Contents lists available at ScienceDirect

### **Visual Informatics**



journal homepage: www.elsevier.com/locate/visinf

# A visual analytics design for studying rhythm patterns from human daily movement data

Wei Zeng<sup>a,\*</sup>, Chi-Wing Fu<sup>b</sup>, Stefan Müller Arisona<sup>c,a</sup>, Simon Schubiger<sup>c,a</sup>, Remo Burkhard<sup>a</sup>, Kwan-Liu Ma<sup>d</sup>

<sup>a</sup> Future Cities Laboratory, ETH Zurich, Switzerland

<sup>b</sup> The Chinese University of Hong Kong, Hong Kong

<sup>c</sup> University of Applied Sciences and Arts Northwestern Switzerland FHNW, Switzerland

<sup>d</sup> University of California-Davis, United States

#### ARTICLE INFO

Article history: Received 2 June 2017 Accepted 12 July 2017 Available online 31 August 2017

Keywords: Movement rhythm Event sequence Visual analytics

#### ABSTRACT

Human's daily movements exhibit high regularity in a space-time context that typically forms circadian rhythms. Understanding the rhythms for human daily movements is of high interest to a variety of parties from urban planners, transportation analysts, to business strategists. In this paper, we present an interactive visual analytics design for understanding and utilizing data collected from tracking human's movements. The resulting system identifies and visually presents frequent human movement rhythms to support interactive exploration and analysis of the data over space and time. Case studies using real-world human movement data, including massive urban public transportation data in Singapore and the MIT reality mining dataset, and interviews with transportation researches were conducted to demonstrate the effectiveness and usefulness of our system.

© 2017 Zhejiang University and Zhejiang University Press. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND licenses (http://creativecommons.org/licenses/by-nc-nd/4.0/).

#### 1. Introduction

In transportation and geographical information systems (GIS), human movements are usually presumed to engage in certain activities, e.g., work, studying, and shopping. Hence, humans' daily movements can be described as "a scheduling of activities in time and space" (Primerano et al., 2008), e.g., home  $\rightarrow$  work  $\rightarrow$  home and home  $\rightarrow$  school  $\rightarrow$  tuition  $\rightarrow$  home. The movements can be further generalized as network motifs by abstracting the activity information (Schneider et al., 2013), e.g., home  $\rightarrow$  work  $\rightarrow$  home can be generalized as  $A \rightarrow B \rightarrow A$ , and home  $\rightarrow$  school  $\rightarrow$  tuition  $\rightarrow$  home as  $A \rightarrow B \rightarrow C \rightarrow A$ .

In this work, we denote these motifs as movement rhythms, each of which basically describes *a sequence of locations visited in time and space*. A better grasp of human movement rhythms can be highly beneficial for various applications, e.g., travel demand management. For instance, by studying individuals' activity and travel schedule, transportation researchers derived an integrated discrete choice model to analyze travel demands at different times of a day (Bowman and Ben-Akiva, 2000).

\* Corresponding author.

E-mail address: zeng@arch.ethz.ch (W. Zeng).

Peer review under responsibility of Zhejiang University and Zhejiang University Press.

To explore the movement rhythms over space and time, an interactive visual analytical tool that facilitates transportation experts' exploration is preferred. Nonetheless, there are several challenges to overcome. First, a direct plot of all the human daily movements, such as to display the changes of geospatial positions in time in 3D space (Kapler and Wright, 2004), can easily lead to visual clutter. Second, the movements of human daily movements can exhibit many different rhythms, e.g.,  $A \rightarrow B \rightarrow A$  and  $A \rightarrow B \rightarrow C \rightarrow A$ , etc. Appropriate data modeling should be developed to efficiently classify these movement rhythms. And lastly, human movements involve many different types of activities, which are happening at different locations in space, and take different times to finish the activities and to travel between locations. The visualization should present the spatial and temporal perspectives of information in an intuitive way.

In our previous work (Zeng et al., 2016b), we presented a visual analytics design for studying movement rhythms from massive public transportation data. This work presents an extended version by applying the approach on another human daily movement data, i.e., the MIT reality mining dataset (Eagle and Pentland, 2006). We first describe an efficient movement modeling method to identify movement rhythms based on the movement's spatial and temporal characteristics (Section 4.3). All movement rhythms are organized into a hierarchical tree structure with a new tree construction algorithm (Section 4.4) devised from the association rule concept.

http://dx.doi.org/10.1016/j.visinf.2017.07.001

2468-502X/© 2017 Zhejiang University and Zhejiang University Press. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).



We show that our algorithm can preserve more details about movement rhythms than typical methods, and can be generalized to aggregate event sequence data in a level-of-detail style. We then present the *Rhythm Sequence View* to depict the temporal perspective of movement rhythms, together with the *Rhythm Density View* plotting movement origin and destination distributions in spatial dimension, and the *Rhythm Statistic View* over-viewing the statistics of frequent movement rhythms (Section 5). In the end, we apply our approach to the study of real-world human daily movement data, i.e., massive urban public transportation data in Singapore and the MIT reality mining dataset (Section 6). Interviews with transportation researches are conducted to demonstrate the effectiveness and usefulness of our system (Section 7).

The major contributions of this work are:

- a new event aggregation algorithm, which can preserve more details of event sequence data than typical methods, and can be generalized to achieve a level-of-detail visualization;
- a visual analysis approach integrating human interactions and data processing to explore human movement rhythms;
- case studies on real-world human daily movements showing interesting movement rhythms, especially the identification of 12 frequent movement rhythms extracted from massive public transportation data in Singapore.

#### 2. Related work

We review related researches in three topics: time geography, movement data and event sequence visualization.

#### 2.1. Time geography

Human movement over space and time can be conceptually formulated by time geography (Hägerstrand, 1970). Time geography employs a space-time cube to present the movement trajectory in 3D: the 2D horizontal plane for spatial and vertical axis for temporal.

Many studies have extended time geography framework to portray and analyze human movements. For example, some works developed new visual representations, such as GeoTime (Kapler and Wright, 2004) and stacking-based trajectory wall (Tominski et al., 2012). Others employed the framework to explore human movement patterns, e.g., individual's activity dairy data (Chen et al., 2011) and student travel behavior (Kamruzzaman et al., 2011).

These works show that space-time cube can be a prominent solution from a visualization perspective. However, directly plotting the movement data in 3D can easily lead to visual clutter problem when the dataset size increases. This limits the applicability of the approach for our problem, where human daily movements in Singapore involve over 30 millions of trips. Therefore, we employ the *pattern extraction* and *summarization* approach (Andrienko et al., 2008) by identifying movement rhythm patterns and visualizing the aggregated characteristics.

#### 2.2. Movement data visualization

One key challenge for movement data visualization is to effectively present the spatio-temporal movement patterns posted by the large data size and to support the complex analytical tasks demanded by the domain users, see Andrienko and Andrienko (2012) for a systematic review. Below, we only discuss a few representative works and summarize them in the following categories.

Visual display: We can design novel visual structures to reveal movement patterns in the data, e.g., waypoints-constrained OD view (Zeng et al., 2016a). Then, the visualizations can organize visual structures to build up a user interface, which can present movements in 3D space, e.g., stacking-based trajectory wall (Tominski et al., 2012), or linked views with multiple perspectives, e.g., TripVista (Guo et al., 2011), or an integrated view, e.g., occlusion-free temporal maps (Sun et al., 2014).

