A Visual Analytics Design for Studying Crowd Movement Rhythms from Public Transportation Data

Wei Zeng[1]* Chi-Wing Fu[2] Stefan Müller Arisona[1,3] Simon Schubiger[1,3] Remo Burkhard[1] Kwan-Liu Ma[4]
[1] Future Cities Laboratory, ETH Zurich [2] The Chinese University of Hong Kong
[3] University of Applied Sciences and Arts Northwestern Switzerland FHNW [4] University of California-Davis

Abstract

Human lives involve various daily movements in a space-time context, which exhibit high regularity that typically forms circadian rhythms. Understanding the rhythms for human daily movements of massive crowds can be highly beneficial for a variety of applications, such as traffic demand management and urban planning. In this paper, we propose an interactive visual data analysis approach, which provides not only quantitative analyses, including frequent human movement rhythms identification, but also visualization supported with a family of user interactions. We also devise a set of interactive visual query methods for users to easily explore the movement rhythms over space and time. Case studies with real-world massive urban public transportation data in Singapore, and interviews with transportation researches are carried out to demonstrate the effectiveness and usefulness of our system.

Keywords: movement rhythm, event sequence, visual analytics

Concepts: •**Human-centered computing** → **Visual analytics;**

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* → *work* → *home* and *home* → *school* → *tuition* → *home*. The movements can be further generalized as network motifs by abstracting the activity information [Schneider et al. 2013], e.g., *home* → *work* → *home* can be generalized as $A \rightarrow B \rightarrow A$, and *home* → *school* → *tuition* → *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].

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, from the perspectives of traffic management and urban planning, only to study the movements of massive crowds makes sense, rather than that of single or small groups of people. A direct plot of all the massive crowds' movements can easily lead to visual clutter. Second, the movements of massive crowds 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, the movements of massive crowds 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 this work, we firstly employ an efficient movement modeling method to identify movement rhythms based on the movement's spatial and temporal characteristics (Section 4.1). All movement rhythms are organized into a hierarchical tree structure with a new tree construction algorithm (Section 4.2) devised from the association rule concept. 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* overviewing the statistics of frequent movement rhythms (Section 5). In the end, we applied our approach to the study of real-world massive urban public transportation data in Singapore (Section 6), and conducted interviews with transportation researches 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 the massive public transportation data in Singapore showing interesting human movement rhythms, especially the identification of 12 frequent movement rhythms.

2 Related Work

We review related researches in the following two relevant topics: movement data and event sequence visualization.

2.1 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.

*e-mail:zeng@arch.ethz.ch

SA '16 Symposium on Visualization ,, December 05-08, 2016, , Macao
ISBN: 978-1-4503-4547-7/16/12
DOI: http://dx.doi.org/10.1145/3002151.3002152

Visual display: We can design novel visual structures to reveal movement patterns in the data, for example, e.g., waypoints-constrained OD view [Zeng et al. 2015]. 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. 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].

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].

This work considers all these techniques. We first adapt 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.

2.2 Event Sequence Visualization

Generally, most of event sequence visualizations adapt graph representation techniques, where each event constitutes a node and each transition between events constitutes an edge. In our case, the locations of human movements can be represented as nodes, and movements between locations can be represented as edges. Then, the 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 the event sequences onto the 2D plane using dimension reduction algorithms [Wei et al. 2012], or to mine frequent event sequences for visualization [Vrotsou et al. 2009]. These approaches are effective for high-dimensional event sequence data. Nevertheless, since most movements in the input data consist of less than 5 journeys (see Figure 1), the movement rhythms in this study do not exhibit so many dimensions.

Considering this, we adapt the second approach, i.e., to aggregate the 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 traffic incidents [Guerra-Gómez et al. 2011]. Many techniques have also been proposed to simplify tree structure [Monroe et al. 2013] and sort layout to improve legibility [Wongsuphasawat and Gotz 2012].

This approach firstly constructs a representative tree structure that can effectively organize all the event sequences. A typical tree construction algorithm has been described in LifeFlow [Wongsuphasawat 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 found that this method may over-aggregate the event sequences, and thus lead to wrong and missing information (Section 5.4). To overcome this issue, we devise a new rhythm tree construction mechanism (Section 4.2).

