Visualising User Interaction History to Identify Web Map Usage Patterns

Ali Tahir[1], Gavin McArdle[2], and Michela Bertolotto[3]

[1,3]School of Computer Science and Informatics,
University College Dublin, Belfield, Dublin 4, Ireland
[2]National Centre for Geocomputation,
National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland
{ali.tahir, michela.bertolotto}@ucd.ie, gavin.mcardle@nuim.ie

ABSTRACT

Map personalisation offers a solution to spatial information overload by removing map features which are irrelevant for particular users while highlighting content which is of interest. Determining interests by monitoring and interpreting map interactions is important in this process. This paper describes a Web based application to assist analysts with this task by providing visual analysis tools. A new technique which generates and displays trajectories formed from mouse movement data is presented. When the data from multiple sessions and several users is considered, visualising all trajectories simultaneously introduces its own information overload issues for the analyst. To resolve this and identify salient trends in the movement data, our tool groups trajectories into clusters based on similar movements. This paper describes these techniques and how they have been applied to a sample of mouse movement data.

INTRODUCTION

Spatial information overload has emerged in recent years and is now a well known problem in the Web mapping domain (Yang, 2005). To address this issue, map personalisation can be used to adapt map contents based on user preferences and interests (Mac Aoidh et al., 2009). The challenge is to implicitly determine user interests by monitoring and interpreting map interactions. While traditional data mining techniques can be used for this purpose, a relatively new domain called Visual Analytics, which is defined as "the science of analytical reasoning facilitated by visual interactive interfaces" (Thomas and Cook, 2005) has proven useful in detecting patterns not apparent via data mining alone. By visualising user map interactions and portraying them on a map, patterns can readily be identified. These interaction trends indicate user interests and can be used to improve map personalisation.

This paper describes an approach to visualise user interaction history with a map. In order to identify salient trends, a density based clustering algorithm called Ordering Points to Identify the Clustering Structure (OPTICS) (Ankerst et al., 1999) is applied to the recorded mouse movements. The approach assumes mouse movements form complex and arbitrary trajectories. A trajectory is a path taken by an object in geographical space. Within Web mapping applications, the task of identifying such trajectories becomes challenging as users can alter the map scale (zoom-in/out) depending on their interests. To deal with this situation, a bounding box corresponding to the visible map at each scale is generated. Spatial clustering is then performed on each individual bounding box of the map where spatial interaction has taken place. The output of the clustering algorithm is a reachability plot i.e., a graph that shows the distance between trajectories. This graph visually shows how individual trajectories form clusters and is used to extract clusters based on given threshold values. Finally, the clusters are visualised on a Web interface which is developed using an open source architecture relying on Web 2.0 technologies.

The remainder of this paper is structured as follows: We describe the relevant literature in the related work section. Next, we present the approach we used, to define the mouse trajectory. The functionality of the OPTICS algorithm on which our approach is based is described in detail. We then explain the visualisation of clusters of mouse trajectories over multiple sessions. Finally, we present an approach to evaluating the technique along with a discussion and some conclusions.

RELATED WORK

Visualisation techniques are widely combined with computational analysis to identify patterns in movement datasets in a variety of domains including ecology and meteorology (Horne et al., 2007). Andrienko et al. (2003a) present a catalogue of existing visualisation techniques that are helpful for application developers to appropriately select and combine existing tools depending on their requirements. These exploratory techniques can be applied to a range of spatio-temporal data and various types of analysis tasks. Within the map personalisation domain, one such task is studying user behaviour when interacting with Web maps. In this regard, mouse movements act as implicit interest indicator. These movements on spatial interfaces were investigated by (Mac Aoidh et al., 2007; Mac Aoidh et al., 2008), who developed a visualisation tool called Geospatial Interactions Visualizer (GIViz), which is a stand-alone application that analyses the behaviour of a single user with geospatial datasets. The research presented in this paper extends this work by considering the trajectories formed by mouse movements at different map scales by expanding the approach and also from desktop mapping to Web.

User interaction with a map can involve a large amount of mouse movements, especially when considering the interaction history of multiple users over several sessions. This results in a large number of trajectories being generated. In order to assist the analyst to identify patterns and trends within the data, aggregation and clustering techniques are required (Andrienko et al., 2007; Andrienko and Andrienko, 2010). These methods essentially combine similar trajectories using an appropriate clustering algorithm. The results can be used to create a generalised view which resolves the problem of trajectories obstructing the view of overlapping trajectories.