Interactive techniques allow users to filter and explore the movement data based on user demands. FromDaDy (Hurter et al., 2009) supports to brush trajectories, and pick and drop the brushed information into juxtaposed views. TrajectoryLenses (Krüger et al., 2013) allows users to filter trajectories based on their origins, destinations or waypoints. A more elaborated interaction tool can be found in Scheepens et al. (2016), which allows for selection of area, directions, additional attributes, etc.

*Computation processing* leverages machine analysis capability to help explore the movement data. Various works have been carried out in this direction, e.g., clustering trajectories (Andrienko et al., 2010), and inferring mobility patterns (Zeng et al., 2014).

In this work, we first employ an efficient modeling method to identify movement rhythms, and map the spatial and temporal perspectives of information into linked views with a set of interactive query methods. Thus, our system combines the advantages of powerful computation analysis with human's domain knowledge and cognitive abilities.

#### 2.3. Event sequence visualization

Since human daily movements are from locations to locations, movement rhythms can be considered as a sequence of events. In this sense, our work is considered to be closely related with event sequence visualization. Most of event sequence visualizations employ graph representation techniques, where each event is represented as a node and each transition between events is represented as an edge. In our case, locations of human movements can be represented as nodes, and movements between locations can be represented as edges. Then, movement rhythms can be considered as a set of event sequences, which can be generally visualized in two ways:

First, we can apply algorithms to map event sequences onto 2D plane using dimension reduction algorithms (Wei et al., 2012), or to mine frequent event sequences (Vrotsou et al., 2009). These approaches are effective for high-dimensional event sequence data. Nevertheless, since most daily movements consist of less than 5 journeys (see Fig. 1 for Singapore public transportation data), movement rhythms in this study do not exhibit so many dimensions.

Considering this, we employ the second approach, i.e., to aggregate event sequences to construct a hierarchical tree structure, and then visualize the tree structure. This approach can be identified in many fields, including patient medical history (Wongsuphasawat et al., 2011), eye movement traces (Tsang et al., 2010), and public transportation mobility (Zeng et al., 2014). Many techniques have also been developed to simplify tree structure (Monroe et al., 2013) and sort layout to improve legibility (Wongsuphasawat and Gotz, 2012). More recently, a visual interface applying data anonymization operations on Sankey diagram-like visualization has been designed to preserve privacy in event sequence data (Chou et al., 2016).

This approach firstly constructs a representative tree structure that can effectively organize all event sequences. A typical tree construction algorithm has been described in LifeFlow (Wong-suphasawat et al., 2011), where the algorithm starts from the root node and iteratively groups the events into the same category until the leaves. However, we find that this method may over-aggregate event sequences, and thus lead to wrong and missing information (Section 5.5). To overcome this issue, we devise a new rhythm tree construction mechanism (Section 4.4).



Fig. 1. Percentages of number of journeys in the Singapore public transportation data on weekdays and weekends, respectively.

#### 3. Overview

This section firstly introduces relevant terminologies in transportation, and then describes the input movement data. After that, we summarize a set of analytical tasks, followed by the system overview.

#### 3.1. Terminologies

Here, we explain some basic concepts employed in this work to facilitate the discussion:

- A *journey* refers to the movement taken by a person from her/his origin to destination.
- A *stay* refers to the stay of a person at a location between two consecutive journeys. The person can do certain activities during the stay, e.g., working or shopping.
- A *itinerary* is a sequence of daily movements and activities of a person, which can consist of multiple journeys and stays.

For instance, considering an employee's home  $\rightarrow$  work  $\rightarrow$  home movements, they comprise the *itinerary* of the employee. The movements of home  $\rightarrow$  work, and work  $\rightarrow$  home are two journeys. The activities performed at the working place is considered as a *stay*. Notice we also consider the periods before home  $\rightarrow$  work and after work  $\rightarrow$  home as stays, where the stay after work  $\rightarrow$  home is specifically denoted as *itinerary* end.

#### 3.2. Data description

Our interactive visual analytics system is applied to the following two movement datasets:

Singapore Public Transportation Data: The data is a one-week passenger movement data over the public transportation in Singapore, including both subway and bus rides. When a passenger makes a ride, the system will record various information, including anonymous card ID, journey ID, tap-in/-out times, tap-in/out stops, etc. If two or more rides are happening within 30 min consecutively, the system will assign the same journey ID to these rides. By ordering these rides based on their tap-in times, we can rebuild a *journey*. By referring to the card ID, we can group these journeys and order them based on their journey starting times, and thus the interval between two consecutive journeys forms a *stay*. These journeys and stays further make up an *itinerary* of the passenger.

In total, there are over 30 million rides made by  $\sim$ 1.8 million individual passengers over the week, with each passenger takes  $\sim$ 2.3 journeys on average every day. Besides, we have geographic information of  $\sim$ 4.8k subway and bus stops, and thus we can retrieve the movement origins and destinations by referring to the recorded tap-in/-out stops.

The MIT Reality Mining Data: The project was conducted on 94 subjects at MIT Media Laboratory from September 2004 to June 2005. Each subject was given a cell phone for tracking their communication, proximity, location and other information. The location information is indicated through sequence number of cellular towers, which has 100–200 m accuracy in urban areas. In this work, we select the locations and times of individual subject, and organize them into *itineraries*. If locations in two consecutive time stamps are the same (the same cellular tower id), we put them together and hence form a *stay* event. A transition between two consecutive *stays* is considered as a *journey*. In total, we extract ~11 thousand *itineraries* from over three million records in the dataset.

#### 3.3. Analytical tasks

Fig. 1 presents detailed percentages of the number of journeys on weekdays and weekends extracted from the Singapore public transportation data. From the figure, we can see that most itineraries consist of less than 5 journeys, and there are not much differences between weekdays and weekends. For instance, on both weekdays and weekends, there are  $\sim$ 50% passengers making two journeys. Nonetheless, transportation researchers would like to explore more details and find differences between these movements.

In discussions with a group of transportation researchers, we find some frequent questions, e.g., "what movement rhythms can be frequently found?", "what is the percentage of a specific movement rhythm, e.g.,  $A \rightarrow B \rightarrow A$ ?", "what differences exist between weekday and weekend?", etc. Based on these questions, we identified a family of analytical tasks:

- *T1 Frequent movement rhythms*: The experts would like to grasp an overview of frequent movement rhythms: what are the frequent movement rhythms? what is the percentage of each movement rhythm?
- *T2* Spatial movement distribution: Movement distribution in the spatial dimension is always of interest when studying movement data. In particular, the experts would like to know the origins and destinations of journeys, i.e., where are the journeys originated from & ended at? How many journeys are originated from or ended at a specific location?
- *T3 Temporal movement flows:* The experts would also like to explore the movement flows in the temporal dimension: How long do people stay at a location? How much time is needed to travel between two locations?

In addition, it would be necessary that our system allows for interactive filtering of movements over space and time, and then information for *T*1, *T*2 & *T*3 will be updated.

