AN AUTOMATIC GENERALIZATION METHOD OF SYMBOLIZED POINT

Clusters and Its Application on Web Feature Service

Ma Zhaoting, Sun Wei, Li Chengming, Wang Jizhou
(Chinese Academy of Surveying and Mapping, Beijing 100039, China)

Abstract:
Web Feature Service (WFS) is an open standard of data interoperation and is commonly used in web spatial data service. For point clusters, such as distribution of street lamps, shopfronts or sewage lids, in municipal management application, raw data without symbols is not intuitionistic to users. When the point clusters are excessive and dense, the efficiency of transporting and rendering them on the web client is debased; the behavior of representation is also spoiled since many points overlapping with each other. To solve the bottleneck of WFS in data transportation and rendering, an automatic generalization method of symbolized point clusters is put forward in this paper and the generalizing results are cached in the data server. The automatic generalization method is based on the size of the symbols, the density of the point clusters and the scale of the map. By adjusting the generalization ratio, the almost overlapping of each two points at some a scale is avoided. With the tiled caching of generalization result, the generalization result is reused and the request of WFS is only affected by bandwidth and web server. Practice shows that the efficiency and representation of WFS with generalized points is grateful.

Keywords:
Point clusters, Symbolic, Generalization, WFS, Cache

Introduction

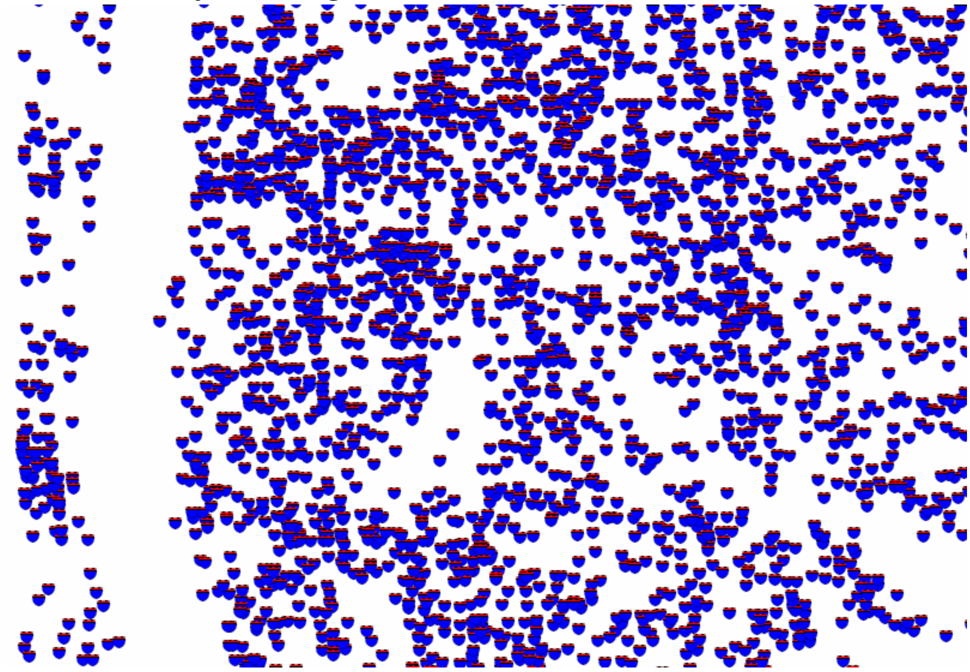
Web Feature Service (WFS) [1] is an open standard of data interoperation defined and maintained by Open Geospatial Consortium(OGC). The Web Feature Service Interface Standard provides an interface allowing requests for geographical features across the web using platform-independent calls. Since the Web Map Service (WMS) [2] interface or online mapping portals like Google Maps[3] return only an image, which end-users cannot edit or spatially analyze. The XML-based GML(Geography Markup Language)[4] furnishes the default payload-encoding for transporting the geographic features, but other formats like shapefiles can also serve for transport. WFS is commonly used in web spatial data service.

There are numerous commercial and open source implementations of the WFS interface standard, including some open source reference implementation, such as GeoServer[5], MapServer[6] and so on. The WFS specification defines interfaces for describing data manipulation operations of geographic features, such as discoverying, querying and retrieval of features. A transactional Web Feature Service (WFS-T) allows creation, deletion, and updating of features. The client generates the request and posts it to a web feature server using HTTP. The web feature server then executes the request. The WFS specification uses HTTP as the distributed computing platform, although this is not a hard requirement.

