SPATIAL GENERALIZATION: AN ADAPTIVE LATTICE MODEL BASED ON SPATIAL RESOLUTION

P. Wang [a,*], T. Doihara [a], W. Lu [a]
[a] Research Institute, Asia Air Survey, Tamura 8-10, Atsugi, Kanagawa, JAPAN
(pt.wang, ta.Doihara, luwei)@ajiko.co.jp

Commission IV, Working Group IV/3

KEY WORDS: Generalization, Clustering, Aggregation, GIS - Geographic Data, Simplification, Adaptive Lattice Model

ABSTRACT:

The paper presents an Adaptive Lattice Model (ALM) based on spatial constraints or resolution such as minimum object size and minimum spatial interval. An ALM represents GIS data by latticing adaptively the spatial space at a certain abstraction level or spatial resolution and omitting all details smaller or finer than the resolution. One implementation of ALM is called CELL, which considers each lattice as a cell whose characteristics are statistically measured as a whole. And another one is called GRID, which moves all vertexes of the objects to the nearest point or one of the corners of lattices in ALM. This paper concentrates on the CELL based generalization including adaptive clustering, adaptive aggregation, and adaptive simplification. Adaptive clustering uses the target resolution to decide which features of the same type can be clustered or not. Adaptive aggregation, following the adaptive clustering, aggregates features in a cluster to form a new feature with the same type. And, adaptive simplification merges adjoining vertexes and deletes small loops and kickbacks by rounding up all vertexes of linear features in the input dataset. Experimental results are also presented to verify the effectiveness of the proposed model.

1. INTRODUCTION

A geographic dataset is a resolution-dependant representation of the original data, and generalization of geographic data means changing the representation resolution. In other words, some features or their attributes in the original dataset may be hidden, aggregated, simplified, or deleted, depending on the designated or target resolution. Although there are three types of resolutions, including spatial resolution, thematic resolution, and temporal resolution (Peng 2000), only spatial resolution will be discussed in this paper.

Map generalization usually involves a great deal of analysis of the geographic data for deciding what and how to generalize, and how to resolve conflicts that might occur during the generalization process. Manual generalization is a labor-intensive process and depends mainly on operator's knowledge, experience and skills. Recently, there many researches regarding general problems about generalization (Johnston 1999; Peng 2000) and concrete operations of generalization (Kreveld 1998; Sester 2000) for building automated generalizing algorithms to shorten generalization time and relieve operators from hard decision-making and repeated operations. For geographic dataset mainly consisted of vectors, polyline generalization is one of the most widely studied problems in the cartographic literature (Johnston 1999). An effective polyline generalization method, called Douglas-Peucker-Ramer algorithm (DPR) and proposed independently by Ramer (1972) and Douglas and Peucker (1973), is widely used. DPR creates a new generalized polyline by calculating the perpendicular distance between any vertex and the result polyline, and comparing the distance with a predefined threshold. DPR is based on the geometric distortion of every polyline in a geographic dataset.

This paper proposes an Adaptive Lattice Model (ALM) based on spatial constraints or resolution, which focuses on spatial generalization including clustering, aggregation, and simplification of geographic data such as roads, rivers and buildings, etc. An ALM represents geographic data by latticing adaptively the spatial space at a certain abstraction level or spatial resolution and omitting all details smaller or finer than the resolution.

This paper is organized as follows: after the explanation of ALM in chapter 2, adaptive clustering, adaptive aggregation, and adaptive simplification are discussed in chapter 3, followed by some experimental results in chapter 4. Finally, a summary concludes the paper in chapter 5.

2. ADAPTIVE LATTICE MODEL

In this chapter, we shall introduce Adaptive Lattice Model from the imaging principal by CCD camera, and two types implementation of ALM, GRID and CELL, consequently.

2.1 From CCD to Adaptive Lattice Model

A geographic dataset is a kind of representation of the real world, and its generalization is nothing more than changing its representation resolution. Although geographic data may be represented on maps with a range of different scales, the validity of observations or inferences drawn from such maps depends upon the resolution of the representation and the observer's resolution. Considering the representation of a geographic dataset as a imaging system with a CCD camera, what the observer can see are pixels arranged in two-dimension, where each pixel corresponds to and represents statistically a region of the geographic data. Figure 1 shows a typical example. Here, the left graph is the original data latticed with a higher resolution (smaller interval), and the right one shows the same data latticed with a lower resolution (larger interval). Considering each rectangle as a cell of a CCD, an observer cannot identify the details within a rectangle. That is to say, an observer can only take a cell as the overall representation of the corresponding region of the geographic data. Similar to the intensity of cells of the CCD device, the overall representation may be whether there are any vertexes in the cell, or whether there are any segments crossing the cell.