#### 3.4. System pipeline

Fig. 2 shows the pipeline of our system. The system starts with data modeling phase (Section 4). To enable interactive filtering of movements over space and time, we firstly index all journeys in spatial and temporal dimensions. We also identify movement rhythms for all itineraries in this phase. These two steps are performed offline when we load the movement data.

In the second phase, users can perform interactive visual exploration of movement rhythms (Section 5). Here, our interface presents three coordinated views: *Rhythm Statistics View* presents statistic overview of frequent movement rhythms (*T1*), *Rhythm Density View* plots the movement origin and destination distribution in spatial dimension (*T2*), and *Rhythm Sequence View* presents the movements traveling times and stay durations in temporal dimension (*T3*). The views complement each other and work together to support various analytical tasks. A series of spatial and



Fig. 2. System pipeline: In the data modeling phase, we perform spatial and temporal indexing, movement rhythms identification, and rhythm tree construction. In the visual exploration phase, three coordinated views are linked to support the analytical tasks.

temporal query techniques have also been integrated. Details of this phase can be found in Section 5.

#### 4. Data model

In order to support the identified analytical tasks, we perform the following steps to model the movement data. In Step 1 *Spatial and Temporal Indexing*, we build a hashing-based indexing method to support interactive filtering of the itineraries over space and time, specifically for Singapore public transportation data. After that, we map the filtered itineraries to certain movement rhythms in Step 2 *Movement Rhythm Identification*. Lastly, in Step 3 *Rhythm Tree Construction*, we organize all identified movement rhythms into a hierarchical tree structure for visualization.

#### 4.1. Spatial and temporal indexing

The Singapore public transportation data contains too many passenger movements (over 30 million rides in one week). To support interactive spatial and temporal queries against user-defined time period  $\Delta t$  and area of interest *AOI*, we use a hashing-based indexing method rather than scanning through every itinerary in the massive movement data.

First, to filter relevant itineraries within  $\triangle t$ , we partition a day from 06:00 to 24:00 (over 99% journeys happen during this period) into 15-minutes time bins. Here, we choose 15-minutes as the minimal time scale, driven by both a common practice in transportation and a recent study (Zhong et al., 2016), which shows that movement regularity dramatically decreases when the temporal scale is less than 15 min. Hence, we have 72 time bins for each day, and 504 (72×7) time bins in total for a week.

Second, to filter relevant trajectories starting from or ending at *AOI*, we create two hash tables with journey IDs as keys, and starting and ending times as values at each stop for each time bin. Notice here we index journeys instead of itineraries, as transportation researchers would like to explore movement origins and destinations at each sequence of movement rhythms; see Section 6.2 for an example. In the end, we have  $\sim 4.8k \times 504 \times 2$  hash tables in total. In the query time, we firstly search for time bins within  $\Delta t$ , and stops within *AOI*, and then filter journeys based on user-defined interactions.

#### 4.2. Movement model

We can model an itinerary *I* as a sequence of  $n \in \mathbb{N}$  stays or n-1 journeys:

$$I := S_1 \to S_2 \to \dots \to S_n, \quad or \tag{1a}$$

$$I := J_1 \to J_2 \to \dots \to J_{n-1},\tag{1b}$$

where  $J_i := S_i \rightarrow S_{i+1}$ . From the movement data, we can associate each  $J_i$  with two locations and two timestamps:

$$J_i \coloneqq (l_{o_i}, t_{o_i}) \to (l_{d_i}, t_{d_i}) \tag{2}$$

where  $(l_{o_i}, t_{o_i})$  represents the journey's origin location and starting time, and  $(l_{d_i}, t_{d_i})$  represents destination location and ending time.



**Fig. 3.** Three cases when we want to assign a region label to a new stop  $l_{i+1}$ : *separate*, *split* and *merge*.

In the Singapore public transportation data,  $l_{o_i} \otimes l_{d_i}$  belong to the input subway/bus stops in the public transportation system, while in the MIT reality mining data,  $l_{o_i} \otimes l_{d_i}$  refer to the cell tower id. *t* follows the rule of:

$$t_{o_i} < t_{d_i} < t_{o_{i+1}}, \quad \forall i \in \mathbb{N} : 1 \le i < n-1.$$

Hence a stay S<sub>i</sub> can be modeled as:

$$S_{i} := (l_{d_{i-1}}, t_{d_{i-1}}) \to (l_{o_{i}}, t_{o_{i}}), \quad \forall i \in \mathbb{N} : 1 < i < n$$
(3)  
and specifically for  $i = 1$  and  $i = n$ :  
$$S_{1} := (l_{o_{1}}, 06 : 00) \to (l_{o_{1}}, t_{o_{1}}) and$$
$$S_{n} := (l_{d_{n-1}}, t_{d_{n-1}}) \to (l_{d_{n-1}}, 24 : 00).$$

#### 4.3. Movement rhythm identification

*Region Mapping:* Notice that a stay should be happening in a region. In the Singapore public transportation data, the locations refer to public transportation stops. It is typical that  $l_{d_{i-1}}$  and  $l_{o_i}$  may not be the same stop, but rather located close to each other. Considering this, we map the bus stops into regions first for the Singapore public transportation data.

$$S_i := (R_i, t_{d_{i-1}} \to t_{o_i}), \quad \forall i \in \mathbb{N} : 1 \le i \le n.$$

$$\tag{4}$$

Here, we consider an area within a 10-minutes walking distance at average 5 km/h speed around the destination stop  $l_{d_{i-1}}$  as the region  $R_i$ . This distance is selected since the max ideal stop spacing is 800 m (approx. 10 min × 5 km/h) in an urban environment (Department of Transport and Main Roads, Queensland, 2016).

For each itinerary, when we already have  $\mathbb{R} := \{R_1, R_2, \ldots, R_i\}$  regions labeled, and to find the region for a new location  $l_{i+1}$ , we may encounter the following conditions as illustrated in Fig. 3. Here, we firstly find the 10-min. at 5 km/h walkable region around  $l_{i+1}$ , and mark it as  $R_{temp}$  (green circle).

- Separate: If  $l_{i+1}$  is located outside  $R_i$  and  $R_{temp}$  does not overlap with  $R_i$ , we label  $R_{temp}$  as  $R_{i+1}$ , mark it as the surrounding area of  $l_{i+1}$ , and add  $R_{i+1}$  into  $\mathbb{R}$ .
- *Split*: If  $l_{i+1}$  is located outside  $R_i$  but  $R_{temp}$  overlaps with  $R_i$ , we firstly split  $R_i$  and  $R_{temp}$  using Voronoi tessellation (the dashed line), and then label  $R_{temp}$  as the surrounding region  $R_{i+1}$  of  $l_{i+1}$ , and lastly add  $R_{i+1}$  into  $\mathbb{R}$ .
- Merge: If l<sub>i+1</sub> is located inside R<sub>i</sub>, we firstly update R<sub>i</sub> by merging it with R<sub>temp</sub>, and then we label R<sub>i</sub> as the surrounding region of l<sub>i+1</sub>.



**Fig. 4.** Illustrations of movement rhythms of a itinerary consisting of two journeys: (a)  $A \rightarrow B \rightarrow A$ , (b)  $A \rightarrow B \rightarrow C \rightarrow A$ .