Clustering is a traditional knowledge discovery and data mining technique which groups similar objects. Density-based clusters are popular for spatial clustering as they have advantages over traditional k-means and hierarchical clustering. Nanni and Pedreschi (2006) have proposed an empirical comparison to identify the natural clusters present in a data source. Density-based clustering successfully locates these, while other methods either fail or obtain less accurate results. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is an early clustering technique (Ester et al., 1996). The principle of DBSCAN is that the density associated with a point is obtained by counting the number of points in a region of specified radius around the point. Points with a density above a specified threshold form clusters. DBSCAN was further refined and improved by Ankerst et al. (1999) with the design of Ordering Points to Identify the Clustering Structure (OPTICS). OPTICS addresses the limitation of DBSCAN in detecting meaningful clusters when density and the sensitivity to input parameters vary in the data. We have opted to use the OPTICS algorithm to perform spatial clustering on mouse trajectories since density-based clusters are well suited for the purpose of trajectory clustering as supported by (Ester et al., 1996; Ankerst et al., 1999).

For clustering to be effective, it is necessary to define an appropriate similarity metric to compare trajectories. There are several approaches in literature for measuring the similarity of trajectories and to group them appropriately. The approaches vary, depending on the analyst's focus. For example, trajectories can be aggregated based on their similarity in geographic, temporal or attribute space (Meratnia and de By, 2002). Rinzivillo et al. (2008) describe four possible distance measures to compute similarity between trajectories which differ in the computational complexity and in the resulting precision in measuring similarity. The two simpler functions use starting and ending points of two trajectories plus several intermediate check points and compute the average spatial distance between the corresponding points. A more complex distance function, *route similarity*, deals with incomplete trajectories and with more significant positioning errors. It is also more suited to unequal time intervals between records. The function repeatedly searches for the closest pair of positions in the two trajectories. It computes the mean distance between the corresponding positions along with a penalty distance for the unmatched positions. Another approach, *route similarity + dynamics*, takes into account the relative times of arrival at corresponding positions. The relative times are the temporal distances to some reference time moments, which are individual for each trajectory. The research presented in this paper utilises the *route similarity* measure for computing the spatial distance between mouse trajectories and ignores the temporal distances as this type of trajectory follows random movements.

While clustering has been applied in several different scenarios and to different movement datasets, to the best of our knowledge it has not been adapted to deal with multiple scales introduced by mouse trajectories on a Web map. Andrienko et al. (2003b) have produced a software package called CommonGIS, which supports spatio-temporal visualisation analysis. CommonGIS can visually analyse generic movement datasets by providing different visualisation techniques, data transformations and spatial decision support facilities. We build on this work and adapt the techniques they have developed for our special case of mouse trajectory analysis. Furthermore, while CommonGIS is developed for a stand-alone application, we propose to use a Web-based approach which is platform independent.

APPROACH

Our Web application called VizAnalysisTools is an extended version of a Web architecture presented by (McArdle et al., 2010). VizAnalysisTools (Tahir et al., 2011) is a suite of tools to analyse mouse movements in order to identify specific usage patterns and behaviours which indicate important user intentions and highlight user interests. The Web application is supported by a visualisation interface, where mouse movements and other associated maps are visualised. Open Geospatial Consortium (OGC) standard such as Keyhole Markup Language (KML) and The World Wide Web Consortium (W3C) recommendation such as Extensible Markup Language (XML), are used. An embedded reporting interface, which uses Google Visualisation API to generate a number of interactive charts (as illustrated in Figure 1), also forms part of the Web application. In addition to the visualisation engine, the Web platform also includes a clustering engine which conducts trajectory analysis on the server side by using Web services to request the trajectory data and to perform geo-computation. At the moment, our two Web applications, GeoAdapta, that records users interactions (McArdle et al., 2010) and VizAnalysisTools, that visualises those interactions (Tahir et al., 2011) work independently. Because of their open nature based on Web services and interoperable formats such as XML and KML, our applications can be embedded in any Web interface.