GML is an XML dialect which can be used to model geographic features. For rendering vector data in IE browser, some need ActiveXs or plugins, such as VML, SVG and SWF. Other do not need such as GML, which renders vector use divs with html and JavaScript. Since the efficiency of rendering GML is low, large geospatial data showing will spend a lot of time in browser and cause the browser stop working.

Background

In our project, all municipal facilities in one city would serve as WFS and be requested by different departments through the Internet. In municipal management application, raw data without symbols is not intuitionistic to users. So we provide standard symbols to these data according to mapping standards. But when the data is dense, most symbols are overlapping with each other which brings burden to visualization and also causes to slowly rendering in IE browser.



The overlapping of symbolized point set

To promote the efficiency of transportation and rendering points set, I want to generate the points set and extract points according to the scale, the symbol size and the distance between each two points. Nowadays there are a lot of paper emphases on the data set of line and surface [7-9]. A little paper about points set generation method considered mostly on the tendency but not the symbol size or the dynamic scale of the map [10-11].

Solution

Since too much points rendering are not visible for users, why not consider extracting some to representing. In our project, all massive geospatial data are point cluster, so our solution is focus on point data.

The solution is emphasis on extracting some points which can represent original points. Since raw data without symbols is not intuitionistic to users, most point data are rendering with different symbols according to mapping standards and industry rules. The extracting is limited by threes factor as following:

1) The scale of the map
2) The density of the points
3) The size of symbols

In our project, we defined a generation rule with overlapping ratio, which means how much overlapping cannot express in the map. The ratio can be adjusted by the users according a certain data. After data generating, one or more points in the map may be represented by one point. To express the tendency of the data, the color of the point symbol is changed automatically according the number of point it represented and the number will show as a tip when the mouse move over the point symbol.

For one scale of S, the distance (D) of two points with symbols of size s, the overlapping ratio(R) can be calculated with the following formula:

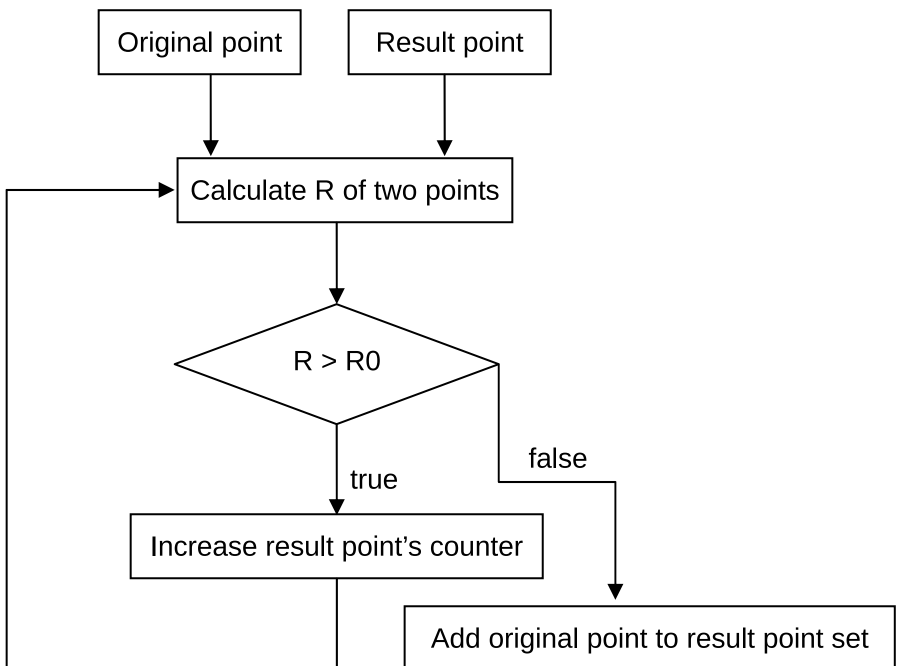
R=0, if s <= D/S

R=D/S/s, if s > D/S

Define a criterion of generating ration of R0, then when R is larger than R0, one of the two points is avoided and the counter of another point is increased. For all points in the vector data, the generation procession is descripting as follow:

1) Add the first original point to the result point set.
2) From the second original point, calculate the distance of the original point to a result point and the overlapping ratio(R) of certain scale with given symbol size.
3) If R is larger than R0, the second point is avoided and the first point’s counter is increased.
4) If R is equal to or less than R0, the second point is added the result points.
5) When the number of result points is larger than one, original point will compare with all the result points and only all of them two points satisfy R less than R0, the original point can be added to the result point set . Otherwise the original point is avoided and the first result point’s counter is increased.
6) Buffer the result points in GML format.