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 a system overview.

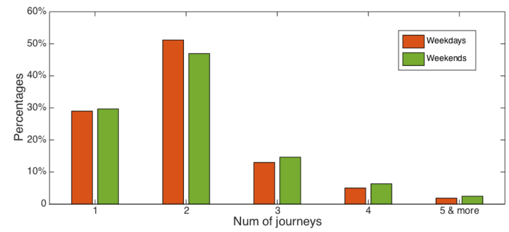


Figure 1: *Percentages of number of journeys in the public transportation data on weekdays and weekends, respectively.*

3.1 Terminologies

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

- A *trip* is a single public transportation ride, including both subway and bus, taken between two stops.
- A *journey* consists of a single or multiple trips, successively taken by a passenger from her/his origin to destination.
- A *stay* refers to the stay of a passenger at a location between two consecutive journeys. The passenger can do certain activities during the stay, e.g., working, shopping, studying, etc.
- A *itinerary* is a sequence of daily movements and activities of a passenger, which can consist of multiple journeys and stays.

3.2 Data Description

The movement data employed in this work is a one-week passenger movement data over the public transportation in Singapore, including both subway and bus rides. When a passenger makes a *trip*, the system will record various information about it, including anonymous card ID, journey ID, tap-in/-out times, tap-in/-out stops, etc. If two or more trips are happening within 30 minutes consecutively, the system will assign the same journey ID to these trips. By ordering these trips 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 trips made by $\sim$1.8 million individual passengers over the week, with each passenger takes $\sim$2.3 journeys on average every day. Figure 1 presents detailed percentages of the number of journeys on weekdays and weekends. 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.

3.3 Analytical Tasks

In discussions with a group of researchers who have been studying the Singapore public transportation data in the past five years, we found some frequent questions, e.g., "what movement rhythms can be frequently found in the transportation data?", "what is the percentage of a specific movement rhythm, e.g., $A \rightarrow B \rightarrow A$?", "what differences exist between weekday and weekend?", etc.

In summary, we identified a family of analytical tasks as follows:

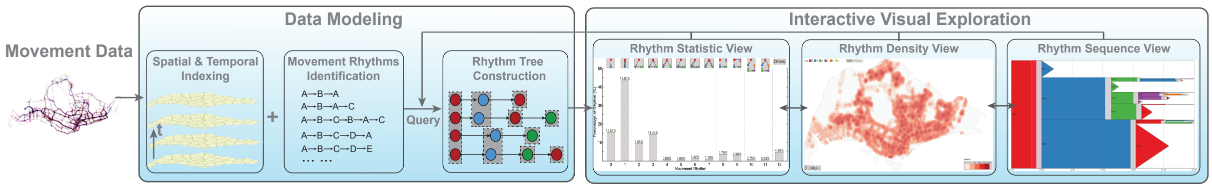


Figure 2: *System pipeline: In the data modeling phase, we perform spatial and temporal indexing, movement rhythms identification, and rhythm tree construction. In the visual exploration phase, three coordinated views are linked to support the analytical tasks.*

T1 Frequent movement rhythms: The experts would like to grasp an overview of frequent movement rhythms in the transportation data: what are the frequent movement rhythms? what is the percentage of each movement rhythm?

T2 Spatial movement distribution: The experts would like to know the origins and destinations of passenger journeys, i.e., where are the journeys originated from & ended at? How many journeys are originated from or ended at a location?

T3 Temporal movement flows: The experts would also like to explore the movement flows in temporal dimension: for how long do passengers 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 **T1**, **T2** & **T3** should be updated correspondingly (**T4**).

3.4 System Pipeline

Figure 2 shows the pipeline of our system. The system starts with the data modeling phase (Section 4). To enable interactive filtering of movements over space and time (**T4**), we firstly index all the passenger journeys in spatial and temporal dimensions. We also identify the 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 and analysis of movement rhythms (Section 5). The visual interface integrates three linked views: the *Rhythm Statistic View*, the *Rhythm Density View* and the *Rhythm Sequence View*, which complement each other and work together to support the various analytical tasks. A series of spatial and temporal query techniques have also been integrated.