**Fig. 5.** Comparing the tree construction differences between (a) LifeFlow and (b) our approach: our approach employs more constraints on the aggregation, and hence can preserve more details.

We do the above checking against all regions  $R \in \mathbb{R}$ . After we do the labeling for all stops, we can substitute Eq. (4) into Eq. (1a), and we can model an itinerary as:

$$I := (R_1, 06: 00 \to t_{o_1}) \to (R_2, t_{d_1} \to t_{o_2}) \to \dots$$
  
 
$$\to (R_n, t_{d_{n-1}} \to 24: 00), \quad \forall i \in \mathbb{N} : 1 \le i \le n$$

If two consecutive stays  $S_i := (R_i, t_{d_{i-1}} \to t_{o_i}), S_{i+1} := (R_i, t_{d_i} \to t_{o_{i+1}})$  have the same region label, we will merge them together as one stay  $S_{merge} := (R_i, t_{d_{i-1}} \to t_{o_{i+1}})$ .

*Rhythm Identification:* After this, we can replace  $R_i$  with a character A - 1 + i, and then the itinerary can be denoted as  $A \rightarrow B \rightarrow ... \rightarrow X$ . The final sequence of characters will be considered as the movement rhythm of the itinerary. For instance, the itinerary illustrated in Fig. 4(a) will be identified as  $A \rightarrow B \rightarrow A$  movement rhythm.

Notice that in the Singapore public transportation data, the origin stop of a successive journey  $l_{o_{i+1}}$  may not be in the same region  $R_i$  of the destination stop  $l_{d_i}$  of previous journey. In this case, we simulate an additional journey J' from  $l_{d_1}$  to  $l_{o_2}$ . Traveling time  $t'_{travel}$  of J' is interpolated as the average traveling time of movements from green to yellow region, and stay durations in the green and yellow regions are interpolated by average stay durations of movements at the two regions versus  $t_{o_2} - t_{d_1} - t'_{travel}$ . In this way, the itinerary illustrated in Fig. 4(b) will be identified as  $A \rightarrow B \rightarrow C \rightarrow A$ .

This step is performed when our system loads the movement data, then our system assigns a string label of the identified movement rhythm to each itinerary.

#### 4.4. Rhythm tree construction

Fig. 5 (top) illustrates four itineraries, where the nodes represent people stay at locations, and arrows represent people travel from one location to another. To simplify the discussion, we assume all itineraries begin at the same time, and stay at locations for the same time period, while traveling times from locations to locations are different. For instance, traveling times from *A* to *B* in the first two sequences are shorter in the top two itineraries than those in the bottom two itineraries.

Fig. 5(a) presents the tree construction algorithm described in LifeFlow (Wongsuphasawat et al., 2011). Here, since all movements start from A, and go to B, LifeFlow will group all itineraries at the first two sequences. Then at the third sequence, the top two itineraries going to A will be grouped together, while the bottom two going to C form another group. Here, since all four movements are grouped together at the first two sequences, the visualization will present averaged traveling times from A to B for all these four movements. Nonetheless, such aggregation misses traveling time difference from A to B between the top two itineraries and bottom two itineraries.

This situation is quite common in reality. For instance, let us assume two employee movements, EM1: home  $\rightarrow$  work  $\rightarrow$  home and EM2: home  $\rightarrow$  work  $\rightarrow$  lunch  $\rightarrow$  work  $\rightarrow$  home. We can imagine that stay duration at work location of EM1 is ~8 h, while EM2 stays at work location for ~3 h before lunch. In this sense, we should not group the second sequence work in EM1 and EM2.

To address this problem, we employ the association rule concept from the Apriori algorithm (Agrawal and Srikant, 1994), and devise a rhythm tree construction algorithm that works as follows:

- 1. Add a \$ symbol at the end of each movement rhythm, e.g., the top two movement rhythms in Fig. 5 (top) become  $I_1 := A \rightarrow B \rightarrow A \rightarrow \$$ , and  $I_2 := A \rightarrow B \rightarrow A \rightarrow C \rightarrow \$$ .
- 2. Associate an event with its successive event to form an association rule till we come to the \$\$ symbol, e.g., the top two rhythms become  $I_1 := (A, B) \rightarrow (B, A) \rightarrow (A, $) \rightarrow $$, and <math>I_2 := (A, B) \rightarrow (B, A) \rightarrow (A, C) \rightarrow (C, $) \rightarrow $$.$
- Aggregate two event sequences iteratively till we come to a difference or the end of one sequence, e.g., we can group the first two sequences of *I*<sub>1</sub> and *I*<sub>2</sub>, and at the third sequence, *I*<sub>1</sub> is (*A*, \$) while *I*<sub>2</sub> is (*A*, *C*). Hence, the aggregation between *I*<sub>1</sub> and *I*<sub>2</sub> stops at the third sequence.

We do this aggregation for all movement rhythms. In the end, we can construct a hierarchical tree structure that organizes all itineraries. In this way, our algorithm is able to form two different groups for itineraries presented in Fig. 5 (top) at the second sequence, as illustrated in Fig. 5(b), and hence our visualization is able to present the traveling time differences.

#### 5. Visualization design

In this section, we first discuss the principles of our visualization design. Then, we describe three visualization modules together with interactions implemented in our system.

#### 5.1. Design rationale

After modeling the movement rhythms, we further derive a set of design rationales for the visualization design to meet the analytical tasks *T1* to *T3*.

- Overview+Details: Our system follows the "Overview first, zoom and filter, then details on demand "mantra (Shneiderman, 1996). First, an overview should be provided to give analysts a broad overview of movement rhythms over space and time. Then analysts can further explore the detailed information using filtering and selection.
- *Interactive Exploration:* Since movement rhythms are spatially and temporally dependent, our system should provide intuitive interactions to support analysts' demands on exploring movement rhythms in a specific time period and area of interest.



**Fig. 6.** *Rhythm Statistic View* presents percentages of 12 most frequent movement rhythms based on the common color scheme used by all the views (a). Glyphs are employed to indicate the corresponding movement patterns: (b)  $A \rightarrow B \rightarrow C \rightarrow A$  and (c)  $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$ .

• *Multi-perspective Analysis:* To accomplish the analytical tasks *T1–T3*, analysts need to probe the data from multiple perspectives. Our system should provide multiple views, and establish links between them to present the information in different dimensions, such that to enable an efficient multi-perspective analysis.

To support *Multi-perspective Analysis*, we design three linked views: *Rhythm Statistic View*, *Density View* and *Sequence View*. They are arranged in a main and two sub windows; see Fig. 9. Below, we describe the details of each view.

#### 5.2. Rhythm statistic view

For *T1*, we aim to overview frequent movement rhythms. During the offline processing, our system summarizes all movement rhythms over one week from the Singapore public transportation data, and sorts these rhythms in descending order. In the end, we identify 12 most frequent movement rhythms that sum up to 95% of all itineraries.