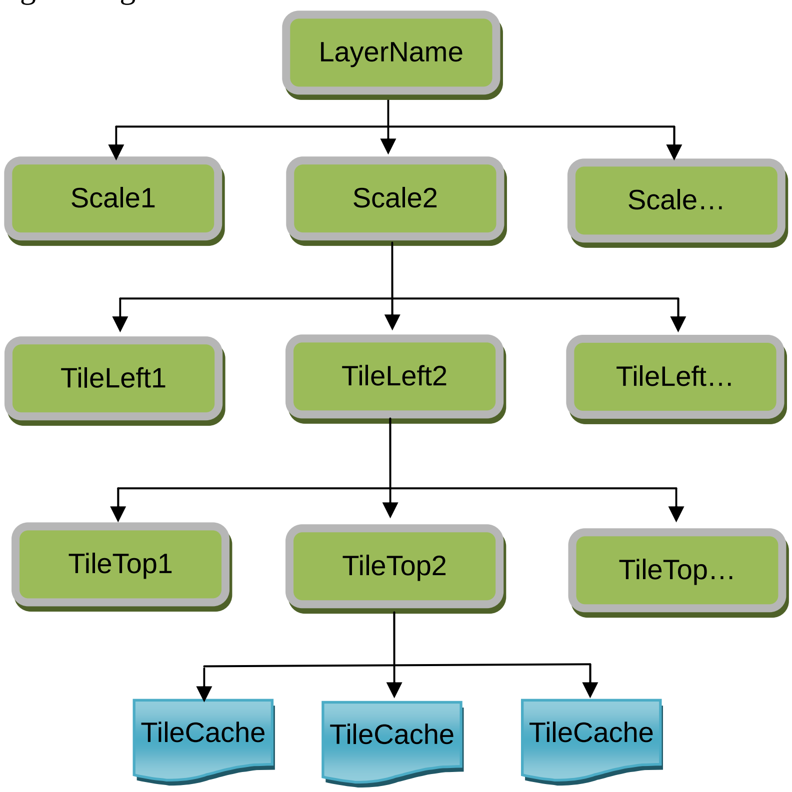
All the step of generating can be expressed in the following procedure:



Procedure of auto generation method of each scale

This procession is calculated in the server and can be buffered for future using. For the server buffering and client requesting efficiency, tiled extracting and buffering is needed. After tiled buffering, end users can request some tiled buffer in the current viewport, which is important when the scale is larger. In our project, tile schema is defined for easily using for client user. The tile size is 256 in screen unit, that is pixel and for avoiding the overlapping of neighboring tiles, a meta-tile of 6 tiles in wide and 6 tiles in height is adopted in the auto generation procedure. Tiles are named with the map coordinate in horizontal and vertical of a certain scale.

Tiled buffer structure in server machines is described as follow. The buffer may be created in advanced or be created in the first time of requesting. In the latter, when end users set WFS request according to the tiled rule to the server, tile cache is checked and created tile buffer if it doesn't exist. Otherwise the tiled buffer is send to the client users directly from the cache server. Only tiles in the viewport are requested and transported to the clients. Points in every tile are generated under the extracting rule with the overlapping ratio. Almost all points do not break the rule expect ones on the borders of neighboring meta-tiles.



The tiled buffer structure in the server

Result

Based on the method described above, our team developed software to extract typical points from original ones automatically. In the software we made a series practices with different points set of standard symbols. The following table shows the relationship between the overlapping ratio and the result points in full view status.

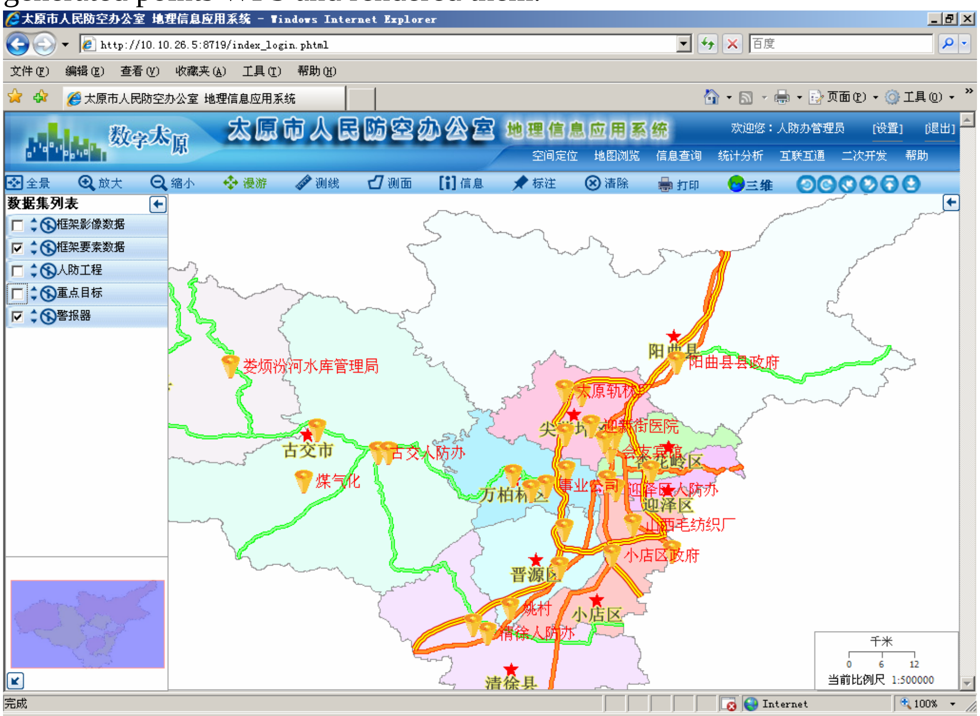
Points set	Origin point	R0=75%	R0=50%	R0=25%
alarms	*298*	*231*	*159*	103
monitor	*1390*	*857*	*634*	433
street lamps	*21476*	*17945*	*12433*	854
Save ratio	**23164**	**82.2%**	**57.1%**	**6.0%**