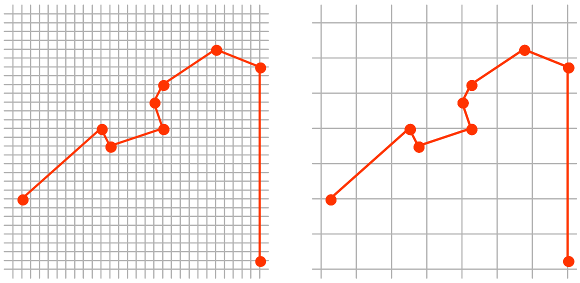

Figure 1. Lattice model for generalization

From the similarity between a CCD camera and the latticed geographic data, we define Adaptive Lattice Model (ALM) as follows: when a geographic dataset is confined to a predefined lattices where every edge's length of cells is equal to the target resolution (or spatial interval), a generalized result of the geographic dataset will be obtained by calculating the overall representation of the corresponding region of the data for every cell. For example, we can check if there are any vertexes of features in a rectangle, or calculate the overlapping area's percentage between a cell and a face feature in the geographic dataset. Different calculating methods may be used to generalize various types of geographic data. Here, two kinds of implementations of ALM are discussed below.

2.2 Implementation of ALM: GRID

In ALM, we can measure a cell by counting the number of vertexes in the cell. All vertexes in the same cell cannot be distinguished from each other, because the distance between any two of those vertexes is smaller than the target resolution, and should be represented by only one vertex in accordance with the principle of generalization. To merge all vertexes in the same cell to one point, moving those vertexes to the centre of the cell is a straightforward solution. In this way, the coordinates of all vertexes of features in the input dataset are rounded up with the cell's size or target resolution, similar to positioning graphs onto grids in a word processor. For this reason, the implementation of ALM mentioned here is called GRID. Since the translation of vertexes is within the cell, the statistic characteristics of the cell are not changed.

With GRID, the generalization process can be simplified to rounding up all vertexes' coordinates to the integral multiple of the target resolution and analysing the relationship among all the resulted vertexes. For example, a pair of adjacent vertexes will be merged if their rounded coordinates are the same as each other; a linear feature may be split if a middle vertex of the feature is merged with a vertex of another feature within the rounding procedure; etc. The algorithms based on GRID are discussed in Doihara 2002.

2.3 Implementation of ALM: CELL

In comparison to merging all vertexes in the same cell by GRID implementation, calculating the overlapping ratio between a cell and the portions of features that fall within the cell is another representation of the cell. GRID considers only vertexes of features and is effective for simplifying linear features. On the other hand, overlapping ratio should be defined as the percentage of the overlapping area between the cell and the portion of the features contained in the cell's area. That is to say, the representation of overlapping ratio may be used to generalize face features, and is called CELL to be distinguished with GRID. CELL can be used to aggregate face features by analysing the relationship of adjoined features. More details are discussed in the following chapter.

3. GENERALIZATION BASED ON ALM

Automated generalization usually focuses on geometrical features such as buildings, rivers, roads, etc. These features are commonly represented by polylines and/or polygons. Figure 2 shows a processing flow for generalizing geographic data by using ALM. Firstly, Selection is used to select features to be generalized. Continuously, we have two processing routes which are designed for selected areal and linear features separately. One is simplification based on GRID for generalizing linear features, which is discussed in another paper (Doihara 2002). Another is designed for generalizing areal features, including clustering and aggregation. We shall show the details of every procedure below.

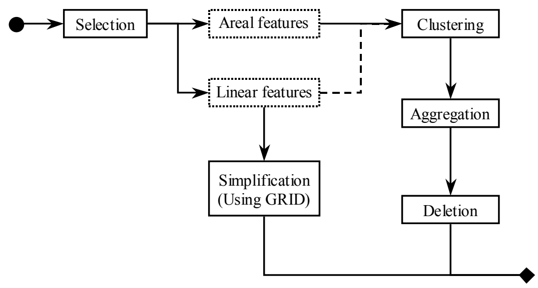

Figure 2. Generalizing Flow with ALM

3.1 Selection

Selection procedure is introduced to filter features to be generalized or not. With the decrease of the resolution, selection procedure ignores features to be omitted and reserves features to be generalized. Selection procedure can also be used to restrict the features to be generalized at the same time for avoiding erroneous data. For example, roads and buildings may be generalized simultaneously, because roads cannot cross with buildings at any target resolution. On the other hand, land boundaries may cross with some roads. In this case, selection procedure let the roads and buildings be generalized in the first phase, and the land boundaries at the next phase. Selection is commonly a manual procedure or a procedure based on predefined database, and will not be discussed more in this paper.

