Visualizing Bike Share Traffic

Josua Krause

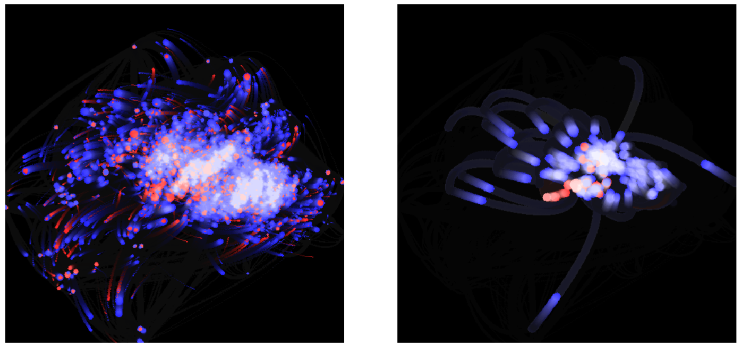

Fig. 1. One month worth of bike share users with and without thresholding aggregated journeys. Blue dots are subscribing bike share users and red dots are casual users. On the right picture only trips that are done at least 200 times are shown.

Abstract— Bike share programs provide a continuous stream of human movement data. Finding patterns in this data can be beneficial for bike share companies and useful for improving infrastructure. We present a tool for analyzing temporal origin-destination data which represents trips as animated dots. Trips can be aggregated and filtered by frequency. The continuous stream of trips is split into time slices and the slices can be split further to focus on different parts of the slices. We show the usefulness of our tool with bike share data from Washington, DC.

Index Terms— Origin-destination data, bike share, flows

1 Introduction

Analyzing movement patterns of people is a task that just recently became feasible. With the help of GPS trackers the daily routine of people can be analyzed [21, 19, 20], although the number of participants is limited. A larger number of trips can for example be acquired by installing GPS trackers in taxi cabs [8, 11]. With the inception of bike sharing programs in various cities large amounts of journey data from people daily utilizing bicycles have been automatically created.

For companies providing shared bicycles it is interesting to see who uses the bikes, when they are used, and which routes have been taken. For maintenance reasons it is also important to know whether the usage of the bikes are balanced between stations or whether there are some stations that are used more often as start or destination. Tourists or one time users may produce such imbalances, so it would be beneficial to detect whether casual users follow any patterns. Since bike sharing is popular among commuters, using it daily to get to work, it is interesting to see how the usage patterns change during the rush-hour. This information can be used to create infrastructure to make commuting by bike easier.

In this paper we introduce a tool for analyzing temporal origin-destination data and use it on data provided by the bicycle sharing program of Washington, DC: Capital Bikeshare [1]. We will first present various approaches that have been used to visualize origin-destination data (see Section 2). Then we will show the data set (see Section 3) and the tool (see Section 4). After that we present results (see Section 5).

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

2 Related Work

Multiple approaches have been used to visualize origin-destination data with temporal features.

Rae [14] use heat-maps based on the density of migration flows of cities in the UK. The user can also select a city to see its flow in a node-link representation. By computing density based clusters, Guo *et al.* [11] define regions to show the net-flow of taxis in Hong Kong as choropleth maps. A kernel density representation shows the distribution of destinations. A multi-color kernel density representation is used by Ferreira *et al.* [8] to show the distribution of origins and destinations of taxi trips in New York.

Guo *et al.* [10] use an origin-destination matrix to show companies relocating within the US. Rows and columns can be reordered to identify clusters and patterns, but provide no spatial context. Also, the resolution of origins and destinations is limited to states due to the use of a matrix. In order to address the lack of spatial context Wood *et al.* [17] explore nested matrices that are laid over a geographic map. Every cell of the matrix aggregates the origins of the geographic map and contains a second matrix showing aggregated destinations of a smaller scale version of the map.

Becker *et al.* [3] compare origin-destination matrices to node-link representations and propose an animated dynamic node-link representation where the user can define time intervals to aggregate the data, set thresholds to limit the number of visible links, and choose the regions that are displayed.

The use of animation can make keeping track of changes difficult (Tversky *et al.* [16]). However, animation can be used to identify trends (Heer and Robertson [12], and Robertson *et al.* [15]) and patterns (Griffin and MacEachren[9]). Boyandin *et al.* [7] show that animated node-link representations of flows make spatial and short temporal patterns easier to detect than in a similar small multiples representation.