In the interactive visual exploration stage, our system presents *Rhythm Statistic View* as illustrated in Fig. 6. The view lists these 12 rhythms as rhythm 0 - 11, and sums up the remaining in *Others* category as rhythm 12. For each rhythm, the view presents a unique glyph on the top, with nodes and directed links to illustrate the movement pattern. The nodes are colored based on a common color scheme used by all views in our system (Fig. 6(a)). Specifically, red color is reserved for stays at *A*, blue for *B*, green for *C*, purple for *D*, golden for *E*, yellow for *F* and afterwards. Hence, the glyphs in Fig. 6(b) & (c) represent  $A \rightarrow B \rightarrow C \rightarrow A$  and  $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$  rhythms, respectively. Then, after users filter the itineraries, our system recomputes the percentages of each movement rhythm, and depicts the statistics as bar charts.

*Basic Interactions*: In Fig. 6, all glyphs' backgrounds are gray, meaning that all these rhythms are selected. Analysts can also specify only a few specific rhythms for comparison, and the deselected glyphs' backgrounds will be marked as white.

#### 5.3. Rhythm density view

For *T2*, we aim to present spatial distribution of journey origins and destinations. To accomplish this, our system presents *Rhythm Density View* as shown in Fig. 7. The view is basically a density map overlaid on top of a base map with road network shown as connected lines. Specifically, subway lines are colored according to the subway color scheme, e.g., green from east–west line and purple for north-west line in Singapore. The density map indicates the



Fig. 7. Rhythm Density View presents the journey origins/destinations in the spatial dimension.

number of journey origins/destinations in a given period, which is generated with kernel density estimation (KDE) (Silverman, 1986) that has been successfully applied in presenting movement distributions, e.g., Scheepens et al. (2011) and Slingsby and van Loon (2016). The movement distributions along journeys are not within the scope of this work, as they do not affect movement rhythm analysis.

In detail, after analysts specify  $\triangle t$  and *AOI*, we are able to filter a set of  $n \in \mathbb{N}$  stops  $S := \{s_1, s_2, \ldots, s_n\}$  within *AOI*, where  $s_i := (s_{x_i}, s_{y_i}) \in \mathbb{N} \times \mathbb{N}$ . Each  $s_i$  is also associated with a number  $v_i \in \mathbb{N}$  indicating the number of journeys starting from or ending at the stop during  $\triangle t$ . Then we compute the density at location  $l := (l_x, l_y)$  as

$$f(l) = \frac{1}{n} \sum_{i=1}^{n} K(\frac{|l-s_i|}{h}) \times v_i,$$
(5)

where  $|l-s_i|$  represents the Euclidean distance between l and  $s_i$ , i.e.,  $\sqrt{(s_{x_i} - l_x)^2 + (s_{y_i} - l_y)^2}$ . h is bandwidth fixed at 10-min. walking distance is at average 5 km/h speed. And K is a normal distribution kernel:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$
 (6)

After computing all the densities, we can get a maximum density value  $v_{max}$ , and then divide  $v_{max}$  into 5 exponentially divided ranges  $[0, v_{max}^{1/5}), [v_{max}^{2/5}, v_{max}^{2/5}), [v_{max}^{3/5}, v_{max}^{4/5}), [v_{max}^{4/5}, v_{max}]$ . Each range has a corresponding gray-scale color; see Fig. 7(b). Lastly, the density field is mapped into the five gray-scale colors based on their values, and thus makes up a density map for rendering.

*Basic Interactions*: Analysts can interactively select a stay sequence, and origin/destination, by clicking on a corresponding stay & "origin/destination" icons shown in Fig. 7(a). A combination of the stay sequence and origin/destination filters journeys' origins/destinations to visualize. For instance, for  $A \rightarrow B \rightarrow A$  movement rhythm, it consists of  $A \rightarrow B$  and  $B \rightarrow A$  journeys: when analysts specify "A" and "origin", our system will count the origin of  $A \rightarrow B$  journey; when analysts specify "A" and "destination", our system will count the destination of  $B \rightarrow A$  journey.

#### 5.4. Rhythm sequence view

For *T3*, we aim to present the temporal perspective of information about the movement rhythms. We design the *Rhythm Sequence View* as shown in Fig. 8(a) with the following visual elements:

 Timeline: To effectively present the traveling times and stay durations, a timeline on the bottom of the view is displayed



(a) Rhythm Tree Construction.

(b) Lifeflow Tree Construction.

**Fig. 8.** Visual comparison between *Rhythm Sequence View* constructed through (a) rhythm tree construction and (b) LifeFlow tree construction mechanisms: our algorithm is able to present more meaningful results. For instance, our algorithm (a) reveals that people of  $A \rightarrow B \rightarrow A$  rhythms go back to A at ~18:30, while LifeFlow (b) wrongly indicates that they go back at ~17:00.

to show the time period from 06:00 to 24:00. We draw a timeline for every 15 min, and a dashed line from bottom to top for every 3 h, to facilitate the visual examination and comparison of times.

- *Tree Structure:* The tree is represented as a flow map, where the nodes representing stays or itinerary ends at locations, and links representing journeys from locations to locations. Each stay node is depicted as a rectangle with its height proportional to itinerary volume and length proportional to the average time that the movements stay at the location. For itinerary ending nodes, we represent them as semi circles. Since journey traveling time is also a period of time as the stay duration, we also depict the links as rectangles at the stay nodes to keep consistence.
- Node Layout and Color: The root node is placed on the left end, with its height the same as that of the display view, and ranges from left side (i.e., 06:00) to average starting time. For its child nodes, we sort them in descending order based on their itinerary volumes, and arrange them from bottom to top. We repeat this process for all child nodes until reaching itinerary ending nodes. We then color all nodes according to the common color scheme as shown on the top of the figure. We also reserve a unique gray color for the links.
- Interactive Time Slider: We also provide an interactive time slider as a pink rectangle overlaid on the view. Analysts can change the querying time interval  $[t_{min}, t_{max}]$  with the following options: (1) dragging the start/finish slider left and right to change  $t_{min}/t_{max}$ , respectively; (2) dragging the whole slider left and right to change both  $t_{min} \otimes t_{max}$ ; and (3) double click on the slider to specify  $t_{min}$  to 06:00 and  $t_{max}$  to 24:00.

#### 5.5. Design alternative

A design alternative in this work is the generation of *Rhythm Sequence View.* As discussed in Section 4.4, we can employ the LifeFlow tree construction mechanism to generate a rhythm tree structure that organizes all the movement rhythms. Fig. 8(b) presents an example visualization generated using this approach, with the same spatio-temporal query parameters as in Fig. 8(a). By comparing them, we can easily observe that this alternative design can lead to inaccurate information.

First, by averaging stay durations at *B* in  $A \rightarrow B \rightarrow A$ ... and  $A \rightarrow B \rightarrow C$ ... rhythms, Fig. 8(b) shows that both rhythms stay at *B* from ~09:00 to ~16:30. Nevertheless, such aggregation misses the differences between the two groups of rhythms, as people of



**Fig. 9.** Our visual interface arranges the three views in a main window and two sub windows. Analysts can interactively switch between the views to explore the spatial & temporal perspectives, and statistics of the movement rhythms on demand. Each view supports a number of interactions to facilitate the exploration.

 $A \rightarrow B \rightarrow A...$  rhythms stay ~2.5 h longer at *B* than those of  $A \rightarrow B \rightarrow C...$  rhythms, as shown in Fig. 8(a).