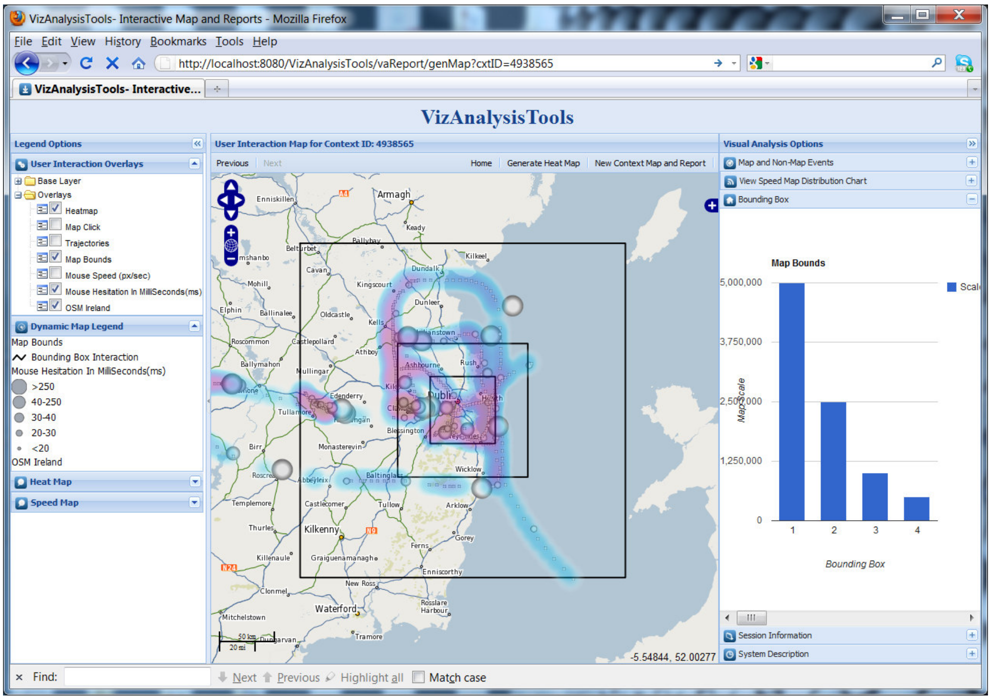

Figure 1: Web interface showing mouse movements and hesitations and the heat map of areas of interests. Corresponding bounding box chart represents user interaction on multiple scales.

A mouse trajectory, in its simplest form, is the mouse movement when a user interacts with an interface. In the spatial context, accurately computing this type of trajectory is challenging since a user can frequently change the map scale (zooming-in/out) or view (panning) based on their interest. As the scale or area of the map being viewed changes, the user generates different bounding boxes. It is not possible to create a trajectory crossing more than one bounding box as to alter a bounding box involves an additional input from the user (such as zoom or pan) which interrupts a trajectory.

Additionally, in a user session, both map and non-map (for example, toggling between geographical layers) movements are recorded. Initially, the trajectories are considered only for the mouse movements on the map. Furthermore, each scale is used to generate a bounding box. Depending on the total length of user session, zoom and pan levels, the amount of trajectories and bounding boxes increase. When trajectories from multiple user sessions are visualised together occlusion can occur making interpretation difficult. To resolve this, clustering is applied on each bounding box to visualise patterns of user mouse movements within a particular scale. In our system, we have applied the OPTICS algorithm (Ankerst et al., 1999) to perform the clustering of mouse trajectories. This algorithm is described in the next section.

OPTICS OVERVIEW

OPTICS is a density-based clustering algorithm that searches for a core distance and reachability distance of an object with respect to its predecessor. OPTICS is a three step process. Firstly, a random object *p* is chosen from the dataset. Then at each iteration *i*, the next object *q* is selected from the dataset with the smallest reachability distance with respect to the already visited core objects. Finally, the process is repeated until all objects in the dataset have been considered. Intuitively, the reachability distance of a point *p* corresponds to the minimum distance from the set of its predecessors. As a consequence, a high value of reachability distance approximately means a high distance from all other objects. The goal of the OPTICS algorithm is to produce an ordering of a dataset and storing the core distance and a suitable reachability distance of each object. OPTICS outputs a reachability plot. From the output plot, clustering can be obtained by choosing an appropriate threshold value of reachability distance.