4 Data Model

In order to support the identified analytical tasks, we perform the following steps to model the movement data.

4.1 Movement Rhythm Identification

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

$$I := S_1 \to S_2 \to \ldots \to S_n, \quad or \tag{1a}$$
$$I := J_1 \to J_2 \to \ldots \to J_{n-1}, \tag{1b}$$

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

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

where $(l_{o_i},\, t_{o_i})$ represents the journey's origin location and starting time, and $(l_{d_i},\, t_{d_i})$ represents the journey's destination location and ending time. l_{o_i} & l_{d_i} belong to the input subway and bus stops in the public transportation system, and t follows the rule of:

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

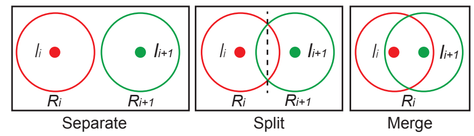


Figure 3: *Three cases when we want to assign a region label to a new stop l_{i+1}: separate, split and merge.*

Hence a stay S_i can be modeled as:

$$S_i := (l_{d_{i-1}},\, t_{d_{i-1}}) \to (l_{o_i},\, t_{o_i}), \quad \forall i \in \mathbb{N} : 1 < i < n \tag{3}$$

and specifically for $i = 1$ and $i = n$:

$$S_1 := (l_{o_1},\, 06:00) \to (l_{o_1},\, t_{o_1}) \textit{ and}$$
$$S_n := (l_{d_{n-1}},\, t_{d_{n-1}}) \to (l_{d_{n-1}},\, 24:00).$$

Notice that in reality, $l_{d_{i-1}}$ and l_{o_i} may not be the same stop, but rather located close to each other. Considering this, we should denote a stay that happens at a region, instead of two isolated stops:

$$S_i := (R_i,\, t_{d_{i-1}} \to t_{o_i}), \quad \forall i \in \mathbb{N} : 1 \leq i \leq n. \tag{4}$$

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

Identification: Notice that 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 Figure 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 its surrounding 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_{i+1} is located inside R_i, we firstly update R_i by merging it with R_{temp}, and then we label R_i as the surrounding region of l_{i+1}.

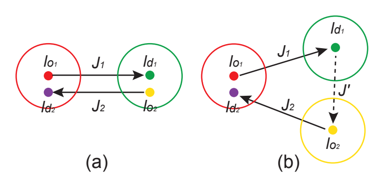


Figure 4: *Illustrations of movement rhythms of a itinerary consisting of two journeys: (a) $A \to B \to A$, (b) $A \to B \to C \to A$.*

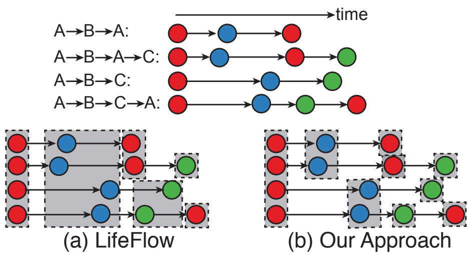


Figure 5: *Comparing the tree construction differences between LifeFlow (a) and our approach (b): our approach adopts more constraints on the aggregation, and hence can preserve more details.*

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

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

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

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

Notice that in reality, 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 the previous journey. In this case, we simulate an additional journey J' from l_{d_1} to l_{o_2}. The traveling time t'_{travel} of J' is interpolated as the average traveling time of movements from the green to yellow region, and the stay durations in the green and yellow regions are interpolated by the average stay durations of the movements at the two regions versus $t_{o_2} - t_{d_1} - t'_{travel}$. In this way, the itinerary illustrated in Figure 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 passenger itinerary.

4.2 Rhythm Tree Construction

Figure 5 (top) illustrates four itineraries, where the nodes represent passengers stay at locations, and arrows represent passengers 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 much shorter in the top two itineraries than those in the bottom two itineraries.