Second, since  $A \rightarrow B \rightarrow A$  rhythm comprises of mostly employee movements on a working day, the aggregation makes a wrong impression that employees arrive home very early at ~17:00. In addition, the aggregation further adds on to incorrect temporal information for the movements after *B*. For instance, Fig. 8(b) shows that people of  $A \rightarrow B \rightarrow C \rightarrow A$  rhythms go back to *A* at ~20:00, which should be ~18:30 as shown in Fig. 8(a).

#### 5.6. User interactions

As illustrated in Fig. 9, our visual interface arranges the three views in a main window and two sub windows. Analysts can click on one sub window to switch the views in the main and the selected sub window. Apart from this and the basic interactions supported by each view, our system supports the following interactions:

*Spatial Filtering*: In the *Rhythm Density View*, analysts can filter itineraries starting from one or multiple regions, by using a lasso tool or selecting one from the administrative regions, see the pink regions in Fig. 9 for examples.

*Temporal Specification*: Analysts can also specify a time period for exploration by adjusting the interactive time slider implemented in the *Rhythm Sequence View*.

*Rhythm Selection*: A specific movement rhythm can be selected by clicking on the corresponding rhythm glyph in the *Rhythm* 



Fig. 10. Study 1: Analyzing statistics of frequent movement rhythms identified from all movements in the one-week Singapore public transportation data.

*Statistic View*, or by connecting the filtered regions sequentially in the *Rhythm Density View*, as illustrated in Fig. 9.

All these interactions are linked among the three views, to facilitate *Multi-perspective Analysis* of movement rhythms. For example, if analysts select a particular rhythm in *Rhythm Statistic View*, *Rhythm Density View* will only present origin/destination densities of the movements in that particular rhythm, and *Rhythm Sequence View* will also update correspondingly.

#### 6. Case studies

We have conducted four case studies with our system running on an Intel Core i7 2 2.8 GHz MacBook Pro with 16GB memory and an AMD Radeon R9 M370X graphics board. It takes about 20 min to index and identify rhythms for all movements in the Singapore public transportation data, and less than one minute for the MIT reality mining data. After that, our system supports interactive exploration.

#### 6.1. Analyzing frequent movement rhythms

With *Rhythm Statistic View*, our system provides the analysts an overview of the statistics about frequent movement rhythms (*T1*). Fig. 10 presents the statistics for the frequent movement rhythms identified from all itineraries in the Singapore public transportation data.

From the figure, we can firstly identify that 12 frequent movement rhythms contribute to over 95% of all movements. Secondly, by summarizing the movement rhythms based on the number of places visited (separated by the dashed line), we can notice that they can be approximated with an exponential distribution. Specifically, the rhythms with 2 places visited occupy ~64.5%, with 3 places ~23.8%, with 4 places ~6%, and so on. We can also discover that journeys of daily movements are limited, with ~90% of the movements having less than 3 journeys. These observations support the findings about human's daily mobility revealed in Schneider et al. (2013), where the authors identified 17 representative mobility motifs by analyzing survey and mobile phone data.

#### 6.2. Exploring spatial distribution of journey origins and destinations

In Study 2, we are interested in exploring the spatial distribution of journey origins and destinations (*T2*). In particular, we firstly select a particular movement rhythm,  $A \rightarrow B \rightarrow A$ , and specify the itineraries' starting time within 07:00–09:00. We choose this rhythm and time period since it correlates with the pendulum movement pattern, which describes movements of employees commuting between residence and work locations in a working day (Rodrigue et al., 2013).

Then, we explore origin and destination distributions of the first and second journeys in the *Rhythm Density Views*, as shown in Fig. 11. In particular, Fig. 11(a) & (b) presents origin and destination distributions of first journeys, i.e.,  $A \rightarrow B$ , while Fig. 11(c) & (d) presents these distributions of the second journeys, i.e.,  $B \rightarrow A$ . Notice that to keep consistency, we have fixed the maximum values in all four density maps the same as 19,956, but actual maximum volumes in Fig. 11(a) & (d) are only ~5000.

By carefully examining the figures, we can make the following observations. First, we find that Fig. 11(a) matches well with Fig. 11(d), while Fig. 11(b) matches well with Fig. 11(c). Hence, with the hypothesis of pendulum movements, we can consider that Fig. 11(a) & (d) presents the residence locations, while Fig. 11(b) & (c) presents work locations. Second, all figures show that origins and destinations are mostly following subway stations, showing the subway system plays an important role in public transportation; see the correspondence between the density hotspots and subway lines. Lastly, Fig. 11(b) & (c) shows a more centralized distribution of work locations, as more deep blue colors can be identified. Specifically, the areas highlighted in (e) & (f) show more densities in Fig. 11(b) & (c) than in Fig. 11(a) & (d). This observation is verified by some transportation researchers, as (e) is a central commercial area, and (f) is an industrial area.

## 6.3. Comparing temporal perspective of movement rhythms differences for different time periods

In Study 3, we compare the temporal perspective of movement rhythms differences, which is related to *T3*. Here, we firstly specify a morning period as 07:00–09:00, and afternoon period as 13:00–15:00, and then filter the itineraries starting in the Monday morning, Monday afternoon, Sunday morning, and Sunday afternoon. Lastly, we compare their corresponding *Rhythm Sequence Views* as illustrated in Fig. 12(a)–(d). To facilitate the comparisons, we scale the heights of root nodes based on their volumes.

By comparing the views between Monday and Sunday, i.e., comparing Fig. 12(a) & (b) with Fig. 12(c) & (d), we can find: first, on Monday, the volume of movements starting in the morning period (634,308) is much higher than that in the afternoon (143,992); whilst on Sunday, the volumes are nearly equal (213,827 vs. 196,276). This difference is likely caused by vast amount of employee movements in Monday morning. Second, these employees spend a longer duration at their work locations, as we can see the *B* nodes in Fig. 12(a) are  $\sim$ 2 h longer than that those in Fig. 12(c); whilst those in Fig. 12(b) & (d) are almost the same.

By comparing the views between morning and afternoon, i.e., comparing Fig. 12(a) & (c) with Fig. 12(b) & (d), we can further find that the percentages of movement rhythms change, whilst these differences are not obvious between Monday and Sunday. Specially, the percentage of  $A \rightarrow B$  rhythm has increased quite a lot in both afternoons, showing that more people starting their journeys in the afternoon do not go back. In addition, through these movements start later, they leave *B* at nearly the same times as these in Fig. 12(a) & (c). This indicates that evening peak traveling demand is not only caused by employee movements starting in the morning, but also by other movements starting in the afternoon.

#### 6.4. Exploring movement rhythms in the MIT reality mining dataset

In Study 4, we explore movement rhythms in the MIT reality mining dataset. As described in Section 3.2, the total number of individual itineraries is  $\sim$ 11K. However, there exist many stays with durations in short times, such as 10 s and less. We suspect the following two reasons: (1) locations generated by mapping cellular towers have low accuracy, and (2) the subjects moved frequently. To address this issue, we set a minimum stay duration of 10 min:



**Fig. 11.** Study 2: Exploring the spatial origin and destination distributions of journeys with their corresponding itineraries in the movement rhythm  $A \rightarrow B \rightarrow A$  and itineraries' starting time within 07:00–09:00: (a) & (b) present the origins and destinations of the first journeys  $A \rightarrow B$ , and (c) & (d) present those of the second journeys  $B \rightarrow A$ .



