DIVAD: A Dynamic and Interactive Visual Analytical Dashboard for Exploring and Analyzing Transport Data

Tin Seong Kam, Ketan Barshikar, and Shaun Tan

Abstract—The advances in location-based data collection technologies such as GPS, RFID etc. and the rapid reduction of their costs provide us with a huge and continuously increasing amount of data about movement of vehicles, people and goods in an urban area. This explosive growth of geospatially-referenced data has far outpaced the planner's ability to utilize and transform the data into insightful information thus creating an adverse impact on the return on the investment made to collect and manage this data. Addressing this pressing need, we designed and developed DIVAD, a dynamic and interactive visual analytics dashboard to allow city planners to explore and analyze city's transportation data to gain valuable insights about city's traffic flow and transportation requirements. We demonstrate the potential of DIVAD through the use of interactive choropleth and hexagon binning maps to explore and analyze large taxi-transportation data of Singapore for different geographic and time zones.

Keywords—Geographic Information System (GIS), Movement Data, GeoVisual Analytics, Urban Planning.

I. INTRODUCTION

AS city-wide urban infrastructures such as buses, taxis, subways, public utilities and roads become digital, the datasets obtained can be used as a framework for tracking movement patterns through space and time. This is particularly true with the recent trend of massive deployment of pervasive computing technologies such as GPS and RFID on the vehicles [1, 2, 3]. For example, routes and economic behaviour data were collected with the use of Global Positioning System (GPS) devices available on the taxi vehicles. These massive movement data collected are likely to contain structure and patterns that provide useful information about characteristics of the measured phenomena. The identification, analysis and comparison of such patterns will provide greater insights on human movement and behaviours within a city. These understandings will potentially contribute to a better urban management and useful information for urban transport services providers both from the private and public sector to formulate informed decision to gain competitive

Dr. Kam Tin Seong is Associate Professor of Information Systems at School of Information Systems, Singapore Management University, Singapore. (phone: +65 6828 0932; e-mail: tskam@ smu.edu.sg).

Ketan Barshikar is a postgraduate student (MITB SSA Progrmme) at the School of Information Systems, Singapore Management University, Singapore (e-mail: ketanb.2012@mitb.smu.edu.sg).

Shaun Jun Hua Tan C. Author is is a postgraduate student (MITB SSA Progrmme) at the School of Information Systems, Singapore Management University, Singapore (e-mail: shaun.tan.2012@mitb.smu.edu.sg).

advantage.

In real-world practices, the use of these massive locational-aware data, however, tend to be confined to simple tracking and mapping with GIS applications. This is mainly due to a general lack of functions in conventional GIS which is capable of visualising and analysing spatio-temporal data effectively. Furthermore, the use of desktop GIS to analyse these data has limited the access of wider user groups.

In order to meet these two requirements, we present DIVAD – a client-side dynamic and interactive geovisual analytics dashboard. Our aim is to provide transport planners and transport service operators with a distributed geovisual analytics tool for them to detect the expected and to discover the unexpected patterns from massive spatio-temporal movement data.

This paper reports on our research and development effort to design and implement a web-enabled client-based geovisual analytics tool for supporting the analysis and visualisation needs of urban and transport planning professional. It consists of six sections. Section 1 provides a general introduction of the paper. This is followed by an overview of the motivation and objectives of our research effort. Section 3 provides a review of analytical mapping techniques for visualising and analysing point data. The hexagon binning mapping technique used to design DIVID is also discussed in this section. The interface design and software selection of DIVID are discussed in section 4 and 5 respectively. In next section, the practical use of the DIVID is demonstrated using a set of taxi movement data. This includes the data preparation and cleaning process. Lastly, the paper concludes by highlighting the future direction of the research.

