View- and Scale-Based Progressive Transmission of Vector Data

Padraig Corcoran[1], Peter Mooney[1],
Michela Bertolotto[2], and Adam Winstanley[1]

[1] Department of Computer Science,
National University of Ireland Maynooth
[2] School of Computer Science and Informatics,
University College Dublin
padraigc@cs.nuim.ie
http://www.cs.nuim.ie/~padraigc/

Abstract. Progressive transmission represents an effective means for the transmission of spatial data over the web. We classify current implementations of this paradigm as either view- or scale-based, which transmit data in a manner which is a function of the required map region or scale respectively. Although many such approaches exist, the concept of one which is both view- and scale-based has received little attention. This can be attributed to the difficulty in maintaining topological equivalence between the transmitted and original map. In this paper we propose a novel methodology for progressive transmission which overcomes this difficulty. Results demonstrate that significant reductions in the amount of data transmitted are achieved when compared to view- and scale-based approaches.

Keywords: View-Based, Scale-Based, Progressive Transmission, HTML5.

1 Introduction

It has been widely recognized that the Internet represents an important means to disseminate spatial data and that in the future more and more developments in the field of Geographical Information Science (GIS) will centre on the Internet [15,20]. Spatial data is generally transmitted over the Internet using either an image- or vector-based representation. A vector-based representation offers a number of benefits over an image-based representation [1]. These benefits include the facility for greater map interaction, spatial analysis and adaptive visualization. Despite these benefits, modern web-browsers offer greater support for the transmission, memory management, and visualization of image-based representations [16]. Consequently this is the representation used by all major mapping services such as Google Maps and OpenStreetMap. The work presented in this paper focuses on vector-based representations. Irrespective of which representation is used, as the size of spatial datasets increase, a trade-off emerges between

the requirements to deliver data of high detail and to deliver it within reasonable time [3]. One widely used approach to optimize this trade-off is progressive transmission. We claim that any transmission methodology may be classified as being progressive in nature if it exhibits both of the following properties: (1) Data is transmitted incrementally in the form of refinements which may be integrated and used by the client without the requirement for a complete dataset transmission. (2) Wherever possible, data previously transmitted is used to fulfill current client requests and is not re-transmitted. This reduces redundancy which exists if the same data is transmitted multiple times. Yang and Weibel [23] state that "The progressive transmission of vector data is beginning to receive more attention as it provides a promising solution for improved efficiency of data delivery in the web environment". Despite the potential benefits, the research underlying progressive transmission is not mature and is still being developed. Yang and Weibel also remark that progress in this research domain is "relatively slow". The work presented in this paper contributes to this research domain in the following way. We introduce the concepts of view- and scale-based approaches to progressive transmission and describe how their fusion can significantly reduce the amount of data transmitted. Ensuring topological equivalence between the transmitted and original map represents the greatest challenge in implementing such an approach [4]. Subsequently we propose an implementation of this paradigm which overcomes this difficulty.

The layout of this paper is as follows. Section 2 introduces the concepts of view- and scale-based approaches to progressive transmission. Section 3 presents an analysis of the benefits and limitations of existing approaches to progressive transmission which attempt to fuse view- and scale-based methodologies. A novel approach which overcomes these limitations is presented in section 4. Finally in sections 5 and 6 we present results and draw conclusions respectively.

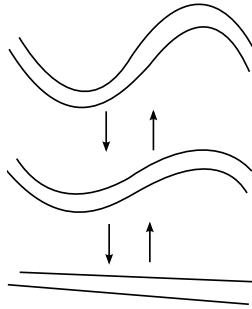## 2   View- and Scale-Based Progressive Transmission

Using the above definition of progressive transmission, we propose a classification of most current implementations as view- or scale-based approaches. We now discuss each of these in turn. View-based approaches progressively transmit data in a manner which is a function of the map region being viewed. Many tile based approaches to image- and vector-based transmission may be classified as view-based [19,1]. Tiles are transmitted progressively and once received may be used. Generally only those tiles contained within map region being viewed are transmitted. If this region changes by only a small amount only the refinements, in the form of new map tiles, required to compute the new region from the previous are transmitted. Given that a client will generally only view a small region of a larger map view-based progressive transmission approaches significantly reduce the amount of data transmitted as compared to transmitting the entire dataset. When a client requests data they have previously received at a different scale, view-based approaches show major inefficiencies. Instead of downloading only the map refinements required to transform the previously downloaded scale

into the new one requested, view-based approaches download the entire dataset corresponding to the required scale. The individual scales of the map in question are computed using map generalization, whose goal is to reduce detail subject to a set of objectives [12]. Topological properties are considered the most important feature of any map and therefore many researchers have developed generalization techniques which produce maps topologically equivalent to the original map [17].