Fig. 12. Study 3: Comparing the temporal perspective of movement rhythm differences for different time periods: (a) Monday morning, (b) Monday afternoon, (c) Sunday morning, and (d) Sunday afternoon.

the stays with durations below the threshold are ignored. We also ignore itineraries with only one stay, to better compare with movement rhythms extracted from the Singapore public transportation data. In the end, we filter 2331 itineraries, and the corresponding *Rhythm Sequence View* is shown in Fig. 13.

Comparing with the movement rhythms in the Singapore public transportation data, we can identify many differences. From the figure, we can easily find some frequent rhythms, e.g.,  $A \rightarrow B$ ,  $A \rightarrow B \rightarrow C$ , and  $A \rightarrow B \rightarrow A$ . However, the rhythms have different percentages compared with these in the public transportation data. For instance,  $A \rightarrow B \rightarrow A$  occupies only 15%. Besides rhythm percentage differences, traveling times and staying durations also show different patterns. Apparently, all traveling times are much longer while stay durations are shorter than these in the public transportation data. For instance, it takes on average  $\sim 5$  h to travel from *A* to *B*, and people stay in *B* for only  $\sim 2$  h.

We suspect many reasons that lead to the differences: (1) The data in Singapore only records the passengers' movements through public transportation, while travels through other modes are ignored, such as taxi trips. (2) The movements recorded in the MIT reality mining data could be noisy (only 100–200 m accuracy), while mapping spatial information to public stops in the Singapore public transportation data is much more accurate. (3) The data cleaning approach employed in the MIT reality mining dataset, i.e., filter out stays with durations less than 10 min, is not very appropriate. (4) The MIT reality mining dataset records movements of only students and laboratory staff, while the public transportation data covers much more diverse populations in Singapore.

#### 7. Expert interviews

The research is motivated by discussions with a team of transportation researchers (denoted as *Experts A*). We also conducted one-on-one interviews with two independent experts. One of them



Fig. 13. Study 4: Exploring movement rhythms identified from the MIT reality mining dataset.

is from a research institute (*Expert B*) with a research focus on human mobility analysis, and the other (*Expert C*) had 3 years working experience on public transportation management in Singapore.

In the interviews, we firstly explained our interface design and visual encodings when our system is loading and modeling the data. We then demonstrated how our system works and showed them the case studies. Lastly, the experts explored the system by themselves. Each interview lasted for 40 min to 1 h, and their feedbacks are summarized as follows.

Visual design and interactions. Overall, all the experts appreciated our visual analytics system. They thought our visual designs are simple to follow, meanwhile informative for the analytical tasks. *Expert B* was very impressed by the *Rhythm Density View*, as most density maps he saw in mobility analysis papers are to "divide the city into cells and color each cell". Our density map "looks more visual appealing than these maps". The experts were especially impressed by the design of the *Rhythm Sequence View*, as it can "clearly present major movement patterns and stay durations".

They also agreed that the three views are well linked. *Experts A* emphasized that the ability to switch between different views can greatly enhance analysts' exploration of human rhythms. *Expert B* appreciated our system's capability of filtering movements over space and time, as it is "really helpful to explore details about human movements". Without these interactions, the movement rhythms presented in Study 3 are nearly impossible to identify.

*Limitations and improvements.* The experts also gave fruitful suggestions to improve our system. Through the development of our system, we had continuous discussions with *Experts A* and refined our designs based on their comments. Actually in the early design stage, we employed the alternative design as shown in Fig. 8(b), and *Experts A* immediately pointed out the problems. Besides, *Expert B* recommended to present temporal distribution of traveling times and stay durations in the *Rhythm Sequence View*, as such information can help "identify mobility outliers". Nonetheless, we consider this a common issue for all existing works that aggregate events when visualizing event sequence data. *Expert C* also suggested that the system may be more useful if we can incorporate land-use data in the analysis, e.g., residence, shopping. In this way, users can know the environments and better understand "why people travel between locations".

*Usability.* Our tool can be helpful for researches on human mobility analysis. From his experience, *Expert B* pointed out that "interactive exploration tool is really needed when analyzing human mobility." *Expert C* also agreed that our system is applicable in urban planning and transportation management. He confirmed the residence locations around subway lines in Study 2, as "Singapore aims to improve the convenience to travel through public transportation for all residents."

#### 8. Discussion

Applicability. Though this research focuses on movement data, we believe our system can be extended to explore any spatially distributed event sequence data, e.g., eye movement traces and patient medical history. In particular, our rhythm tree construction algorithm can be generalized for aggregating any event sequences by associating an event with *N* successive events. In LifeFlow (Wongsuphasawat et al., 2011), *N* equals to 0; in our work, *N* equals to 1. A larger *N* may be more appropriate in some other applications, such as genome sequence and chess game, where two and more sequential events have more meanings. In addition, from visualization perspective, we can achieve a multiscale aggregation of event sequences by adjusting *N*. In this sense, Fig. 8(a) can be considered as a detailed view of Fig. 8(b).

*Future Work.* There are multiple directions for future work. First, in the data modeling step, we index the movements spatially on stops, which costs too much memory and is not scalable. In the future, we plan to implement an advanced indexing mechanism, such as Ferreira et al. (2013), to improve the query efficiency. Second, as pointed out by *Expert C*, our system lacks semantic context of the environment, and thus cannot explain why people move between locations. Regarding this, we plan to incorporate land-use data in our analysis, and this will require advanced data mining techniques to automatically fuse the two datasets. Lastly, we found most of traveling times and stay durations follow normal distributions with mean values the same as the averaged times in the *Rhythm Sequence View*. In the future, we aim to explore new visual designs that can plot this information intuitively.

#### 9. Conclusion

In this paper, we present a visual analytics system designed to facilitate transportation researchers' work in exploring human daily movement rhythms. We show how to identify movement rhythms from raw movement data, and then depict that movement rhythms can actually be modeled as event sequences. Thus, we can employ event sequence visualizations to present the temporal aspect of information. However, these visualizations can overaggregate the movement rhythms. To address this problem, we devise a new tree construction algorithm based on the association rule concept, which can preserve more details. The algorithm can be applied to any event sequence data, and we also show that the algorithm can be generalized to achieve a level-of-detail visualization.

Based on these, we develop an interactive visual interface to support the various analytical tasks. We use the system to conduct four case studies on studying massive one-week public transportation data in Singapore and the MIT reality mining dataset. The studies show that human daily movements with public transportation can be mostly described with 12 frequent movement rhythms, and the movement rhythms exhibit spatial and temporal variations. The positive expert feedbacks show that an interactive visual analytics system is helpful for domain-specific analysis; meanwhile, it is clear that in-depth domain knowledge could be very helpful for visual design, such as the identification of the over-aggregation problem.