II. MOTIVATION AND OBJECTIVES

Our research and development efforts were motivated by the general lack of effective and easy to use web-enabled client-based geovisual analytics tool for discovering spatio-temporal patterns in massive movement data. It aims to provide transport planners and transport services operator with an analytical tool for visual discovering of human movement in a city and their interaction in time and space. Specifically, it attempts to support the following analysis requirements:

1) To be able to display massive movement data cartographically on an internet map such as Openstreetmap;
2) To create a map visualisation framework that supports both macro and micro views;

3) To provide a graphical visualization framework that can display the distribution of travel over time and other related travel attributes such as taxi fares, travelling speeds and volumes of good carried;

4) To be able to display detail records on-demand.

III. ANALYTICAL MAPPING

There has been work done on the visualization of geographically-referenced movement data through conventional GIS. By and large, these data are displayed statically using point symbols. However, when the dataset gets larger or when a third dimension of time is involved, such standards of data visualization become inadequate. For instance, point markers are typically used to display the origins of movement data. When the number of data points increases, the markers will overlap each other, as shown in Fig. 1. When this happens, there will be no way to tell the density difference in the various locations.

Fig. 1 A point map showing the distribution of taxi locations by origins

One way to overcome this problem is to aggregate the data according to a specific administrative boundary and present their distribution in the form of a choropleth map. Unfortunately, as indicated in Fig. 2, choropleth map tends to simplify the spatial patterns of the point data.
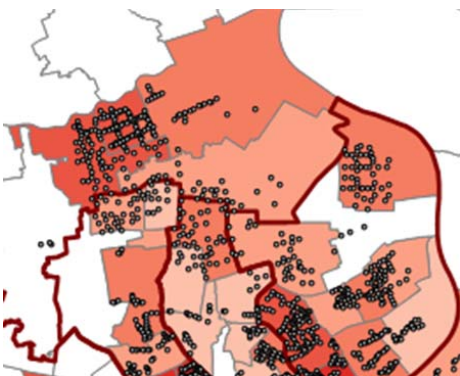

Fig. 2 Choropleth map the distribution of taxi locations by origin

One way to overcome the deficiency of choropleth map is through the use of heat map [4]. Similar to choropleth map, heat map is a type of data visualization that uses color to represent data values in a two-dimensional image. By repeatedly creating kernels of a certain bandwidth over all locations within the study region, then calculating the intensity of data points within the kernel, the final image that is displayed to the user reflects the corresponding color within the gradient (Fig. 3).
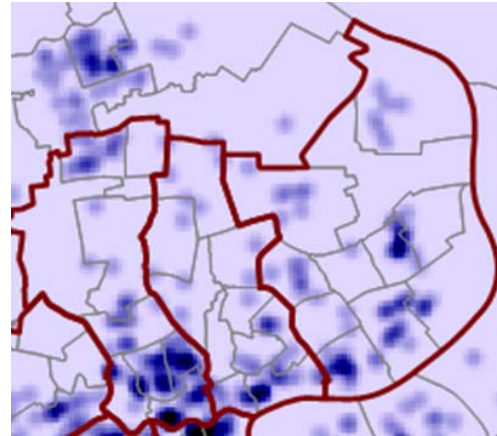

Fig. 3 Heat map created using kernel density method

Despite its usefulness, constructing a heat map with large number of point data (i.e. 1 millions) using kernel density method is computationally demanding. Furthermore, heat map is sensitive to the kernel distance used which is beyond the mathematical skills of urban planners.

In view of this problem, our research aims to explore the potential of hexogen binning method as an alternative mean to visualize and analyse movement data (Fig. 4). Hexagon binning is a form of bivariate histogram for visualizing the structure in datasets with large n [5].
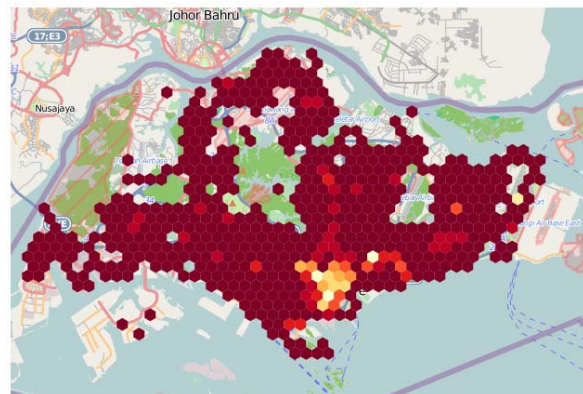