Scale-based approaches progressively transmit data in a manner which is a function of the required map scale [3,23] and in general operate as follows. A progressive generalization of the map is precomputed to produce a series of maps. There exist three forms of generalization. These are object, model and cartographic generalization [21,7]. Object generalization takes place when defining the original database. Since databases represent abstractions of the real world, a degree of generalization must take place during data capture. Model generalization is concerned with data reduction. Cartographic generalization describes map generalization for the purpose of visualization [2]. Since the purpose of generalization in the context of progressive transmission is to reduce map data or detail, model based generalization approaches are used. In model-oriented generalisation, the focus is on the objects and their "importance". Therefore in general every level of generalisation broadly corresponds to a scale but this is not guaranteed. On the other hand cartographic generalisation generates visualizations at a given scale.

The resulting generalized maps are transmitted to the client in the opposite order to which they were created [10,24,22,11]. This transition between scales is achieved through a process of refinement which transmits only the additions required to compute the current scale from the previous. In essence, when using such an approach, the process of refinement is the inverse of the corresponding generalization [18]. By implication of the fact that the maps transmitted equal those produced through generalization, although in a reverse order, progressive transmission will satisfy the same objectives as the corresponding generalization process. For example if the generalization process produces a set of map scales which are all topologically equivalent to the original map then all transmitted maps will be topologically equivalent to the original map also. The concept of progressive transmission is illustrated using a sample data set which contains two line features and is represented by the top diagram of Figure 1. A generalization process, represented by arrows pointing down, consisting of two steps is applied to these features. This results in a map which contains two straight line features and is represented in the bottom diagram of Figure 1. This generalized map is the first representation transmitted to the client. Next, using a process of refinement, details removed during the generalization process are progressively transmitted and integrated by the client; this process is represented by arrows pointing up in Figure 1. Since only the refinements required to transform the current scale from the previous are transmitted, relative to transmitting the entire current scale, scale-based progressive transmission significantly reduce the amount of data transmitted. This methodology cannot be applied to an

image-based representation of spatial data because map generalization cannot be applied to such a representation. When the client requests only a small region of a larger map scale-based approaches introduce major inefficiencies. Instead of downloading only the map refinements corresponding the required map region, scale-based approaches download the set of map refinements corresponding to the complete map at the required scale. This is because the map refinement process is not adaptive and is constrained to be the inverse of the corresponding generalization.


Fig. 1. The arrows pointing down and up represent generalization and progressive transmission processes respectively

3   Existing Fusion Methodologies

As discussed in the previous section both view- and scale-based approaches to progressive transmission reduce the volume of transmitted data in different ways. Therefore to maximise the reduction in transmitted data volume concepts from both approaches must be fused. To the author's knowledge only three works attempting such a fusion exist. These are the works of Ramos et al. [16], Follin [8] and Li [13]. We now describe each of these approaches in turn and illustrate that one of the greatest challenges in implementing such a transmission strategy is maintaining topological equivalence with the original map. In fact Yang [23] identifies this challenge as the number one reason why progress in this research domain has been slow.

The methodology of Ramos et al. [16] operates as follows. The features contained within the client view are determined using a Quadtree data structure and these features alone are transferred progressively. No attempt is made to maintain topologically equivalence with the original map. Follin [8] proposes to transmit those features contained within a client view through the application of refinement techniques which are the inverse of generalization. It was demonstrated by Corcoran et al. [4] that such an approach does not ensure topological equivalence. Instead refinement techniques which are distinct from their generalization counterparts must be applied. Li [13] proposes an alternative methodology which does maintain topological equivalence but is subject to a number of constraints which we now discuss. Li proposes to divide the map