Figure 5(a) presents a tree construction algorithm described in LifeFlow [Wongsuphasawat et al. 2011]. Here, since all the movements start from A, and go to B, LifeFlow will group all the 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 the four movements are grouped together at the first two sequences, the visualization will present the averaged traveling times from A to B for all these four movements. Nonetheless, such aggregation misses the 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's 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 the stay duration at the *work* location of $EM1$ is ~8 hours, while $EM2$ stays at the *work* location for ~3 hours before lunch. In this sense, we should not group the second sequence *work* in $EM1$ and $EM2$.

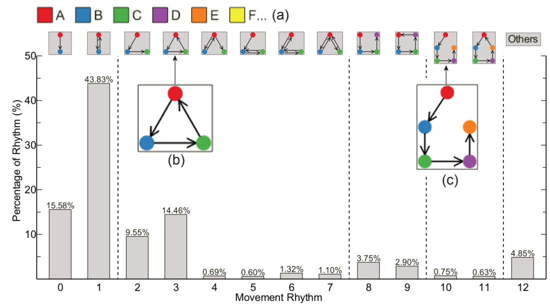


Figure 6: *Rhythm Statistic View presents percentages of 12 most frequent movement rhythms, with glyphs indicating the corresponding movement patterns: (b) $A \rightarrow B \rightarrow C \rightarrow A$ and (c) $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$. (a) shows the common color scheme used by all the views.*

To address this problem, we adapt 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 Figure 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 $I_2 := (A,\ B) \rightarrow (B,\ A) \rightarrow (A,\ C) \rightarrow (C,\ \$) \rightarrow \$$.

3. 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_1 and I_2, and in the third sequence, I_1 is $(A,\ \$)$ while I_2 is $(A,\ C)$. Hence, the aggregation between I_1 and I_2 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 the itineraries. In this way, our algorithm is able to form two different groups for itineraries presented in Figure 5 (top) at the second sequence, as illustrated in Figure 5 (b), and hence our visualization is able to present the traveling time differences.

5 Visualization Design

In this section, we describe three visualization modules together with interactions implemented in our system.

5.1 Rhythm Statistic View

In the interactive visual exploration stage, our system presents the *Rhythm Statistic View* as illustrated in Figure 6. The view lists these 12 rhythms as rhythm 0 - 11, and sums up the remaining in the *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 the views in our system (Figure 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 Figure 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.

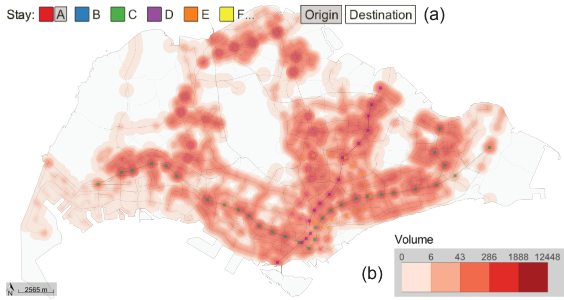


Figure 7: *Rhythm Density View presents the journey origins/destinations in the spatial dimension.*

5.2 Rhythm Density View

After analysts specify $\triangle t$ and *AOI*, we are able to filter a set of $n \in \mathbb{N}$ stops $S := \{s_1, s_2, ..., s_n\}$ 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

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

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 at average 5 km/h speed. And K is a normal distribution kernel:

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

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

5.3 Rhythm Sequence View

We design the *Rhythm Sequence View* as shown in Figure 8(a) with the following visual elements:

- *Horizontal Timeline:* A horizontal timeline on the bottom of the view is displayed to show the time period from 06:00 to 24:00. We draw a vertical short line for every 15 minutes, and a dashed line from bottom to top for every 3 hours, 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 equilateral triangles placed in a horizontal style with their edge length proportional to the ending itinerary volumes, whilst the horizontal length has no meaning. 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 the 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 the child nodes until reaching the itinerary ending nodes. We then color all the 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} & t_{max}; and 3) double click on the slider to change t_{min} to 06:00 and t_{max} to 24:00.