3.2 Adaptive Clustering

Adaptive clustering uses the minimum spatial interval or target resolution to decide which features of the same type can be clustered or not. When linear and face features are selected by the selection procedure, and need to be generalized, the clustering procedure is shown in figure 3. Here, linear features are used as constraints for separating face features compulsorily when necessary, e.g. when two buildings located at the different sides of a main road cannot be merged together. And face features are the targets of clustering, which results in lists of face feature. Some details of figure 3 are given below.

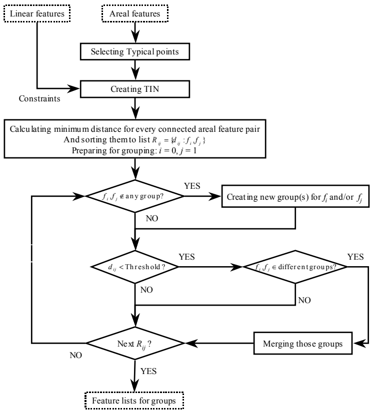

Figure 3. Adaptive clustering diagram

a) Selecting representative points for all face features includes several instances. One is using the centre point of a face feature. The next one is treating all vertexes of the outline as a face feature's representative points. The third one is to split a face feature to Triangulated Irregular Network (TIN) and use the centres of all triangles as the feature's representative points.
b) When creating TIN from all representative points, the outlines of face features and all linear features are used as the constraints or break lines to avoid unfavorable clustering.
c) A link is an edge of a triangle in TIN, and must connect two different face features. If there are multiple links between two face features, only one is reserved for later process and others are ignored, when calculating the minimum distances for all links. Links connecting the same features are ignored in this implementation.
d) After sorting all links with their minimum distance, the shortest link should be at the front, and the longest one at the rear. The clustering procedure must begin from the first (shortest) link.

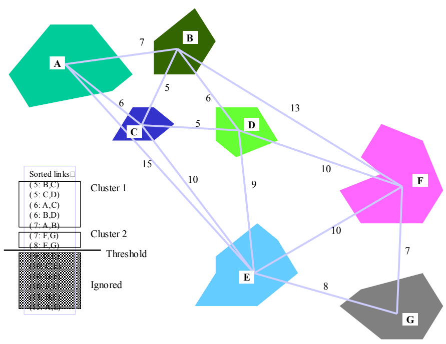

Figure 4. Adaptive clustering image

Figure 4 illustrates an example of adaptive clustering. Here, the centres of face features are used as their representative points; the numbers near the links are minimum distances respectively. Sorted links and the grouped results are listed at the lower left part of the figure. Here, the threshold used to tell whether two face features belong to the same cluster or not is 8.5. Of course, all links which minimum distances are larger than the threshold can be ignored in the clustering procedure.

Adaptive clustering may be used for generalizing buildings, lakes, or other face features.

3.3 Adaptive Aggregation

Adaptive aggregation, following the adaptive clustering procedure mentioned above, aggregates face features in a cluster to form a new face feature with the same type. Figure 5 shows a diagram of adaptive aggregation.

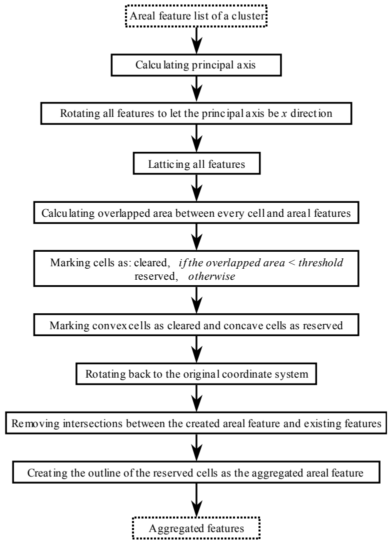

Figure 5. Aggregating diagram

In figure 5, the input data are the face feature lists obtained by clustering procedure mentioned above, and the results are aggregated face features. Some steps of the aggregation procedure in figure 5 are described below.

a) Defining the principal axis of a cluster as the maximum likelihood direction of the cluster. In other words, most segments of features in a cluster are parallel to the principal axis. The principal axis is calculated as follows: first define a histogram of direction from 0 to 180 degrees; then calculate the directions of all segments and vote the obtained direction with the segment's length consequently; lastly find the direction with the peak value in the voted histogram.
b) Rotating all features to let the x-axis parallel to the principal axis. This makes most segments of the features in the cluster being parallel or perpendicular to the x-axis.
c) Removing intersections between created face features and existing features is used to check if there are any intersections and separate them if they are crossed.
d) Creating the outline of all reserved cells as the aggregated face feature was accompanied by representing stair-like cells with slant segments to decrease the vertexes of the aggregated result. Furthermore, holes may be created if there are any cleared cells in the resulted face feature.