into a series of tiles and generalizes each tile individually. A constraint is added such that features which are not contained within a single tile are not generalized. This facilitates the integration of tiles of different scales. A generalization method which maintains topological equivalence with the original map is used. The resulting tiles are then transmitted to the client in a similar manner to existing scale-based transmission strategies described in section 2. The method of Li exhibits a number of disadvantages. Firstly, the failure to generalize features which are not contained within a single tile results in an inconsistent degree of generalization being applied across the map. Secondly, features not contained within a single tile must be segmented before transmission and subsequently merged by the client when received [1]. This concept is illustrated in Figure 2(a) where a single linear feature crosses two tiles. Finally if a feature spans a set of tiles and only a subset of these are transmitted, this can introduce topological inconsistencies in the clients dataset. For example consider Figure 2(b) where a single feature spans four tiles. If only the top row of tiles are transmitted, as represented by Figure 2(c), the feature in question will be represented as two unconnected features in the clients dataset. Antoniou et al. [1] propose to overcome this issue by altering the client data structure to encode the required connectivity between features.

(a)   (b)   (c)
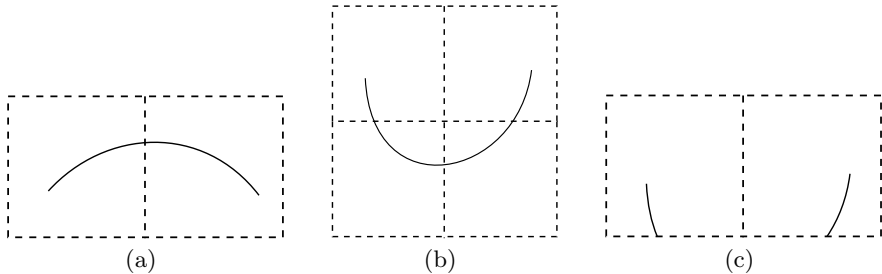
Fig. 2. A line feature (represented by a solid line) spans multiple map tiles (represented by dashed rectangles) in both (a) and (b). The transmission of only the top row of tiles in (b) results in the two unconnected features represented in (c).

Due to the fact that Li [13] uses a progressive transmission strategy which is the inverse of a generalization, a tile-based approach is necessary. For example consider the map containing two features represented by the top diagram of Figure 3(a). This map is generalized in two steps represented by arrows pointing down. If the client only views the feature contained in the right-hand-side of the map, this feature cannot be transmitted without also transmitting the feature in the left-hand-side of the map the feature. This is because such a transmission would not represent the inverse of the corresponding generalization. Instead both features must be transmitted and this is represented by arrows pointing up Figure 3(a). Next consider the case where the map is divided into two tiles and each is generalized individually as illustrated in Figure 3(b). Now if the client views only the feature on the right-hand-side, the server can determine

the tile containing this feature and transmit the inverse of the corresponding generalization as illustrated in Figure 3(c). The inverse of the generalization corresponding to the tile on the left-hand-side is not transmitted. In the following section we propose a novel fusion method which overcomes the limitations of the methodologies proposed by Ramos et al. [16] and Li [13].

(a)   (b)   (c)
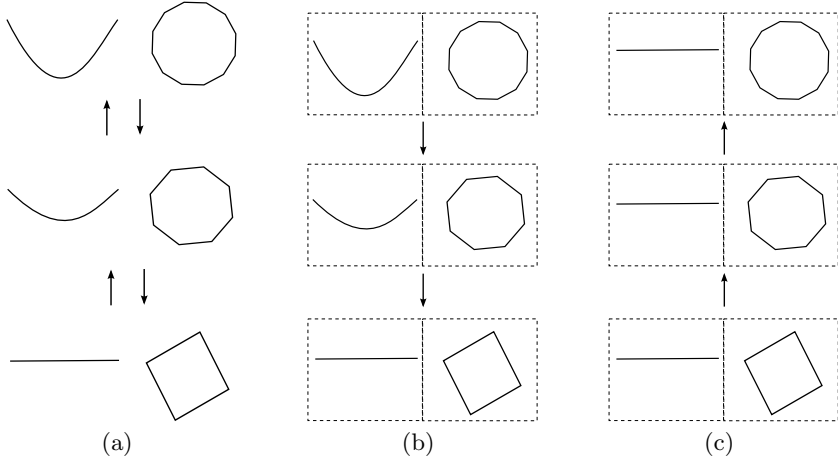
Fig. 3. Individual map tiles are represented by dashed rectangles

4   Proposed Fusion Methodology

This section describes the proposed methodology for fusing view- and scale-based approaches to progressive transmission. Initially the dataset is generalized using the approach presented in [6]. The generalization method in question is a simplification technique which iteratively removes feature vertices while maintaining topological equivalence to the original map. Depending on the size of the dataset this step may require a number of seconds to compute. Although this can be precomputed and therefore does not effect the user experience. Next all features are inserted into an R-tree data structure to allow fast future indexing [9]. Again this step is pre-computed.