5.4 Design Alternative

A design alternative in this work is the generation of *Rhythm Sequence View*. As discussed in Section 4.2, we can adopt the LifeFlow tree construction mechanism to generate a rhythm tree structure that organizes all the movement rhythms. Figure 8(b) presents an example visualization generated using this approach, with the same spatio-temporal query parameters as in Figure 8(a). By comparing them, we can easily observe that this alternative design can lead to wrong and missing information.

First, by averaging stay durations at B in $A \rightarrow B \rightarrow A...$ and $A \rightarrow B \rightarrow C...$ rhythms, Figure 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 passengers of $A \rightarrow B \rightarrow A...$ rhythms stay ~2.5 hours longer at B than those of $A \rightarrow B \rightarrow C...$ rhythms, as illustrated in Figure 8(a).

Second, since the $A \rightarrow B \rightarrow A$ rhythm comprises of mostly employee movements on a working day, the aggregation makes an 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, Figure 8(b) shows that passengers of $A \rightarrow B \rightarrow C \rightarrow A$ rhythms go back to A at ~20:00, which should be ~18:30 as shown in Figure 8(a).

5.5 User Interactions

As illustrated in Figure 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:

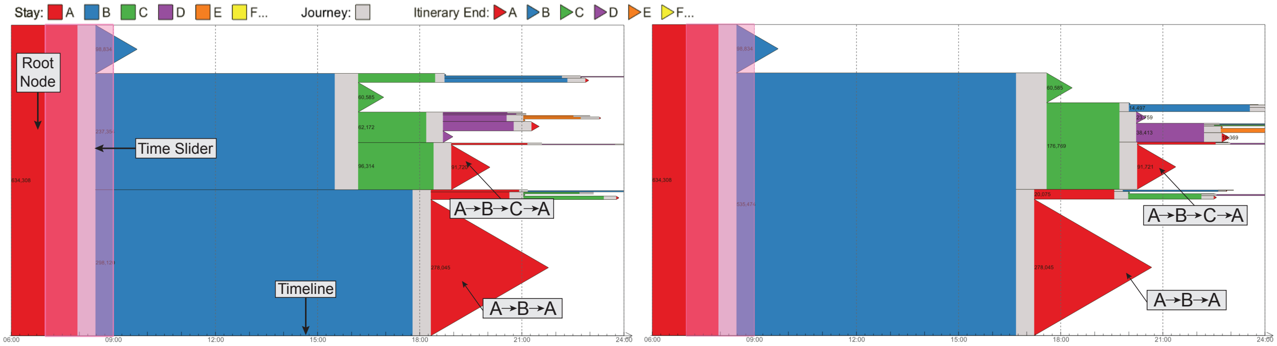
Spatial Filtering: In the *Movement Density View*, analysts can filter the itineraries starting from one or multiple regions, by using a lasso tool or selecting one from the administrative regions, see the pink regions in Figure 9 for examples.

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

6 Case Studies

We have conducted two case studies with our system running on an Intel Core i7 2 2.8GHz MacBook Pro with 16GB memory and an



(a) Rhythm Tree Construction (b) Lifeflow Tree Construction

Figure 8: *Visual comparison between Rhythm Sequence View constructed through our proposed (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 passenger 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.*

AMD Radeon R9 M370X graphics board. It takes about 20 minutes to index and identify rhythms for all the movements over one week, and then 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**). Figure 10 presents the statistics for the frequent movement rhythms identified from all passenger itineraries over one week.

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 passenger daily movements are limited, with ∼90% of the movements having less than 3 journeys. These observations support the findings about passenger 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 Comparing Temporal Perspective of Movement Rhythms Differences for Different Time Periods

In Study 2, 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 Figure 11 (a) - (d). To facilitate the comparisons, we scale the heights of root nodes based on their volumes.