An important aspect while applying clustering is to find a distance measure between two trajectories. As described in the related work section, different distance functions can be applied to different types of trajectories. Mouse trajectories are random and the corresponding positions between two trajectories do not necessarily match as they also account for unequal time intervals as the user intentions can be totally different in each session. Andrienko et al. (2007) present an algorithm for measuring route similarity between two trajectories. The principle is that two trajectories are repeatedly scanned in search for the closest pair of positions. While scanning, two derivative distances are computed: the mean distance between the corresponding positions and a penalty distance. Skipping a position increases the penalty distance. Finding corresponding positions decreases the penalty distance. The final result is the sum of two derivatives distances. In our system, we need to measure the distance between mouse trajectories. Therefore we extended Andrienko's approach for this specific case. In the following section we describe the initial evaluation we have conducted.

CLUSTERING USER INTERACTION HISTORY

We recorded sample user history consisting of 35 different user sessions. Each session corresponds to a single mouse trajectory. The interaction history can be simultaneously visualised on a portion of a map as shown in Figure 2, which displays the mouse movements of a particular user in a specific geographic area. Given the low scale at which the user interacted with this map, the mouse trajectories appear as straight lines. This is highlighted by the bounding box representing the scale at which the user viewed the map. As the session grows, the mouse movements increase and it becomes difficult to interpret the data. To implement clustering, there are key inputs that have to be carefully selected and tested to yield acceptable results.

OPTICS algorithm takes two input parameters in additions to the dataset; a distance threshold and minimum number of neighbours. As Ankerst et al. (1999) suggest the distance threshold influences the number of clustering levels which can be seen in a reachability plot. The smaller the distance, the more objects may have undefined reachability distances. Therefore, the clusters with lower density might not be visible and ideally this situation should be avoided. Similarly, the larger minimum neighbour value will produce better results.

In our implementation of clustering, we have used distance threshold of 1000 meters (based on the Haversine formula which calculates great-circle distances between two points, longitudes and latitudes, on a sphere) and 4 as minimum neighbours. The resulting reachability plot can be shown in Figure 3, where trajectories are shown along x-axis and the corresponding reachability distances are shown along y-axis. Ideally, valleys in the reachability plot visually identify the clusters but since our dataset was small, this is not very apparent in this scenario.

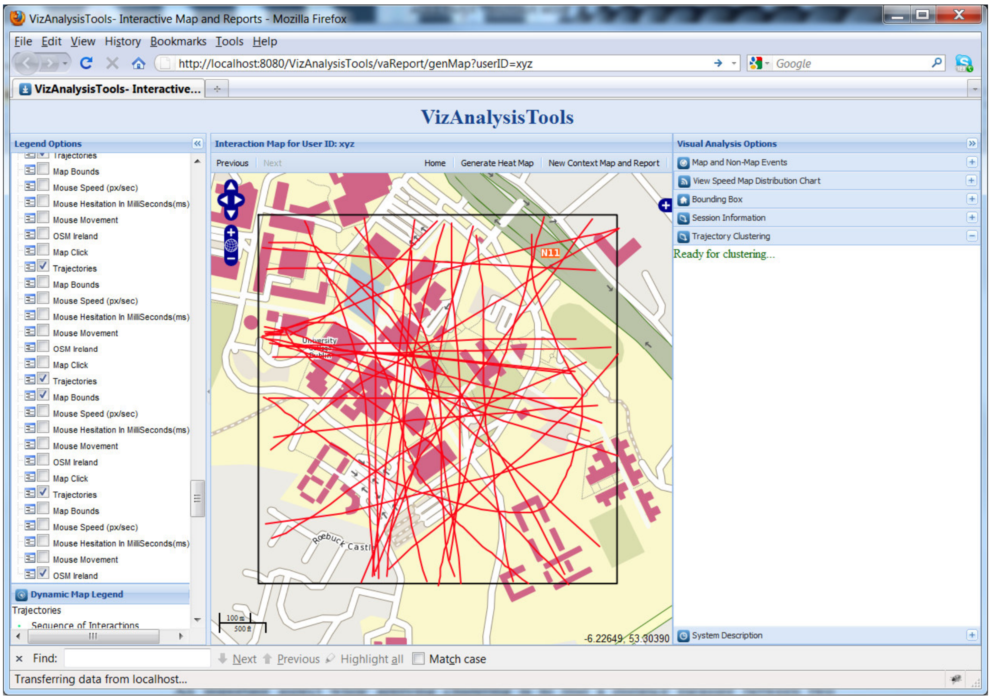

Figure 2: User interaction history visualised over 35 sessions in one bounding box