Using a node-link representation for showing origin-destination data introduces clutter for large data-sets. Boyandin *et al.* [6] use colors to identify sources and sinks of migration flows, time-tables for

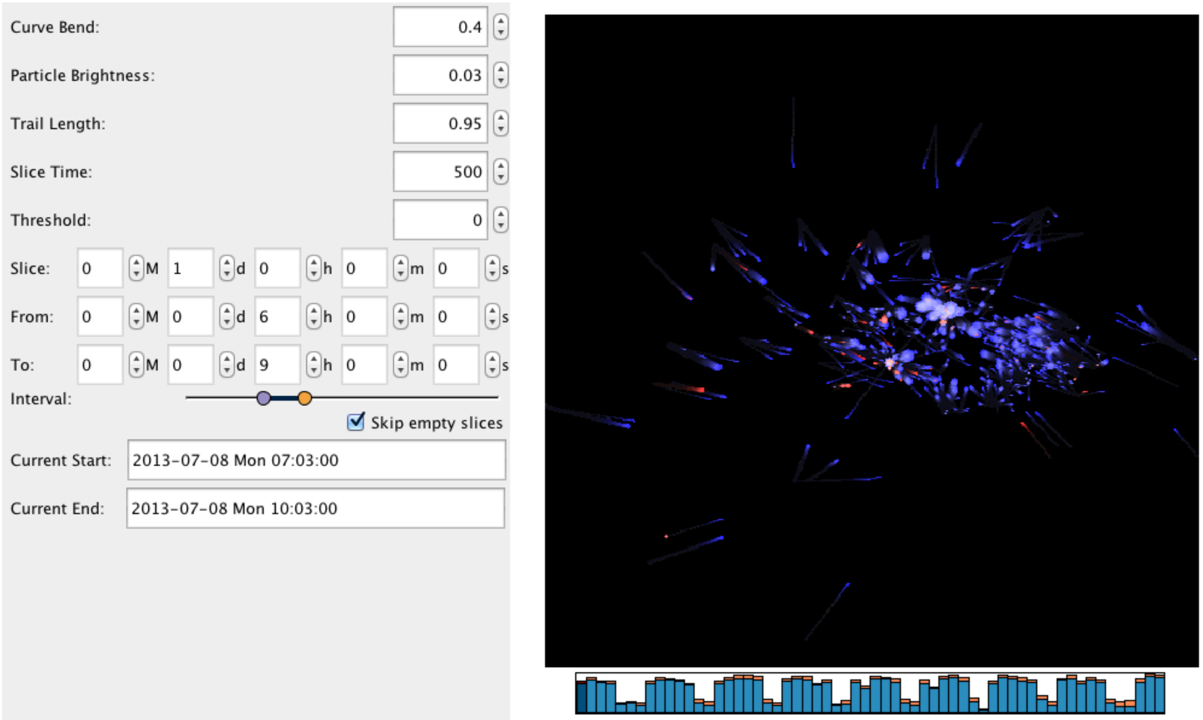

Fig. 2. The tool is split in three parts. On the right side is the dot visualization. On the bottom right is the bar chart representation. On the left is the control panel, where the user can define the excursion of the trips, the brightness of the dots, the length of the trails, the animation speed of the slices and the threshold to remove dots that are too small. With the spinners below this, the user can define the length of a slice and the start and end of the relevant window. The relevant window can also be edited by using the sliders below. The user has also the option to skip slices that contain no trips, however this only occurs when the duration of the slice is chosen very short. The text fields at the bottom show the start and the end time of the currently displayed relevant window.

temporal navigation, and edge bundling to overcome clutter. However, edge bundling may imply hierarchical relations of flows which do not exist in the data.

Boyandin *et al.* [5] displays origins and destinations of migration flows on separate maps connected by links to a time-table showing the magnitude of the flows for a given time. Another way to avoid clutter in node-link flow representations is to aggregate flows with similar origins and destinations. Wood *et al.* [18] does this with bicyclist data from London by showing asymmetric bezier curves as links to avoid over-plotting of opposing flows. Beecham *et al.* [4] improves this by also showing hourly densities in a cycle plot and allows brushing on them to filter input data.

3 Data

The data used in this paper is provided by Capital Bikeshare [1] the bike share company of Washington, DC. The data, reaching back to October 2010, is freely available [2], though we used only data from October 2012 until September 2013 to avoid the need to skip the majority of the time to get to the latest data. The remainder are 2470109 unique bicycle trips through Washington.