The server maintains an up-to-date list of the features, and their corresponding level of generalization, which exist on the client. When the client views a particular region of the map a message is transmitted to the server which contains the Cartesian coordinates of the viewing window in question. On receipt of this message the server performs a window query using the R-tree to determine the set of features which lie within this window. The server then compares this set to the set of features currently stored by the client. This comparison allows the server to determine the required map refinements to be transmitted progressively to the client. This process is repeated each time the client changes their current viewing window and is represented in the UML sequence diagram of Figure 4.
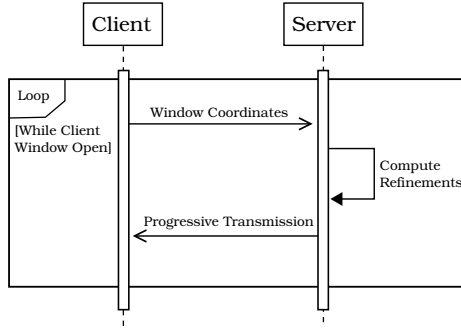
Fig. 4. UML Sequence Diagram

When the server determines which features lie within the clients current viewing window, it only transmits map refinements corresponding to those features. For example consider the sample dataset, containing two linear and one polygon feature, represented by the top diagram of Figure 5(a). This dataset is generalized in two steps represented by arrows pointing down in Figure 5(a). When the client viewing window, represented by a dashed rectangle, only contains the polygon feature on the right-hand-side, refinements corresponding to this feature alone are transmitted. Features which lay in the previous viewing window but not in the current one are determined by the server and subsequently deleted from the clients local dataset. These refinement steps are illustrated in Figure 5(b) by arrows pointing up. This refinement is not the inverse of a corresponding generalization and therefore the maintenance of topological equivalence to the original map cannot be assumed. In order to maintain topological equivalence the methodology of Corcoran et al. [4] is used. This methodology allows the refinement of a map which is not the inverse of generalization to be computed such that topological equivalence is maintained. The number of possible different sets of features which may require refinement will be large. That is, the client may view many different subsets of the map. It is therefore not practical to pre-compute the refinements of all these sets and they must be computed on the fly.

The proposed fusion methodology was implemented using the following technologies. The server used was Python simpleHTTPServer. However any HTTP server capable of running Python can be used. All server-side components were written in Python while all client-side components were written in JavaScript. Extensive use was made of many of the new API's in HTML5 [14]. The transmission of vector data along with generalization and refinement operations was performed using the HTML5 WebSocket API which defines a full-duplex communication channel between client and server. Client map visualization was performed using the HTML5 Canvas API. An overview of the benefits of using HTML5 for the transmission and visualization for spatial data can be found in [5].

(a)   (b)
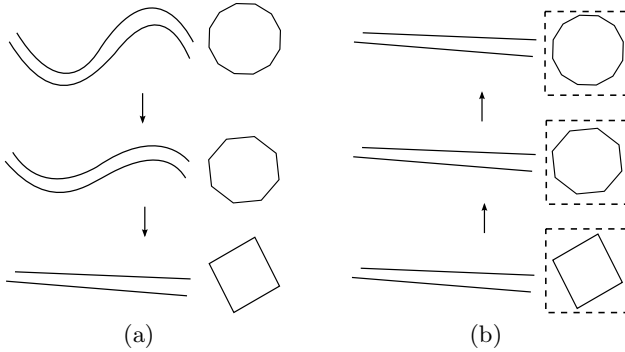
Fig. 5. The arrows pointing down and up represent generalization and progressive transmission processes respectively. The dashed rectangle in (b) represents the clients current viewing window.