Fig 4 Hexagon binning map showing the distribution of taxis origins

Technically, the polygon used in a hexagon binning map does not have to be in the form of regular grid, triangular or hexagon. According to Lewin-Koh [5], hexagons possess at least three advantages over regular grid and triangular. Firstly, hexagon has the highest symmetry and is the most circular of all regular polygons such as square or rectangle that tile the

plane [6]. Secondly, hexagons have the maximum number of sides a polygon can have for a regular; so in terms of packing a hexagon, it is 13% more efficient for covering the plane than squares. Last but not least, hexagons are visually less biased for displaying densities than other regular tessellations. For example with square, our eyes are drawn to the horizontal and vertical lines of the grid. Hexagons, on the other hand, tend to break the lines.

A hexagon binning map can be created using ArcGIS, a popular commercial-of-the-shelf GIS software and its Hexagon Polygon tool [7]. However, it requires an ArcInfo license, which is beyond the reach of at least 85% of the ArcGIS users.

Alternatively, a hexagon binning map can also be constructed using Quantum GIS [8], an open source desktop GIS and its MMQGIS plugin [9], a set of Python plugins for manipulating vector map layers in Quantum GIS (QGIS).

Using a desktop GIS, a hexagon binning map can be constructed by: (1) laying a polygon or lattice atop a 2-domension filed of point data; (2) counting the numbers of data points in each polygon and (3) shading the polygons using point count values [10].

Despite their usefulness, the hexagon binning map generates using desktop GIS such as ArcGIS and Quantum GIS are static. Any update of the underlying point data used requires the GIS analyst to repeat the analysis procedures. Furthermore, to disseminate the hexogen binning map to wider user base, it has to be published as an internal map layer [7, 10], which is beyond the technical capability of an urban planner, not to mention the high cost involved in the setting-up and maintaining an internet map server.

One way to overcome this problem is to adopt the client-based web mapping approach. In the course of our literature search, we are fortunate to come across the hexbin-js [11], a client-based hexagon binning JavaScript library written by Zachary Johnson. In his blog [12], Zachary demonstrated a test case of hexbin-js to visualize the distribution of Walmarts in the USA. The approach he used in the test case inspired our research direction significantly.

IV. USER INTERFACE DESIGN

The design of DIVAD follows closely the data visualization general guidelines suggested by Schneiderman's mantra [13]: over, zoom and filter, details on demand. It consists of three major views namely: map, graphics and table, depicting different aspects of the data to the user (Fig. 5).

At the top most of the application is the Map view. The design of Map view is based on the principle of Overview + Focus view. It is achieved through the use of two sub-views namely: Overview and Focus views. The Overview view (on the right) provides a macro view of the study area and a hexagon binning map. The hexagons are shaded using choropleth mapping technique. Four data classification methods namely: equal interval, natural breaks, quantiles, continuous, are provided in a form of drop-down list. The colour scheme used is based on the well-known ColorBrwewer scheme [14]. A drop-down list is provided for