The cluster ordering of a trajectory is displayed in the plot to the right hand side of the screen shot in Figure 3. There are automatic techniques to extract clustering from this plot based on the valleys in the plot. Another possible approach is to choose an appropriate threshold value of the reachability-distance. We were able to extract 11 clusters from this reachability plot. Once the clusters are identified, they have to be visualised. We opted for a methodology to generate clusters as a KML file representing all clusters including the symbology and the visual style. The cluster width corresponds to the number of trajectories forming a cluster and random colours are assigned to different clusters. Starting and ending symbols are put as markers in each cluster which gives an indication of the direction of the mouse trajectories. The clusters are shown as an overlay on the base map. Additionally, individual trajectories can be switched on and off for visibility.

This sample experiment was performed using our University campus map. The clustering trend in Figure 3 showed that the user was mostly interested in the eastern and southern parts of the campus (corresponding to the interaction window highlighted in the map) where mainly administrative buildings are located. The movement trend can also be used to extract significant places, which are unknown otherwise, based on the user interests.

DISCUSSION & CONCLUSIONS

To produce accurate map personalisation, it is necessary to understand user intentions by interpreting their actions. The research presented in this paper has shown how visual analytics can be used for this approach. Techniques developed for analysing movement patterns are applied to trajectories generated from mouse movements. Although altering map scale introduces challenges for interpreting movements, by applying similarity and clustering metrics to individual and group mouse data based on a bounding box representing a given scale, analysts can identify the salient mouse movement patterns. These can be used to identify areas of high interaction on the map indicating interest in the underlying spatial features. This knowledge can be used to personalise the map by adapting the features presented in the future. A case study demonstrated the approach and reduced 35 mouse trajectories to 11 clusters to indicate the usage pattern within a bounding box.

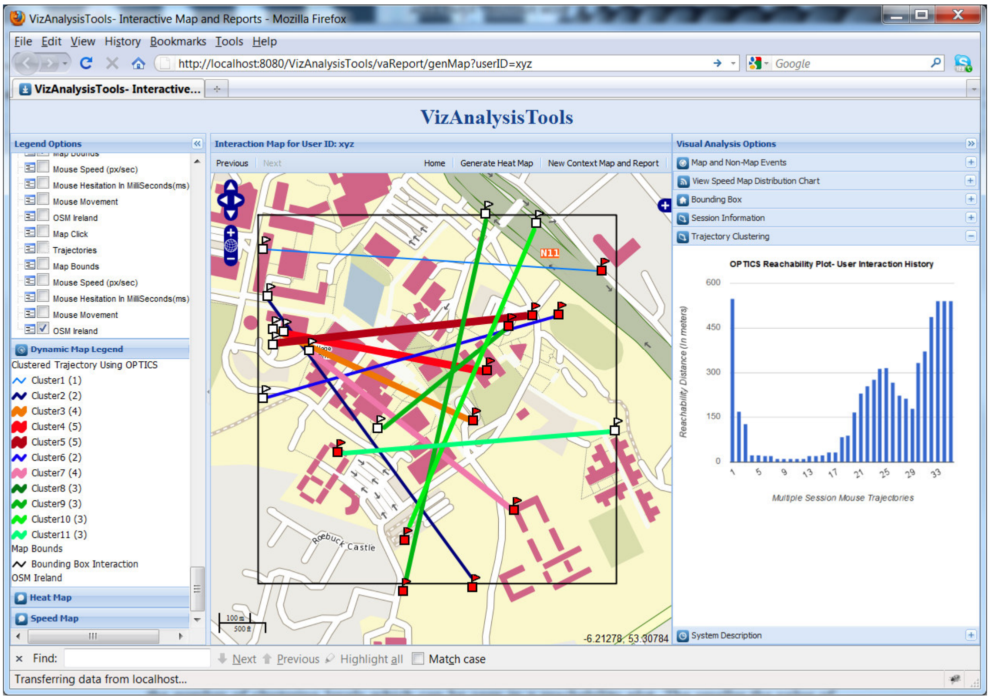

Figure 3: Map showing clustering in a bounding box and the direction of each cluster. The white flag symbol shows the starting point of a cluster while the red flag indicates the ending point. The map legend shows clusters cardinality on the left. The reachability plot is shown on the reporting panel to the right.

