Journal of Visual Languages and Computing 000 (2017) 1–12

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

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

Classification and visualization for symbolic people flow data

Yuri Miyagi[a,*], Masaki Onishi[b], Chiemi Watanabe[c], Takayuki Itoh[a], Masahiro Takatsuka[d]

[a] Ochanomizu University, Japan
[b] National Institute of Advanced Industrial Science and Technology, Japan
[c] University of Tsukuba, Japan
[d] The University of Sydney, Australia

ARTICLE INFO

Article history:
Received 23 August 2017
Accepted 27 September 2017
Available online xxx

Keywords:
People flow
Symbolic trajectory
Movement pattern
Graph visualization

ABSTRACT

People flow information brings us useful knowledge in various industrial and social fields including traffic, disaster prevention, and marketing. However, it is still an open problem to develop effective people flow analysis techniques. We considered compression and data mining techniques are especially important for analysis and visualization of large-scale people flow datasets. This paper presents a visualization method for large-scale people flow dataset featuring compression and data mining techniques. This method firstly compresses the people flow datasets using UniversalSAX, an extended method of SAX (Symbolic Aggregate Approximation). Next, we apply algorithms inspired by natural language processing to extract movement patterns and classify walking routes. After this process, users can interactively observe trajectories and extracted features such as congestions and popular walking routes using a visualization tool. We had experiments of classifying and visualizing walking routes using two types of people flow dataset recorded at an exhibition and a corridor applying our method. The results allow us to discover characteristic movements such as stopping in front of particular exhibits, or persons who passed same places but walked at different speeds.

1. Introduction

Security cameras record many pictures of pedestrians every day. It is worth analyzing the pictures to discover movement patterns of the people, since we can get useful information to solve many social problems. For example, we can establish better evacuation routes, find causes of traffic jams, and come up with product displays that attract more customers, by interpreting the discovered movement patterns. However, it may be difficult to understand overall trends and find important knowledge from large amount of people flow datasets in a short time. For example, sizes of our datasets containing people flow during several weeks may be more than 10GB. We can preserve such datasets; however, it is not easy to observe or edit the datasets shortly (e.g. within a couple of minutes). It is still an open problem to develop compression and data mining techniques and visualize the movement patterns discovered from large-scale people flow datasets.

In this paper, we propose a visualization technique for large-scale people flow datasets. The technique firstly compresses the people flow data applying UniversalSAX [1], an extended implementation of SAX (Symbolic Aggregate Approximation). It converts the numeric position data to sets of smaller sizes of character datasets. It then applies natural language processing algorithms to the character datasets to quickly discover typical movement patterns and classify them. Finally, the technique visualizes the people flow datasets focusing on the movement patterns. Users can observe people flow interactively using three types of views, overview view, detailed view and clustering view.

This paper shows case studies with real-world people flow datasets in an exhibition and a corridor, then discusses effectiveness of the presented tool. In the first case, we used a 669MB people flow dataset and compressed it to only 115KB. Then, we visualized characteristic points such as congestion and walking routes. In the second case, we applied clustering before visualizing trajectories. We compared results of clustering following different conditions.

The remainder of this paper is organized as follows. In Section 2, we introduce related work on analysis of trajectories of pedestrians and other moving things. Section 3 presents the detail of the proposed technique. Section 4 introduces case studies using trajectory datasets recorded at an exhibition and a corridor. Section 5 contains a discussion of visualizations using the technique. Section 6 summarizes this paper.

* Corresponding author.
E-mail addresses: miyagi@itolab.is.ocha.ac.jp, miyagi.yuri@is.ocha.ac.jp (Y. Miyagi), onishi@ni.aist.go.jp (M. Onishi), chiemi@cs.tsukuba.ac.jp (C. Watanabe), itot@itolab.is.ocha.ac.jp (T. Itoh), masa.takatsuka@sydney.edu.au (M. Takatsuka).

https://doi.org/10.1016/j.jvlc.2017.09.005
1045-926X/

2. Related work

This section introduces existing studies to analyze people flow and trajectories of other moving things. Visualizing trajectories using graphs is major technique. Höcker [2] et al. proposed a graph structure to represent paths of walkers and an algorithm which searches for particular trajectories. Çetinkaya [3] et al. compared 4 types of graph visualization techniques using datasets of locations of states. These studies did not focus on summarization of trajectories, or development of interactive visualization tools.