users to choose their preferred colour. The Focus view on the right of the Map view displays detail view of a selected hexagon. In this view, each movement data records within a selected hexagon will be represented as point symbols.
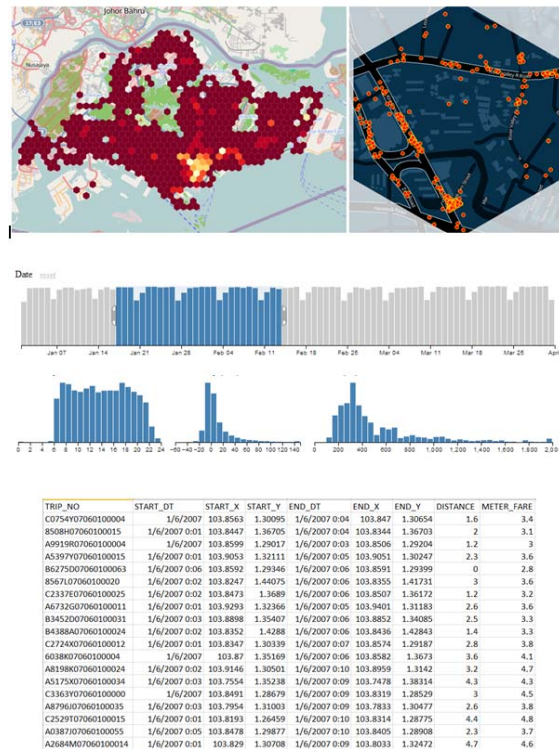

Fig. 5 The interface and components of DIVID

The Graphic view is located at the centre of DIVID. It is made up of four histograms. Each of the histogram displays the distribution of one variable such as the starting time of a taxi journey, hour of the day, travelling speed and meter's fare.

An interactive slider is provided for each histogram that allows user to select a particular temporal period or a value range. For example, a user might want to use the slider in the date histogram to select all the movement data records on Jan 21 and then use the slider of time of the day histogram to further narrow down the selection to movement records at 8.00-10.00am.

The third visual component of DIVID is actually a table. It provides detailed list of the individual point records.

Elements in these three views are linked through coordinated linked view (CLV). CLV is a common data visualization technique that allows the manipulation of visual presentation in different ways as well as coordinating the interaction between the different views [15]. This method to synchronize the presentation of data is very crucial, especially when the user interacts with the application to update the space-time-attribute parameters for an overview-to-detail display.

A combination of these three views along with the ability to select the type of the trip, time of the day, day of the week and hexagon in the hexagon binning map gives the user a great control over the data and more power to analyze and explore.

V. THE APPLICATION

A. System Architecture

Research and development that focus on the design and implementation of data-drive Rich Client web mapping applications or tools are relatively new but it is in the growing trends. Some of the notable examples include the PA CA of GeoVista lab [16], GAV Flash Toolkit of NComVA AB [17] and SpatialKey [18]. By and large, these applications were designed using Flex Builder and required Flash player pre-installed in the web browser. Furthermore, Flash as a runtime is not available on Apple's iOS devices: iPhone, iPod touch and iPad.

After careful consideration, especially by taking into consideration the latest changes in the mobile computing and rich client applications, we have decided to design and development DIVID based on JavaScript.

The architecture of DIVAD was designed and developed with a purpose to make it an open, flexible and re-usable tool which could be used for analysis and exploration of movement point data. As a result of this endeavor, it was decided to make it is a web based tool to enable a wider access to end users.

The system architecture of DIVAD consists of two major components namely web mapping and data visualisation

B. Web Mapping

The web mapping of DIVID consists of two components namely: client-based mapping and a basemap database. The web mapping client of DIVID is developed using Polymap, an open source javascript mapping API jointly developed by Stamen Design and SimpleGeo [19]. Polymaps is a fast and light-weight javascript mapping framework specifically for rendering tiled imagery and vector data quickly and efficiently (using SVG for vectors). Different from conventional server-based internet mapping technology, polymaps is a client-side framework without any server-side dependencies. Unlike Googe's, Microsoft's or Yahoo!'s APIs, the Polymaps API is entirely free and open-source.

Polymaps uses cartography from OpenStreetMap, Bing, and other image-based Web map providers and also supports visual presentations for tiled vector data. The application uses Scalable Vector Graphics (SVG) to display information so one can use CSS rules to easily design a map.