Initially, the approach shows clustering on a small set of user sessions, however, further experiments will show the applicability of the approach for larger sets of trajectories. While our similarity metric only considers geographic distance of mouse trajectories, mouse trajectories can also be grouped based on their temporal distances, which is important when the sequence in which mouse movements occur is important. The approach will be tested in terms of its technical validity while its usefulness and usability for analysts will also be evaluated. We plan to test different scenarios involving several types of users and spatial tasks (e.g., new students searching for administrative blocks at the beginning of the semester, etc.). This will ensure that the techniques described in this paper are beneficial for interpreting mouse movement data to assist map personalisation in support of spatial task completion.

ACKNOWLEDGEMENT

Research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan. The authors gratefully acknowledge this support.

BIBLIOGRAPHY

Andrienko, G. and Andrienko., N., 2010 A General Framework for Using Aggregation in Visual Exploration of Movement Data. The Cartographic Journal 47(1): 22-40.

Andrienko, G., Andrienko, N. and Wrobel, S., 2007 Visual Analytics Tools for Analysis of Movement Data. In ACM SIGKDD Explorations Newsletter 9(2): 38-46.

Andrienko, N., Andrienko, G. and Gatalsky, P., 2003a Exploratory Spatio-Temporal Visualization: An Analytical Review. Journal of Visual Languages & Computing 14(6): 503-541.

Andrienko, G., Andrienko, N. and Voss, H., 2003b GIS for Everyone: The CommonGIS Project and Beyond. Maps and the Internet. 131-146.

Ankerst, M., Breunig, M., Kriegel, H.P. and Sander, J., 1999 Optics: Ordering Points to Identify the Clustering Structure. In Proceedings of ACM SIGMOD International Conference on Management of Data (SIGMOD'99) Philadelphia, Pennsylvania. New York: ACM, pp. 49-60.

Ester, M., Kriegel, H., Sander, J. and Xu, X., 1996 A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases With Noise. In: Proceedings. KDD. vol. 96, pp. 226-231.

Horne, J.S., Garton, E.O., Krone, S.M. and Lewis, J.S., 2007 Analysing Animal Movements Using Brownian Bridges. Ecology, 88(9): 2354-2363.

Mac Aoidh, E., McArdle, G., Petit, M., Ray, C., Bertolotto, M., Claramunt, C. and Wilson, D., 2009 Personalization in Adaptive and Interactive GIS. Annals of GIS 15(1): 23-33.

Mac Aoidh, E., Bertolotto, M. and Wilson, D., 2008 Understanding Geospatial Interests by Visualizing Map Interaction Behavior. Information Visualization 7(3): 275-286.

Mac Aoidh, E., Bertolotto, M. and Wilson, D., 2007 Analysis of Implicit Interest Indicators for Spatial Data (short paper). In Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems. (ACMGIS'07), Seattle, WA, USA, November 7-9, 2007, ACM Press, pp. 336-339.

McArdle, G., Ballatore, A., Tahir, A. and Bertolotto, M., 2010 An Open-Source Web Architecture for Adaptive Location Based Services. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Hong Kong. vol. 38(2), pp. 296-301.

Meratnia, N. and de By, R., 2002 Aggregation and Comparison of Trajectories. In Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems. pp. 49-54.

Nanni, M. and Pedreschi, D., 2006 Time-focused Clustering of Trajectories of Moving Objects. Journal of Intelligent Information Systems 27(3): 267-289.

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N. and Andrienko, G., 2008 Visually Driven Analysis of Movement Data by Progressive Clustering. Information Visualization 7(3): 225-239.

Thomas, J.J. and Cook, K.A., 2005 Illuminating the Path: The Research and Development Agenda for Visual Analytics. Piscataway, NJ. IEEE Computer Society Press. ISBN 0769523234.

Tahir, A., McArdle, G., and Bertolotto, M., 2011 A Web-based Visualisation Tool for Analysing Mouse Movements to Support Map Personalisation, In Proceedings of The First International Workshop on SIM3 (in conjunction with DASFAA2011), Lecture Notes in Computer Science, Hong Kong (in press).

Yang, Y. and Claramunt, C., 2005 A Hybrid Approach for Spatial Web Personalization. Lecture Notes in Computer Science 3833, pp. 206–221.