Practices show that the R0 is the key factor which controls the saving ratio. From the visual effect and the saving ratio, 50% overlapping is a good balance of most points set.

The time of requesting and rendering WFS points of original points, without buffer and tiled buffer are compared in the following table. We can see that original points without any generation spent most time of requesting and rendering since all the points are transported to the client in every scale. When the scale is larger, time is saved for only points in the view are rendered. Real time generation without buffer spent most time in extracting points and can be used in larger scale. Tiled buffer spent the lowest time in every scale since the saving of extracting, transporting and rendering. The responding time is not change dramatically in every scale. Only tiles in the screen viewport are requested and rendered, so the points data is transported and rendered smoothly and efficiently.

scale	Original points	Without buffer	Tiled buffer
1:50000	*103.2S*	101.4S	29.5S
1:10000	*94.3S*	90.1S	32.2S
1:2000	*79.5S*	49.3S	29.1S

The following picture is a real client page of end users which requested the tiled generated points WFS and rendered them.



An application of this generation method of points WFS

Conclusion

In our project, all municipal facilities in one city would serve as WFS and be requested by different departments through the Internet. The tiled automatic generalization method decreased the data follow of network transportation and the real-time responding burden of servers'. Moreover the generated points set save the rendering time in the client. Practice showed when the overlapping equals to 50% is the best balance between the visual effect and the time saving. This method can be popularizing to other points cluster generation.

Since the generation results are buffered in the server, the WFS provider should remove all buffer related when the WFS vector data update.

The generation method put forward in this paper is focus on points set and is only suited for points. In next step, we will research on auto generation methods of lines and surfaces in WFS.

Reference

[1] OpenGIS Web Feature Service (WFS) Implementation Specification, [EBOL], http://www.opengeospatial.org/standards/wfs

[2] OpenGIS Web Map Service (WMS) Implementation Specification, [EBOL], http://www.opengeospatial.org/standards/wms

[3] Google Maps, [EBOL], http://maps.google.com/

[4] OpenGIS Geography Markup Language (GML) Encoding Standard, [EBOL], http://www.opengeospatial.org/standards/gml

[5] GeoServer, [EBOL], http://geoserver.org/display/GEOS/Welcome

[6] MapServer, [EBOL], http://mapserver.org/

[7] Allouche M K, Moulin B. Amalgamation in Cartgraphic Generalization Using Kohonen's Feature Nets [J]. International Journal of Geographical Information Science, 2005, 19 (819): 899-914.

[8] Sester M. Optimization Approaches for Generalization and Data Abstraction [J]. International Journal of Geographical Information Science, 2005, 19 (8/9): 871-897

[9] Liu Songlin, Fan Yuru, Cheng Yi, Wu Zhengsheng, Delaunay Triangulation Applying in the Polygons Generalization and Punctuate Objects Generalization [J], Journal of Zhengzhou Institute of Surveying and Mapping, 2007, 24(4): 303-305.

[10] YUKIO S. Cluster Perception in the Distribution of Point Objects [J]. Cartographica, 1997, 34(1):49-61.

[11] Yi Lu, Jinghai Du, Jingsheng Zhai. A Model of Point Cluster Generalization with Spatial Distribution Features Recognized and Measured [A]. In: Proceedings of 20th International Cartographic Conference [C], Beijing: Publishing House of Surveying and Mapping, 2001: 2123-2128.