Polymaps was developed with an eye on speedy performance. If you've ever used Google Maps over a bad Internet connection and waited endlessly for a map to reposition itself after zooming in, you know how slow the process can be. Polymaps caches tiles so if it is needed again the tile is pulled from the recent cache instead of the server. If the tile is not available in the cache, adjacent tiles are temporarily rescaled until the needed tile becomes available. Polymaps has a bunch of other tricks up its sleeve for making the whole process of zooming and panning a smooth experience for the user.

The basemap database of DIVAD comes from OpenStreetMap or popularly known as OSM [20]. OSM is a collaborative, worldwide and voluntary initiative to create and disseminate free geographical information. The maps are created using data from portable GPS devices, aerial photography, other free sources or simply from local knowledge.

Unlike Google Map, Microsoft Bing and Yahoo! Maps, OpenStreetMap's primary purpose is to provide the user with the underlying map data. The data are available under the Open Database License. This allows anyone to use that data in any way they want. User can create completely new and exciting map-based products without having to pay any high-cost.

C. Data Visualisation

The data visualization elements such as dynamic bar graphs and histograms of DIVID are created using Data-Driven Document (D3.js), a JavaScript library for manipulating documents based on data library [21] written by Michael Bostock [22]. The D3 library allows user to manipulate elements of a web page in the context of a dataset. These elements can be HTML, SVG or Canvas elements and can be introduced, removed, or edited according to the contents of the dataset [23].

The combination of HTML5 and D3 Javascript provides an interactive user interface with rich and seamless user experience with no requirement to refresh the webpage due to asynchronous communication. The images generated on the user interface are in the form of Scalable Vector Array (SVG) providing high quality and distortion free images at any zoom level.

By using a dataset it is possible to create an interactive SVG visualization with smooth transitions and interaction. With minimal over-head, D3 is extremely fast, supports large dataset and dynamic exhibits behaviours for interaction and animation [23].

DIVAD was developed with a strong emphasis on interactivity and usability. Aligning to this theme, the visual model was designed to provide meaningful insights to the end user in an user-friendly manner.

D3 javascript was used to build a highly interactive and attractive user interface. The user interface not only depicts the movement point data in a visually pleasing and understandable manner but it is also capable of changing dynamically based on the changes in data. We took the approach of using D3 JavaScript to process the large dataset as it handles the data efficiently and provides an elegant user experience.

VI. CASE STUDY: SINGAPORE TAXI DATA

There are a variety of different types of vehicular transportation data that could be used to gain insights into human travel behaviour patterns in a city. For example, most of the taxis in the metropolitan cities are equipped with GPS systems that facilitate the tracking of their movement within

the city. Taxi is an unique kind of public transport whereby they are equipped with not only GPS and also other devices to gather rich data. The trip also represents the true starting and ending points of a journey.

The role and significance of taxi transport cannot be understated especially in a large city. It is a convenient mode of travel used by residents for a one-leg trip to travel to schools, workplaces, recreational venues, etc.

In Singapore, taxi services are closely linked with our daily lives as it represents one of the more important modes of public transport. Statistics from Land Transport Authority (LTA), Singapore revealed that as of 2009, there are 24,702 taxis operating with an average of 864,000 taxi trips daily.

The population which utilizes taxi services represents 14.77% of the public transport commuters, with each average trip distance covering some 9.0 km, a comparatively longer distance than other public transport modes such as buses.

Understanding the travelling patterns of a population with taxis during different hours of a day and across the days of the week can help to identify peak hours and bottlenecks on the road and the route patterns or activities with different regions. Because of all these factors, we have decided to choose taxi data to evaluate the effectiveness of DIVID for detecting, analysing and visualising spatio-temporal patterns of human travel behaviours in Singapore.

The data used for the analysis contained close to two million records with each record representing a taxi trip. It contained GPS data, timing of the trip, speed and duration of the trip and billing information. This data was segregated on the basis of 24 time-slots representing the twenty-four hours of the day to add the time dimension to the trip data.