The trips are provided in multiple CSV files that vary in their format per quarter. Each trip consists of the name of the station where the bike was picked up, the name of the station where the bike was put back, the time of starting the trip, the duration, and whether the person was a casual user or a subscribing member. We processed the names of the stations, usually consisting of the intersecting road names, and converted them to geographic coordinates with a Geocoding tool [13]. We stored the trips in a MySQL database with the start time as index to allow for fast ranged trip lookups by start time.

4 The Tool

In our tool we present trips as dots moving on a black background. Every dot can be colored blue (indicating bike share subscription of the driver) or red (indicating a casual user). The dots move on a curved path from its origin to its destination to avoid over-plotting of opposing trips similar to Wood *et al.* [18] and Beecham *et al.* [4]. However, they cycle the animation of a given set of trips continuously, whereas we split the time in slices and advance time constantly. In our approach the complete data set is split into time slices of equal length and within those slices we define a relevant window. At the beginning of each time slice we create dots for trips of the relevant window and start moving them. If a trip lasts longer than the relevant window we let it continue to move during the next slice.

To make the vast number of simultaneous trips for longer slices easier to read we aggregate similar trips based on origin, destination, duration, and type. The aggregated trips are then represented as larger dot. Furthermore, we show a trail for each moving dot to make its route easier to follow and gasp. The user can vary those parameters when needed and can also set a threshold of a minimal number of trips for a dot to be shown at all.

In addition to the animated trails the tool shows a summary of future slices as bar chart. The bar chart shows the total number of trips in the relevant window of the given slices. It starts with the currently displayed slice and shows the next 59 slices. The blue part of the bar shows the number of subscriber trips during the relevant window and the red part shows the number of casual users.

As further information about the current state the start and end times of the current relevant window is always shown. The complete tool can be seen in Figure 2. It is efficiently implemented using a MySQL database as storage and OpenGL accelerated image blending to show the movement of the dots and simulate their trails. This allows for an interactive user experience with immediate feedback.

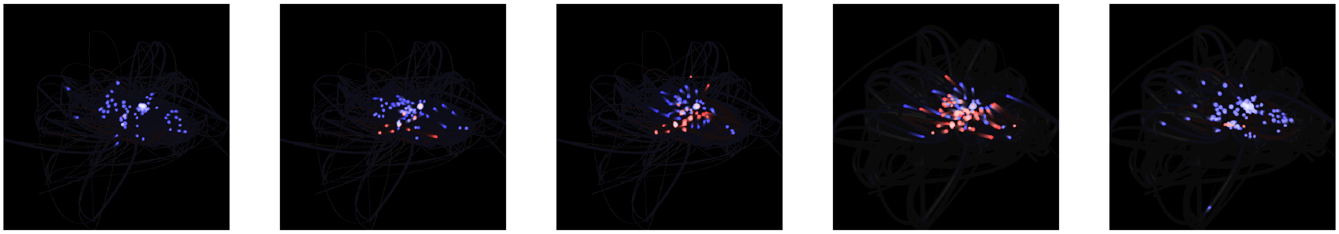

Fig. 3. Different time slices. The leftmost image shows bike usage during the morning rush hour, the second shows bike usage during the day on a week day. The third image shows bike usage during the day on a Sunday. The last two images show the bike usage of two complete days (relevant window is 24 hours), the first being a Sunday and the second being a weekday. All images are during the summer months.

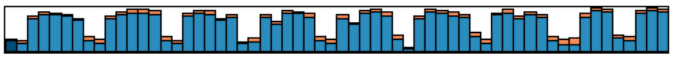

Fig. 4. The bar chart view showing the number of trips per relevant window. In this bar chart the slice length is one day, the relevant window is during the morning rush hour. Weekends can be easily identified by a decreasing number of trips which sometimes start already on fridays. The current (leftmost) slice is a Saturday.

5 Results

By comparing trips during the morning rush-hour (time window from 6am to 9am) to trips during the day until the start of the evening rush-hour (9am to 5pm) one can easily see how different user groups use the bike sharing program. See Figure 3 for some snapshots.

The daily commuters during the rush hour have usually longer journeys than users during the day. The commuters create a temporary imbalance in the morning by bringing bicycles to the center of the city where they work. However, the imbalance is resolved in the evening during this rush hour. During the relatively early time of the morning rush hour the number of casual users is very low.