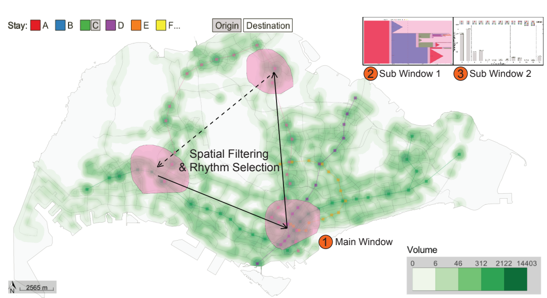


Figure 9: *Our visual interface arranges the three views in a main window and two sub windows.*

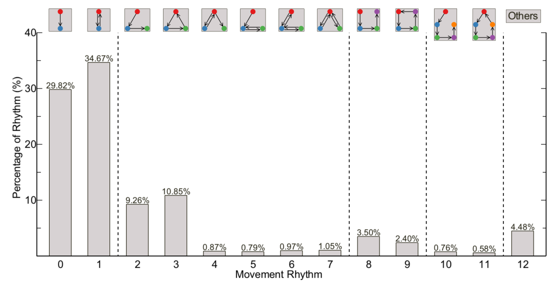


Figure 10: *Study 1: Analyzing statistics of frequent movement rhythms identified from all the movements over one week.*

By comparing the views between Monday and Sunday, i.e., comparing Figure 11(a) & (b) with Figure 11(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 Figure 11(a) are ∼2 hours longer than that those in Figure 11(c); whilst those in Figure 11(b) & (d) are almost the same.

By comparing the views between morning and afternoon, i.e., comparing Figure 11(a) & (c) with Figure 11(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 passengers 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 Figure 11(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.

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 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

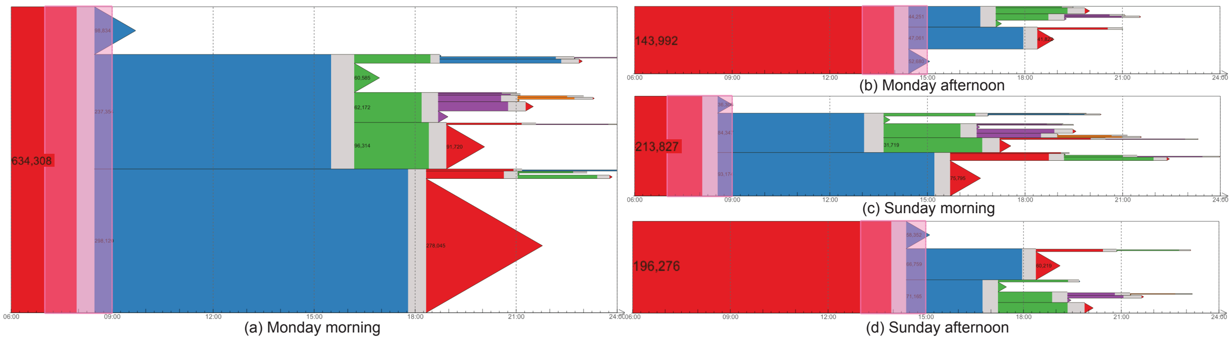


Figure 11: *Study 2: Comparing the temporal perspective of movement rhythm differences for different time periods: (a) Monday morning, (b) Monday afternoon, (c) Sunday morning, and (d) Sunday afternoon.*

case studies. Lastly, the experts explored the system by themselves. Each interview lasted for 40 minutes to 1 hour, 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. 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 2 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 adopted the alternative design as shown in Figure 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 passengers travel between locations".

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 visualization, where two and more sequential events have more meanings. In addition, from visualization perspective, we can achieve a multi-scale aggregation of event sequences by adjusting *N*. In this sense, Figure 8(a) can be considered as a detailed view of Figure 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 passengers 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 data sets. 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 over-aggregate 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 two case studies on studying massive one-week public transportation data in Singapore. 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. We thank anonymous reviewers for the various constructive comments.

References

AGRAWAL, R., AND SRIKANT, R. 1994. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases*, Morgan Kaufmann Publishers Inc., VLDB'94, 487–499.

ANDRIENKO, N., AND ANDRIENKO, G. 2012. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization 12*, 1, 3–24.

ANDRIENKO, G., ANDRIENKO, N., BREMM, S., SCHRECK, T., VON LANDESBERGER, T., BAK, P., AND KEIM, D. 2010. Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. *Computer Graphics Forum 29*, 3, 913–922.