In order to get reliable and good quality results and to have efficient performance of analysis, it is of utmost importance to prepare the data well for it to be suitable for analysis. The taxi-transport data was carefully examined for various types of anomalies such as incorrect GPS data, unexpected values of distance, time-duration etc. Unnecessary and redundant data were removed to keep the data light-weight and to improve the performance of the analysis and exploration. Aggregation of data was avoided to maintain the raw power of data for precise analysis. The data was sub-divided based on origin and destination of a trip, time of the trip.

Our assessment shows that DIVID possesses several notable advantages over conventional desktop GIS and server-based internet GIS approaches in visualizing and analyzing spatio-temporal patterns of the taxi data. With the Overview + Focus view of DIVID, user can quickly detect usual hotspots from the Overview panel and by rooming the mouse over the hotspot hexagon. The detail distribution in the form of point symbol appears in the Focus panel (Fig. 6).

Fig. 6 Detecting and investigating unusual hotspot with Overview and Focus feature of DIVID

Another very important analysis experience discovered in the assessment is the dynamic and interactive features of DIVID. For example, after detecting a general pattern as shown in Fig. 6, an analyst would like to further explore if this patterns is similar throughout different time-interval of a day. He can easily accomplish this task by using the interactive histogram feature of DIVAD. Due to the unique client-based rendering approach used by DIVID, he will not experience the web refreshing and slow rendering of image commonly experienced in the server-based internet GIS. This will definitely encourage the analyst to be more prepared to slide and dice the taxi data for more in depth exploration and analysis.

VII. Conclusion

The demonstration of DIVAD's potential using the taxi transport data of Singapore underlined its ability to enable the users to explore and analyse the traffic flow data to gain valuable insights regarding the speed, time, density, direction and the purpose of traffic flow which in turn could be used for making policy decisions and for planning transport infrastructure.

DIVAD has a lot of potential to be extended and enhanced further. The stored transport data could be replaced with stream of real-time traffic data in future for a real-time analysis and exploration. The data model could be extended to accommodate public transport data such as bus or metro railway etc. The scope of DIVAD could also be extended to cover larger geographic territory such as a state or a country and for multiple modes of transport. Some of these possible enhancements would be taken up in the near future.

Acknowledgment

References

[1] Michael Wegener, "Current and Future of Land Use Planning" *Land Use Model Conference organized by the Texas Transportation Institute*, Dallas, 19-21 February 1995.
[2] Wenhuan Shi, Qing-Jie Kong, Student Member, IEEE and Yuncai Liu, "A GPS/GIS Integrated System for Urban Traffic Flow Analysis", *The 11th International IEEE Conference on Intelligent Transportation Systems Beijing*, China, October 12-15, 2008.