During the day the trips are much shorter and concentrate on the center of the city. Casual users drive from the White House to the Lincoln Memorial and vice versa implying this service is mainly used by tourists visiting the city. The trips, however, are not equally distributed in both directions which leads to slight (due to the relatively small number) imbalance of bicycles in the touristy area.

Comparing weekdays to weekends shows that during the weekend the total number of trips far less than during the week (This can also easily be seen by looking at the bar chart representation. See Figure 4). Also the ratio of casual users is much higher. This is likely due to day tourists visiting over the weekend and the working commuters staying at home on the weekend. However, trips on a weekend are not as much focused on the center as trips during the mid day on the week. This may come from people that normally do not use the bike share but use it on weekends to get to the city.

6 Future Work

The tool can be expanded to show multiple views of different relevant windows at once to make comparison easier. Making the bar chart view clickable in order to directly jump to a time slice would make navigation through the data set easier. A cycle plot as used by Beecham *et al.* [4] can be used to show the distribution of trips within a slice which would make it easier to choose the relevant window.

Also, using other data sets without fixed origins and destinations can create new challenges of how to aggregate trips effectively.

References

[1] Capital bikeshare. `http://capitalbikeshare.com/`.

[2] Capital bikeshare trip history data. `http://capitalbikeshare.com/trip-history-data`.

[3] R. Becker, S. Eick, and A. Wilks. Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics*, pages 16–28, Mar. 1995.

[4] R. Beecham, J. Wood, and A. Bowerman. Identifying and Explaining Inter-Peak Cycling Behaviours within The London Cycle Hire Scheme. *Workshop on Progress in Movement Analysis: Experiences with Real Data*, 2012.

[5] I. Boyandin, E. Bertini, P. Bak, and D. Lalanne. Flowstrates: An Approach for Visual Exploration of Temporal Origin-Destination Data. *Computer Graphics Forum*, 30(3):971–980, June 2011.

[6] I. Boyandin, E. Bertini, and D. Lalanne. Visualizing Migration Flows and their Development in Time: Flow Maps and Beyond. 2008.

[7] I. Boyandin, E. Bertini, and D. Lalanne. A Qualitative Study on the Exploration of Temporal Changes in Flow Maps with Animation and Small-Multiples. *Computer Graphics Forum*, 31(3), 2012.

[8] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban data: a study of new york city taxi trips. *IEEE Trans. Vis. Comput. Graph.*, 2013.

[9] A. Griffin and A. MacEachren. A comparison of animated maps with static small-multiple maps for visually identifying space-time clusters. *Annals of the Association of American Geographers*, 96:740–753, 2006.

[10] D. Guo, J. Chen, A. M. MacEachren, and K. Liao. A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE transactions on visualization and computer graphics*, 12(6):1461–74, 2006.

[11] D. Guo, X. Zhu, H. Jin, P. Gao, and C. Andris. Discovering Spatial Patterns in Origin-Destination Mobility Data. *Transactions in GIS*, 16(3):411–429, June 2012.

[12] J. Heer and G. Robertson. Animated transitions in statistical data graphics. *IEEE transactions on visualization and computer graphics*, 2007.

[13] S. Morse. Conversions of address to latitude/longitude. `http://stevemorse.org/jcal/latlonbatch.html?direction=forward`.

[14] A. Rae. From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems*, 33(3):161–178, May 2009.

[15] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. Effectiveness of animation in trend visualization. *IEEE transactions on visualization and computer graphics*, 14(6), 2008.

[16] B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57:247–262, 2002.

[17] J. Wood, J. Dykes, and A. Slingsby. Visualisation of Origins, Destinations and Flows with OD Maps. *Cartographic Journal, The*, 47(2):117–129, Jan. 2002.

[18] J. Wood, A. Slingsby, and J. Dykes. Visualizing the Dynamics of London's Bicycle-Hire Scheme. *Cartographica*, 46:239–251, 2011.

[19] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding Mobility Based on GPS Data. *In Proceedings of ACM conference on Ubiquitous Computing*, 2008.

[20] Y. Zheng, X. Xie, and W.-Y. Ma. GeoLife: A Collaborative Social Networking Service among User, location and trajectory. *Invited paper, in IEEE Data Engineering Bulletin*, 2010.

[21] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from GPS trajectories. *In Proceedings of International conference on World Wild Web*, 2009.