5   Results

To demonstrate the effectiveness of the proposed fusion methodology the following approach was used. A number of large datasets were downloaded from OpenStreetMap (OSM). The data is downloaded initially in OSM XML format. One of the datasets which contains 13,755 polygon and line vertices is displayed in Figure 6. A small region of each map was chosen. The region corresponding to the map of Figure 6 is shown in Figure 7(d). Each region was then downloaded progressively using the following three methodologies. Firstly the region was downloaded using a scale-based approach to transmission which is the inverse of the corresponding generalization. In the case of downloading the region of Figure 7(d) this would correspond to downloading the complete dataset and not just the required region. Next the region was downloaded using a view-based approach where the features in the region are downloaded completely for 5 different scales where each scale represents an individual generalization of the dataset. In the case of downloading the region of Figure 7(d), this would correspond to downloading the region completely for each scale and not just the refinements required to compute the current scale from the previous. Finally the region was downloaded using the proposed fusion methodology. Three intermediate stages in the transmission of this region using the propose fusion methodology are displayed in Figure 7(a)-(c). The convergence of this transmission is equal to Figure 7(d). Since view-based approaches pre-compute all map scales, upon receipt of a user request computation is required before transmission. Similarly since view-based approaches pre-compute generalization, upon receipt of a user request the inverse of this is transmitted without any computation before transmission. On the other hand, given that the proposed fusion-based transmission methodology is adaptive and not the inverse of a generalization process, it cannot be pre-computed. This results in a correspond short delay before transmission is initiated. A formal analysis of the computational complexity of the proposed method for computing topological invariant refinements can be found in [4].

Table 1. The first column displays the number of vertices in each dataset. Each of the remaining columns displays the percentage of dataset transmitted in order to send a required small region of the map.

| Map Vertices / Method	Scale-based	View-based	Fusion	Fusion Time
13,755	100%	23%	7%	12 Sec.
15,223	100%	24%	9%	20 Sec.
11,918	100%	19%	5%	8 Sec.
13,142	100%	22%	6%	15 Sec.
14,142	100%	14%	4%	13 Sec.
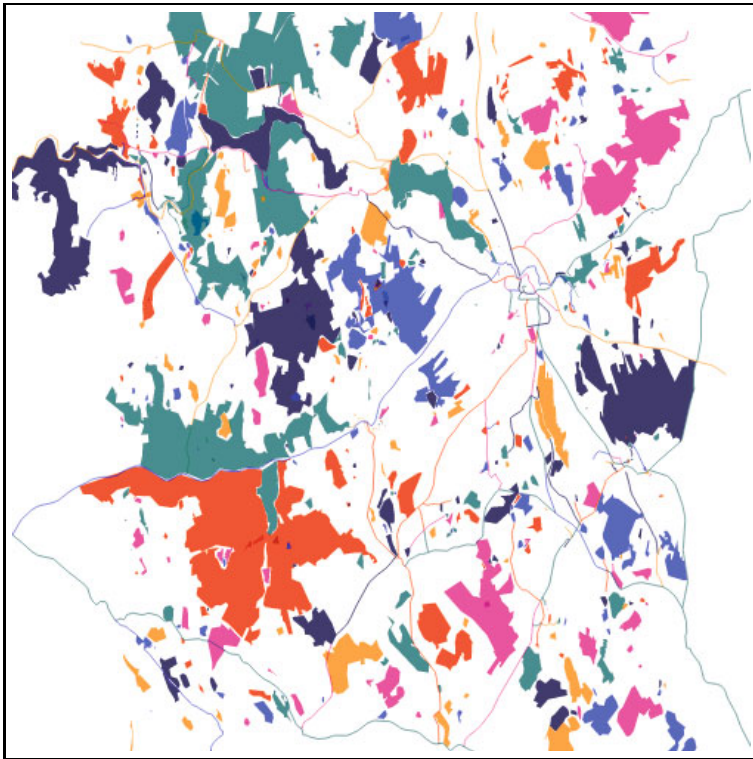
Fig. 6. Example of large scale dataset used in study

Five sample dataset and corresponding map regions were selected. From each region, each of the three transmission methodologies were applied and the corresponding number of feature vertices transmitted determined. Based on this we calculated the percentage of the total map vertices transmitted by each approach. The results of this analysis are displayed in Table 1. The final column of this table shows the time required to compute the required refinement using the fusion methodology. It can be seen from this table that, as expected, in all cases the scale-based approach to transmission requires the transmission of the

entire dataset. The view-based approach reduces this percentage significantly but is not optimal due to the fact that the data for each scale is transmitted completely. The proposed fusion methodology merges the benefits of both transmission methodologies. Consequently it transmits the required map region using the minimum number of vertices of all three approaches.