3.4 Adaptive Simplification

Adaptive simplification is completed by rounding up all vertexes of linear features in ALM. Figure 6 shows the diagram of adaptive simplification, including elimination, separation, collapse, division, and connection, etc. Adaptive simplification is described in detail in Doihara 2002.

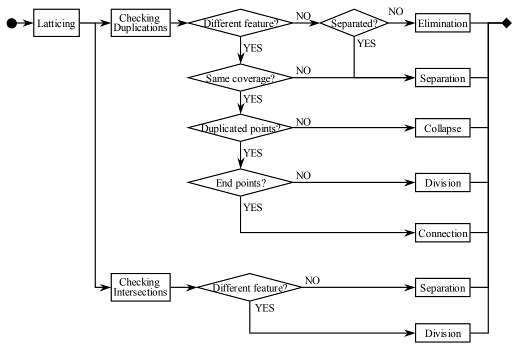

Figure 6. Adaptive simplification

4. EXPERIMENTS

We have implemented a prototype system for generalization of geographic data based on the proposed ALM. Figure 7 shows the main window of the prototype system. This prototype system is still under development and runs on Microsoft windows without using any other applications.

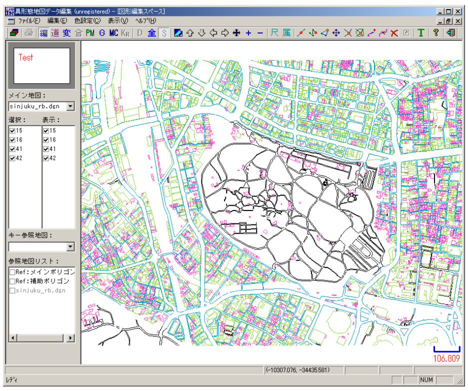

Figure 7. Running window

Figure 8 shows an experimental result, where the input geographic data includes buildings and roads. The statistics about the experiment are listed in Table 1.

Figure 8. Generalized result: Original buildings and roads (up: 1/2500) and aggregated buildings and simplified roads (down: 1/25000)

Features / Points	Generalization	
	Before	After
Buildings (polygons)	5255	1183
Roads (polylines)	8016	7119
Points (roads+buildings)	65922	32034

Table 1. Margin settings for A4 size paper

5. CONCLUSION

This paper has proposed an Adaptive Lattice Model to generalize geographic data. Generalization decreases the resolution of a geographic dataset, and causes indistinguishableness of the original dataset, which is similar to the imaging principle of a CCD camera. The proposed ALM represents GIS data by latticing adaptively the spatial space at a certain abstraction level or spatial resolution and omitting all details smaller or finer than the resolution. The generalization based on ALM mainly includes selection, adaptive clustering, adaptive aggregation and adaptive simplification, etc. A prototype system has also been constructed based on the proposed model, and experimental results show the effectiveness of the proposed model.

ALM is still under development. Both precise aggregation and improvements for exterminating the intersection between generalized features still remain as our future works.

REFERENCES

Doihara T., 2002. An adaptive lattice model and its application on map simplification. In: *Joint International Symposium on Geospatial Theory, Processing and Application*, Ottawa, Canada, ISPRS Vol.34 ().

Douglas, D.H. and T.K. Peucker , 1973. Algorithms for the reduction of the number of points required to represent a digitised line or its caricature. *Canadian Cartographer*, 10(2), pp.112-122.

Johnston, M.R., Scott, C.D., and Gibb, R.G. 1999. Some problems arising from a simple GIS generalisation algorithm. In: *Proceedings of the 11th Annual Colloquium of the Spatial Information*, Research Centre, University of Otago, Dunedin, New Zealand, pp.191- 199.

Kreveld, M. and J. Peschier 1998. On the automated generalization of road network maps. http://divcom.otago.ac.nz/SIRC/GeoComp/GeoComp98/21/gc_21.htm (accessed 12 Mar. 2002)

Peng W., 2000. Database generalization: conflicts, problems, and operations. In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, Holland, ISPRS Vol. XXXIII, Part B4, pp.826—833.

Rumer, U. 1972. An iterative procedure fro the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1, pp.244-256.

Sester, M. 2000, Generalization Based on Least Squares Adjustment, In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, Holland, Vol. XXXIII, Part B4, pp.931—938.