the latter two feature types. Specifically, the proposed method is able to distinguish precisely and objectively between dispersed and clustered patterns for all the geometry types, providing insight to facilitate the analysis of many spatial processes (e.g. transportation infrastructure or urbanization progress). In addition, the results in the previous section show that the proposed Voronoi-like partition overcomes some of the deficiencies reviewed in the Introduction. Because the Voronoi-like partition forms a native and unique subdivision of the geometric space regardless of map scales, there is no need to decide an optimal unit for the partition as in the case of quadrat method or in calculating spatial autocorrelation for raster data [7], [8]. Another advantage is that the computational model can be readily equipped with Geary's C statistic instead of Moran's I, in applications where one concerns more sensitively about local deviations [3]. Since from the computational view, the elements for the two indices are almost the same [3].

While the use of different measures (i.e. *area* and *individual density* of partition units) for the variable Y in the I coefficient affects the calculated I values, especially for curves and polygons, we consider both measures reasonable as the difference lies in how people view the spatial patterns. In the case of *area* measure, the spatial distribution is regarded as a property that concerns mainly about locations, while *individual density* views the spatial distribution as a combination of the shapes of objects and their locations.

The proposed method is not sufficient and necessary for comparing the similarity of different distribution in map generalization, although the result for evaluating the polygon patterns appears that it is more appropriate to use *area* measures for comparing the similarity between different polygon distributions. That is, while two similar distributions may have similar I values (e.g. the I values obtained in Table III using *area* measure), similar values does not necessarily follow that these two distributions are similar (e.g. the I values for dataset No. 5 and No. 6 in Table I are the same, but the two distributions are totally different). Unfortunately, none of the existing approaches satisfy this requirement, and therefore, further research into this area is still needed.

In the context of charactering spatial patterns, the approach can be extended in many ways. In addition to area and individual density measures, one can use the main directions of building polygons to define the variable Y, and then follow the proposed method to characterize orientation patterns. This pattern is of great importance for map generalization in urban areas [5]. Another extension is to improve the weighting function to be more comprehensive. Since the higher order neighbor can be defined on the Voronoi-like structure [16], more weighting functions can be defined according to the discussion in Section II. Despite the descriptive ability, i.e. to characterize the spatial dispersion, the proposed model can be used to identify clusters (in transportation networks) and denser areas (in urban districts) by user specified density thresholds. This is highly useful in e.g. the urban planning and crime analysis.

From a statistical point of view, this current work is still an empirical study illustrating how the underlying idea works and its potential applications. No justified explanations why Moran's I can be much larger than zero even for a perfectly regular point pattern is available at this stage. A possibility is due to the edge effects which can be easily accounted for by excluding the boundary units in most situations. However, this will leave the Moron's I undefined for those perfectly regular point patterns, since consequently all the remaining Voronoi-like cells in such a pattern will have exactly the same area or density and hence the denominator of Equation 1 will be zero. Besides, current results only show the ordinary relationship between different patterns in terms of clustering degree. Further work needs to be done to clarify which values of this extend Moran's I should be considered as indicating clustering in contrast to regularity.

VI. Conclusion

The paper proposes and implements a Voronoi-like model of spatial autocorrelation for characterizing spatial patterns in vector data, including points, curves or networks, and polygons. In this work, Morans I coefficient, a well-known spatial autocorrelation statistic in geography, is extended with this new model to be able to calculate spatial patterns for various geometry types in a uniform manner. The experimental results show that this Voronoi-like partition is successful in characterizing spatial patterns in terms of how dispersed these various geometries are distributed in space. That is, a higher value indicates a higher level of clustering, and vice verse. At the end of the paper, strengths and limitations of the approach are discussed; applications and further investigations are suggested.

Acknowledgment

TD Kadaster who provides data for this research is acknowledged. We are also grateful to the three anonymous reviewers for their critical comments on an early version of this paper.

References

[1] Barry N. Boots and Arthur Getis. *Point pattern analysis*, volume 8 of *Scientific Geography series*. SAGA Publications, 1988. ISBN: 0-8039-2588-3.

[2] Ronald E. Wilson and Katie M. Filbert. Crime mapping and analysis. In Shashi Shekhar and Hui Xiong, editors, *Encyclopedia of GIS*, pages 180–186. Springer, 2008.

[3] Chandana Gangodagamage, Xiaobo Zhou, and Henry Lin. Spatial autocorrelation. In Shashi Shekhar and Hui Xiong, editors, *Encyclopedia of GIS*, pages 32–37. Springer, 2008.

[4] D.A. Moore and T.E. Carpenter. Spatial analytical methods and geographic information systems: Use in health research and epidemiology. *Epidemiol. Rev.*, 21(2):143–161, 1999.

[5] S. Steiniger and R. Weibel. Relations among map objects in cartographic generalization. *Cartography and Geographic Information Science*, 34(3):175–197, 2007.

[6] Yue-Hong Chou. Critical issues in the evaluation of spatial autocorrelation. In *Spatial Information Theory A Theoretical Basis for GIS*, volume 716 of *LNCS*, pages 421–433. Springer Berlin / Heidelberg, 1993.

[7] Yue-Hong Chou. Spatial pattern and spatial autocorrelation. In *Spatial Information Theory A Theoretical Basis for GIS*, volume 988 of *LNCS*, pages 365–376. Springer Berlin / Heidelberg, 1995.

[8] Ashton Shortridge. Practical limits of Moran's autocorrelation index for raster class maps. *Computers, Environment and Urban Systems*, 31:362–371, 2007.

[9] Michael John De Smith, Michael F. Goodchild, and Paul Longley. *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Troubador Publishing Ltd, 2006.

[10] Peter Rogerson. *Statistial Methods for Geography: A Student Guide*. Pine Forge Press, 2 edition, 2006. ISBN: 1412907950.

[11] Radu Stoica, Xavier Descombes, and Josiane Zerubia. A gibbs point process for road extraction from remotely sensed images. *International Journal of Computer Vision*, 57(2):121–136, 2004.

[12] M.I. Karavelas. A robust and efficient implementation for the segment Voronoi diagram. In *International Symposium on Voronoi Diagrams in Science and Engineering (VD2004)*, pages 51–62, 2004.

[13] Laurent Rineau. 2D conforming triangulations and meshes. In CGAL Editorial Board, editor, *CGAL User and Reference Manual*. 3.4 edition, 2008.

[14] Max J. Egenhofer. Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2):161–174, 1991.

[15] Paul Morrison and Ju Jia Zou. Achieving a more accurate skeleton through the refinement of skeletonization triangles. In *Proceedings of the Digital Image Computing: Technqiues and Applications (DICTA ' 05)*, pages 337–342, Dec. 2005.

[16] Tinghua Ai and Peter van Oosterom. Displacement methods based on field analysis. In *Joint Workshop on Multi-Scale Representations of Spatial Data, IAPRS XXXIV(2)*, page 10 p., 2002.