[3] Daniel Mcfadden, "The Measurement of Urban Travel Demand" *Journal of Public Economics*, North-Holland Publishing Company, 1974.
[4] Filippova, D. Lee, J. Olea, A. VanDaniker, M. and Wongsuphasawat, K. "Exploring Clusters in Geospatial Datasets" *Tech Report HCIL-2008-26*, University of Maryland.
[5] Lewin-Koh, N (2011) Hexagon Binning: An Overview. (cran.r-project.org/web/packages/hexbin/).
[6] Yale,P.B. 1968. *Geometry and Symmetry*. Holden-Day, San Francisco, 288 pp.
[7] Field, K. (2012) "Using a binning technique for point-based multiscale web maps". Available at http://http://blogs.esri.com/esri/arcgis/2012/06/08/using-a-binning-technique-for-point-based-multiscale-web-maps/) Accessed 10 Sept 2012.
[8] Quantum GIS. Available at http://www.qgis.org/en/documentation/manuals.html Accessed 30 Oct 2012.
[9] Minn, M (2012) MMQGIS. Available at http://michaelminn.com/linux/mmqgis/ Accessed 28 Oct 2012.
[10] MapBox blog. Binning: An Alternative to Point Maps, May 25 2012. Available at http://mapbox.com/blog/binning-alternative-point-maps/. Accessed 31 Oct 2012.
[11] Johnson, Z. Hexbin-js. Available at https://github.com/indiemaps/hexbin-js. Accessed 20 June 2012.
[12] Johnson, Z. (Oct 18 2011) Hexbins!. Available at http://indiemaps.com/blog/2011/10/hexbins/. Accessed 10 january 2012.
[13] Shneiderman, B. (2005) "The eyes have it: A task by data type taxonomy for information visualization" *IEEE Conference on Visual Languages (VL96)*, pp. 336-343.
[14] ColorBrewer.org Colorbrewer: Color Advice for Maps. Available at http://colorbrewer2.org/. Accessed 20 Sept. 2012.
[15] Robert, J.C. "State of the Art: Coodinated & Multiple Views in Exploratory Visualization", *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*.
[16] MacEachren, A., Crawford, S., Akella, M. and Lengerich, G. (2008) "Design and Implementation of a Model, Web-based, GIS-Enabled Cancer Atlas". *The Cartographic Journal*, Vol. 45, No. 4, pp. 246-260.
[17] NComVa AB. GVA Flash Toolkit. Available at http://www.ncomva.com/technology/technology-overview/toolkit/. Accessed 30 Oct 2012.
[18] SpatialKey.org. SpatialKey. Available at http://www.spatialkey.com/. Accessed 30 Oct 2012.
[19] Stamen Design and SimpleGeo. Polymaps. Available at www.polymaps.org. Accessed 31 Oct 2012.
[20] Wikipedia OpenStreetMap. Available at http://en.wikipedia.org/wiki/OpenStreetMap. Accessed 31 Oct 2012.
[21] Bostock, M. (2012) Data-Driven Document. Available at d3js.org. Accessed 31 Oct. 2012.
[22] Bostock, M. Ogievetsky, V. and Heer, J. (2011) "D3:Data-Driven Documents" *IEEE Trans Vis and Comp Graphics*, 17(12), pp. 2301-2309 Dec 2011.
[23] Dewar, M (2012) Getting Started with D3. O'Reilly Media Inc. USA.
[24] Azzi, M., Caviglia, G., Ciuccarelli, P., Bonetti, E. and Bontempi, L. (2011) "Dust: A Visualization Tool Supporting Parent's School-Choice Evaluation Process". *Parson Journal for Information Mapping*, Vol. III, Issue 4, Fall 2011, pp. 1-7.
[25] Fujino,T. (2007) "SVG+Ajax+R: a new framework for WebGIS" *Computational Statistics*, Vol. 22, pp. 511-520.
[26] Liu, L. Andris, C. and Ratti, C. (2010) "Uncovering cabdrivers' behaviour patterns from their digital traces" *Computer, Environment and Urban Systems*. Vol. 34, pp. 541-548.
[27] J. B. GODDARD, "Functional Regions within the City Centre: a Study by Factor Analysis of Taxi Flows in Central London" *Transactions of the Institute of British Geographers*, Blackwell Publishing, Mar 1970.
[28] Yeo Zhi Shu Eunice, "A Rich Internet Application (RIA) based Geo-Visual Analytics Approach", IS470 – Guided Research in Information Systems, Singapore Management University's school of Information Systems, 2009.

Dr. Kam Tin Seong is currently associate professor of information systems (practice) at the School of Information Systems, Singapore Management University, with interests in designing and implementing geovisual analytics tools for real world applications.

Ketan Barshikar is currently pursuing Master of IT in Business (Service Sector Analytics) at Singapore Management University. He completed his Bachelor of Engineering in Computer Science from University of Pune and thereafter worked at Tata Consultancy Services India Ltd as a System Engineer for close to three years.

Shaun Tan is currently pursuing his Bachelor of Science (Information Systems) and a Master of IT in Business (Service Sector Analytics) at Singapore Management University.