In experiments it was found that computation of the refinements using the proposed fusion methodology introduced a delay of up to 20 seconds before transmission could be initiated. We are currently developing a methodology which will hopefully reduce computational complexity and make the transmission process applicable in real-time.

(a)

(b)

(c)

(d)
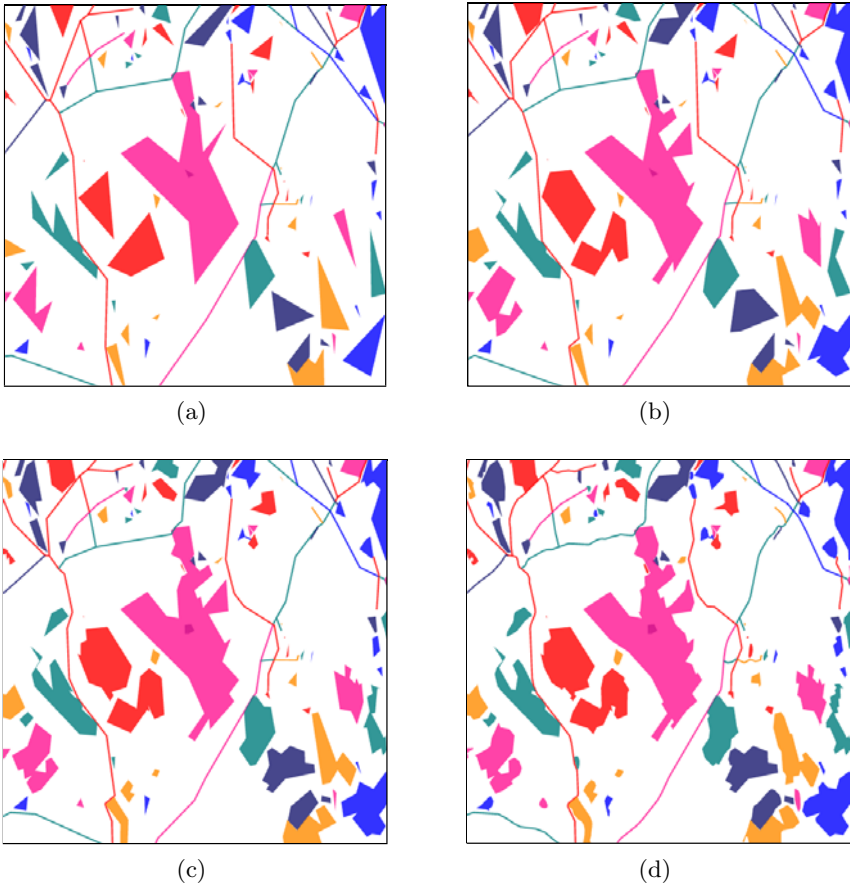
Fig. 7. A small region of Figure 6 is displayed in (d). Three intermediate stages in the transmission of this region using the propose fusion methodology are displayed in (a)-(c). The convergence of this transmission is equal to (d).

6   Conclusions

This paper provides an analysis and clarification of existing methodologies to progressive transmission. This is achieved through the introduction of a novel categorization of such methodologies into three classes of view-based approaches, scale-based approaches and the fusion of these two approaches. Analysis of existing fusion approaches illustrates that despite their potential, existing implementations are inefficient and constrained. To overcome these issues a novel implementation of this paradigm is presented. Results achieved demonstrated that the proposed methodology significantly reduces the amount of data which requires transmission relative to corresponding scale- and view-based approaches. This methodology uses a single generalization and refinement operator. In future work we also hope to implement other operators.