Not only using graphs, there have been many existing methods which classify and visualize spatio-temporal people flow data recorded as real values. Following studies commonly visualized trajectories or their features on maps. Andrienko et al. [4] collected datasets from wide range using GPS, and analyzed properties of various moving objects, using both of drawing specific trajectories and bar charts on maps. On the contrary, our methods supposes that datasets are collected by cameras in small area like one floor in a shop. Andrienko et al. [5] also proposed interactive clustering method for trajectories. Furthermore, they visualized popular walking patterns on maps. Guo et al. [6] developed a composite visualization tool to analyze patterns of various objects such as pedestrians, bicycles, and cars. They adopted not only direct drawing of trajectories on maps, but also other visualization methods including piled polyline charts, scatterplots, and parallel coordinates plots. These techniques do not apply data compression techniques for trajectory datasets. Wang et al. [7] also extended the technique presented in [6], and visualized moving patterns of taxis in wider regions. Krueger et al. [8] presented an improved visualization system for chronologically GPS datasets. The main feature of this system is that users can move a circle looked like a lens, and focus on particular regions to observe detailed information such as speeds and directions. Wang et al. [9] extracted and visualized features of automobiles passing at particular positions, applying datasets collected using many sensors on roads. Lu et al. [10] proposed TrajRank, a visualization system for trajectories for vehicles like taxies. They separated trajectories to segments and calculated their ranks to suggest characteristic travel behaviors.

Other techniques introduced below commonly visualized trajectories on pictures of camera views. Mehran et al. [11] proposed a technique using streamline visualization and abnormality detection methods. They claimed that streamlines are superior to path lines. Yabushita et al. [12] proposed a technique which summarizes pedestrians' trajectories recorded at open spaces where definite routes are not constructed. This technique effectively represents major routes of pedestrians; however, it misses some types of important information including temporal tendency and walking speeds. Ko et al. [13] focused on angles of trajectories of walking people, and extracted irregular movings such as a zigzag. They also tried to transform the trajectories onto pictures of camera views, and visualized them in order for users to understand the angles of the trajectories. Fukute et al. [14] applied a spectral clustering algorithm to pedestrians' trajectories to classify them to meaningful sets of walking patterns, and visualized temporal transition of populations for each cluster by applying a piled polyline chart. Andrienko et al. [15] proposed "trajectory wall" as an extension of space-time cubes. Users can grasp trajectories in a same cluster which have similarities about routes. Guo et al. [16] classified walkers' trajectories according to their speed and direction. They also developed a system to visualize important trajectories using meaningful colors based on HSV model. However, the tool does not support interactive trajectory selection for detail-on-demand visualization.

Following researches focused on observation of moving patterns, though analysis of trajectories is not exactly a main task. Gupta et al. [17] worked on to visualize relationships among small number of pedestrians. They did not visualize particular shapes of trajectories, but applied a gantt chart and visualized places where people stayed. This representation is especially useful for users because they can find places where multiple persons stayed at the same time. Krueger et al. [18] presented an improvement of another visualization system named TravelDiff. The system consumes messages in Twitter instead of applying datasets of trajectories, and visualizes movement patterns and crowded places. They tried to visualize three types of datasets, for pedestrians, taxies, and airplanes, as graphs and heatmaps.

AI-Dohuki et al. [19] developed a visualization system for trajectories of taxies. The system can visualize not only statistical traffic information but also messages which answer to questions input by users. For example, users can select particular streets or time periods, and visualize related datasets.

Also, there have been many techniques on compression and pattern discovery techniques for people flow datasets. Teknomo et al. [20] analyzed moving patterns of customers at a supermarket. They allocated letters to each intersection in a hypermarket and expressed customers walking routes as strings. Also, they analyzed populations and length of walking times. Thack et al. [21] converted the spatiotemporal trajectories preserving the distances in the original space, and then divided the trajectories. They succeeded to distinguish four types of trajectories collected under different conditions, by taking into account both positions and movement patterns. Oates et al. [22] successfully extracted motifs of trajectories from noisy people flow datasets by applying a context-free grammar technique. These studies adopted SAX or TraSAX (an extension of SAX) during compressing and visualizing datasets, as we also adopt. However, a common issue in these studies is that they displayed all recorded trajectories as is. Their visualization results are therefore so complex that it may be difficult to find important routes or places. Contrary to these methods, our technique firstly extracts typical movement patterns and then visualizes them noticeably. Yada [23] analyzed movements of customers in a supermarket, and searched for popular sections. He converted original datasets to sequences of characters that indicate sections where customers moved to. However, important information including times of staying at each section is not preserved after converting.

3. Presented visualization tool

This section presents the processing flow and detailed description of technical components of the presented technique. Fig. 1 shows the whole processing flow of our proposed technique. After the preprocessing including trajectory recording, conversion to strings and feature extraction, users can interact with the visualization.

Section 3.1 defines the people flow datasets. Section 3.2 describes a technique to convert the trajectories into sets of characters, following by the discovery of typical movement features described in Section 3.3. Lastly, we describe a visualization technique which emphatically displays the features in Section 3.4.

3.1. Recording people flow data

We define that a record of people flow data includes the following information:

- Time at which the position of a walker is measured.
- ID of the walker.
- Position of the walker in a 2D space (x, y).

We can construct a trajectory of this walker by collecting the records which have the particular ID corresponding to the walker, and then chronologically ordering the collected records.