BOWMAN, J. L., AND BEN-AKIVA, M. E. 2000. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice 35*, 1–28.

DEPARTMENT OF TRANSPORT AND MAIN ROADS, QUEENSLAND. 2016. *Public Transport Infrastructure Manual: Bus Stop Infrastructure*. 1 – 56.

FERREIRA, N., POCO, J., VO, H. T., FREIRE, J., AND SILVA, C. T. 2013. Visual exploration of big spatio-temporal urban data: A study of New York city taxi trips. *IEEE Transactions on Visualization and Computer Graphics 19*, 12, 2149–2158.

GUERRA-GÓMEZ, J., WONGSUPHASAWAT, K., WANG, T. D., PACK, M., AND PLAISANT, C. 2011. Analyzing incident management event sequences with interactive visualization. *Transportation Research Board 90th Annual Meeting Compendium of Papers*.

GUO, H., WANG, Z., YU, B., ZHAO, H., AND YUAN, X. 2011. TripVista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *Proceedings of IEEE Pacific Visualization Symposium*, 163–170.

KRÜGER, R., THOM, D., WÖRNER, M., BOSCH, H., AND ERTL, T. 2013. TrajectoryLenses – a set-based filtering and exploration technique for long-term trajectory data. *Computer Graphics Forum 32*, 3pt4, 451–460.

MONROE, M., RONGJIAN, L., HANSEUNG, L., PLAISANT, C., AND SHNEIDERMAN, B. 2013. Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics 19*, 12, 2227–2236.

PRIMERANO, F., TAYLOR, M. P., PITAKSRINGKARN, L., AND TISATO, P. 2008. Defining and understanding trip chaining behaviour. *Transportation 35*, 1, 55–72.

SCHEEPENS, R., HURTER, C., VAN DE WETERING, H., AND VAN WIJK, J. J. 2016. Visualization, selection, and analysis of traffic flows. *IEEE Transactions on Visualization and Computer Graphics 22*, 1, 379 – 388.

SCHNEIDER, C. M., BELIK, V., COURONN, T., SMOREDA, Z., AND GONZLEZ, M. C. 2013. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface 10*, 84.

SUN, G., LIU, Y., WU, W., LIANG, R., AND QU, H. 2014. Embedding temporal display into maps for occlusion-free visualization of spatio-temporal data. In *Proceedings of IEEE Pacific Visualization Symposium*, 185–192.

TOMINSKI, C., SCHUMANN, H., ANDRIENKO, G., AND ANDRIENKO, N. 2012. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics 18*, 12, 2565–2574.

TSANG, H. Y., TORY, M., AND SWINDELLS, C. 2010. eSeeTrack - visualizing sequential fixation patterns. *IEEE Transactions on Visualization and Computer Graphics 16*, 6, 953–962.

VROTSOU, K., JOHANSSON, J., AND COOPER, M. 2009. ActiviTree: Interactive visual exploration of sequences in event-based data using graph similarity. *IEEE Transactions on Visualization and Computer Graphics 15*, 6, 945–952.

WEI, J., SHEN, Z., SUNDARESAN, N., AND MA, K.-L. 2012. Visual cluster exploration of web clickstream data. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, 3–12.

WONGSUPHASAWAT, K., AND GOTZ, D. 2012. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12, 2659–2668.

WONGSUPHASAWAT, K., GOMEZ, J. A. G., PLAISANT, C., WANG, T. D., TAIEB-MAIMON, M., AND SHNEIDERMAN, B. 2011. LifeFlow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1747–1756.

ZENG, W., FU, C.-W., MÜLLER ARISONA, S., ERATH, A., AND QU, H. 2014. Visualizing mobility of public transportation system. *IEEE Transactions on Visualization and Computer Graphics 20*, 12, 1833–1842.

ZENG, W., FU, C.-W., MÜLLER ARISONA, S., ERATH, A., AND QU, H. 2015. Visualizing waypoints-constrained origin-destination patterns for massive transportation data. *Computer Graphics Forum*, n/a–n/a.