Acknowledgments

References

1. Antoniou, V., Morley, J., Haklay, M.: Tiled Vectors: A Method for Vector Transmission over the Web. In: Carswell, J.D., Fotheringham, A.S., McArdle, G. (eds.) W2GIS 2009. LNCS, vol. 5886, pp. 56–71. Springer, Heidelberg (2009)
2. Basaraner, M.: Model Generalization in GIS. In: Proceedings of International Symposium on GIS, Istanbul (2002)
3. Bertolotto, M.: Progressive Techniques for Efficient Vector Map Data Transmission. In: Spatial Data on the Web: Modelling and Management, pp. 65–84. Springer, Heidelberg (2007)
4. Corcoran, P., Mooney, P.: Topologically Consistent Selective Progressive Transmission. In: AGILE International Conference on Geographic Information Science. Lecture Notes in Geoinformation and Cartography, Springer, Heidelberg (in press, 2011)
5. Corcoran, P., Mooney, P., Winstanley, A., Bertolotto, M.: Effective Vector Data Transmission and Visualization Using HTML5. In: GIS Research UK (GISRUK) (2011)
6. Corcoran, P., Mooney, P., Winstanley, A.C.: Planar and Non-Planar Topologically Consistent Vector Map Simplification. International Journal of Geographical Information Science (in press, 2011)
7. Foerster, T., Stöter, J.: Towards A Formal Classification Of Generalization Operators. In: International Cartographic Conference. International Cartographic Association, Moscow, Russia (2007)
8. Follin, J., Bouju, A., Bertrand, F., Boursier, P.: Multi-resolution extension for transmission of geodata in a mobile context. Computers & Geosciences 31(2), 179–188 (2005)

9. Guttman, A.: R-Trees: A Dynamic Index Structure for Spatial Searching. In: ACM SIGMOD International Conference on Management of Data, pp. 47–57 (1984)

10. Hamid, A.A., Ahmed, M., Helmy, Y.: Enhanced Progressive Vector Data Transmission For Mobile Geographic Information Systems (MGIS). In: Innovations and Advances in Computer Sciences and Engineering, pp. 61–66. Springer, Heidelberg (2010)

11. Haunert, J.H., Dilo, A., van Oosterom, P.: Constrained set-up of the tgap structure for progressive vector data transfer. Computers and Geosciences 35(11), 2191–2203 (2009)

12. Jones, C., Ware, J.: Map generalization in the Web age. International Journal of Geographical Information Science 19(8-9), 859–870 (2005)

13. Li, Q.: Variable-scale representation of road networks on small mobile device. Computers & Geosciences 35(11), 2185–2190 (2009)

14. Lubbers, P., Albers, B., Salim, F.: Pro HTML5 Programming: Powerful APIs for Richer Internet Application Development. Apress (2010)

15. Peng, Z.R., Zhang, C.: The roles of geography markup language (GML), scalable vector graphics (SVG), and Web feature service (WFS) specifications in the development of Internet geographic information systems (GIS). Journal of Geographical Systems 6(2), 95–116 (2004)

16. Ramos, J., Esperança, C., Clua, E.: A progressive vector map browser for the web. Journal of the Brazilian Computer Society 15(1), 35–48 (2009)

17. Saalfeld, A.: Topologically Consistent Line Simplification with the Douglas-Peucker Algorithm. Cartography and Geographic Information Science 26(1), 7–18 (1999)

18. Ai, T., Li, Z., Liu, Y.: Progressive Transmission of Vector Data Based on Changes Accumulation Model. In: Fisher, P. (ed.) Developments in Spatial Data Handling, pp. 85–96. Springer, Heidelberg (2005)

19. Vun, N., Lin, G., Chai, M.: Digital map system with svg display. In: IEEE Tenth International Symposium on Consumer Electronics, ISCE 2006, pp. 1–5 (2006)

20. Weibel, R.: From geometry, statics and oligarchy to semantics, mobility and democracy: some trends in map generalization and data integration. In: Workshop on Generalization and Data Integration, Boulder, Colorado, June 20-22 (2010)

21. Weibel, R., Dutton, G.: Generalizing Spatial Data and Dealing with Multiple Representations. In: Longley, P., Goodchild, M., Maguire, D., Rhind, D. (eds.) Geographical Information Systems: Principles and Technical Issues, pp. 125–155. Wiley, Chichester (1999)

22. Yang, B., Purves, R., Weibel, R.: Efficient Transmission of Vector Data Over the Internet. International Journal of Geographical Information Science 21(2), 215–237 (2007)

23. Yang, B., Weibel, R.: Editorial: Some thoughts on progressive transmission of spatial datasets in the web environment. Computers and Geosciences 35(11), 2175–2176 (2009)

24. Zhang, L., Kang, Z., Li, J., Yang, L.: Web-Based Terrain and Vector Maps Visualization for Wenchuan Earthquake. International Journal of Applied Earth Observation and Geoinformation 12(6), 439–447 (2010)