#### Acknowledgments

The research was conducted at the Future Cities Laboratory at the Singapore-ETH Centre, which was established collaboratively between ETH Zurich and Singapore's National Research Foundation (FI 370074016) under its Campus for Research Excellence and Technological Enterprise programme. Chi-Wing Fu is supported by the CUHK strategic recruitment fund and direct grant (4055061). Kwan-Liu Ma is supported in part by the U.S. National Science Foundation.

#### References

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules. In: Proc. VLDB, pp. 487–499.

- Andrienko, N., Andrienko, G., 2012. Visual analytics of movement: An overview of methods, tools and procedures. Info. Vis. 12 (1), 3–24.
- Andrienko, G., Andrienko, N., Bremm, S., Schreck, T., Von Landesberger, T., Bak, P., Keim, D., 2010. Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. Comput. Graph. Forum 29 (3), 913–922.
- Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S.I., Wachowicz, M., 2008. Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research. Info. Vis. 7, 173–180.
- Bowman, J.L., Ben-Akiva, M.E., 2000. Activity-based disaggregate travel demand model system with activity schedules. Transp. Res. A: Pol. 35, 1–28.
- Chen, J., Shaw, S.-L., Yu, H., Lu, F., Chai, Y., Jia, Q., 2011. Exploratory data analysis of activity diary data: a spacetime GIS approach. J. Transp. Geogr. 19 (3), 394–404.
- Chou, J.-K., Wang, Y., Ma, K.-L., 2016. Privacy preserving event sequence data visualization using a Sankey diagram-like representation, in: Proc. SIGGRAPH ASIA 2016 Symp. Vis., 3002153, pp. 1:1–1:8.
- Department of Transport and Main Roads, Queensland, 2016. Public Transport Infrastructure Manual: Bus Stop Infrastructure. pp. 1–56.
- Eagle, N., Pentland, A., 2006. Reality mining: sensing complex social systems. Pers. Ubiquitous Comput. 10 (4), 255–268.
- Ferreira, N., Poco, J., Vo, H.T., Freire, J., Silva, C.T., 2013. Visual exploration of big spatio-temporal urban data: A study of New York city taxi trips. IEEE Trans. Vis. Comput. Graphics 19 (12), 2149–2158.
- Guo, H., Wang, Z., Yu, B., Zhao, H., Yuan, X., 2011. TripVista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In: Proc. IEEE Pacific Vis. Symp., pp. 163–170.
- Hägerstrand, T., 1970. What about people in regional science?. Pap. Reg. Sci. Assoc. 24 (1), 6–21.
- Hurter, C., Tissoires, B., Conversy, S., 2009. FromDaDy: Spreading aircraft trajectories across views to support iterative queries. IEEE Trans. Vis. Comput. Graphics 15 (6), 1017–1024.
- Kamruzzaman, M., Hine, J., Gunay, B., Blair, N., 2011. Using GIS to visualise and evaluate student travel behaviour. J. Transp. Geogr. 19 (1), 13–32.
- Kapler, T., Wright, W., 2004. GeoTime Information Visualization. In: Proc. IEEE Symp. Info. Vis., pp. 25–32.
- Krüger, R., Thom, D., Wörner, M., Bosch, H., Ertl, T., 2013. TrajectoryLenses –A set-based filtering and exploration technique for long-term trajectory data. Comput. Graph. Forum 32 (3pt4), 451–460.
- Monroe, M., Rongjian, L., Hanseung, L., Plaisant, C., Shneiderman, B., 2013. Temporal event sequence simplification. IEEE Trans. Vis. Comput. Graphics 19 (12), 2227–2236.
- Primerano, F., Taylor, M.P., Pitaksringkarn, L., Tisato, P., 2008. Defining and understanding trip chaining behaviour. Transp. 35 (1), 55–72.

- Rodrigue, J.-P., Comtois, C., Slack, B., 2013. The Geography of Transport Systems, third ed., Routledge, New York, p. 416.
- Scheepens, R., Hurter, C., van de Wetering, H., van Wijk, J.J., 2016. Visualization, selection, and analysis of traffic flows. IEEE Trans. Vis. Comput. Graphics 22 (1), 379–388.
- Scheepens, R., Willems, N., van de Wetering, H., Andrienko, G., Andrienko, N., van Wijk, J.J., 2011. Composite density maps for multivariate trajectories. IEEE Trans. Vis. Comput. Graphics 17 (12), 2518–2527.
- Schneider, C.M., Belik, V., Couronn, T., Smoreda, Z., Gonzlez, M.C., 2013. Unravelling daily human mobility motifs. J. R. Soc. Interface 10 (84).
- Shneiderman, B., 1996. The eyes have it: a task by data type taxonomy for information visualizations. In: Proc. IEEE Symp. Visual Lang., pp. 336–343.
- Silverman, B.W., 1986. Density estimation for statistics and data analysis. In: Monographs on Statistics and Applied Probability. Chapman and Hall, London.
- Slingsby, A., van Loon, E., 2016. Exploratory visual analysis for animal movement ecology. Comput. Graph. Forum 35 (3), 471–480.
- Sun, G., Liang, R., Qu, H., Wu, Y., 2017. Embedding spatio-temporal information into maps by route-zooming. IEEE Trans. Vis. Comput. Graphics 23 (5), 1506–1519.
- Tominski, C., Schumann, H., Andrienko, G., Andrienko, N., 2012. Stacking-Based visualization of trajectory attribute data. IEEE Trans. Vis. Comput. Graphics 18 (12), 2565–2574.
- Tsang, H.Y., Tory, M., Swindells, C., 2010. eSeeTrack Visualizing sequential fixation patterns. IEEE Trans. Vis. Comput. Graphics 16 (6), 953–962.
- Vrotsou, K., Johansson, J., Cooper, M., 2009. ActiviTree: Interactive visual exploration of sequences in event-based data using graph similarity. IEEE Trans. Vis. Comput. Graphics 15 (6), 945–952.
- Wei, J., Shen, Z., Sundaresan, N., Ma, K.-L., 2012. Visual cluster exploration of web clickstream data. In: Proc. IEEE VAST, pp. 3–12.
- Wongsuphasawat, K., Gomez, J.A.G., Plaisant, C., Wang, T.D., Taieb-Maimon, M., Shneiderman, B., 2011. LifeFlow: visualizing an overview of event sequences. In: Proc. ACM SIGCHI, pp. 1747–1756.
- Wongsuphasawat, K., Gotz, D., 2012. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. IEEE Trans. Vis. Comput. Graphics 18 (12), 2659–2668.
- Zeng, W., Fu, C.-W., Müller Arisona, S., Erath, A., Qu, H., 2014. Visualizing mobility of public transportation system. IEEE Trans. Vis. Comput. Graphics 20 (12), 1833–1842.
- Zeng, W., Fu, C.-W., Müller Arisona, S., Erath, A., Qu, H., 2016a. Visualizing waypoints-constrained origin-destination patterns for massive transportation data. Comput. Graph. Forum 35 (8), 95–107.
- Zeng, W., Fu, C.-W., Müller Arisona, S., Schubiger, S., Burkhard, R., Ma, K.-L., 2016b. A visual analytics design for studying crowd movement rhythms from public transportation data. In: Proc. ACM SIGGRAPH ASIA 2016 Symp. Vis., pp. 4:1–4:8.
- Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., Schmitt, G., 2016. Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. PLoS One 11 (2), e0149222.