Visual Interface for Exploring Caution Spots from Vehicle Recorder Big Data

Masahiko Itoh*, Daisaku Yokoyama†, Masashi Toyoda‡, and Masaru Kitsuregawa‡

*The University of Tokyo, National Institute of Information and Communications Technology, Email: imash@tkl.iis.u-tokyo.ac.jp
†The University of Tokyo, Email: {yokoyama,toyoda}@tkl.iis.u-tokyo.ac.jp
‡National Institute of Informatics, The University of Tokyo, Email: kitsure@tkl.iis.u-tokyo.ac.jp

Abstract—It is vital for the transportation industry, which performs most of their work by automobiles, to reduce its number of traffic accidents. Many local governments in Japan have made potential risk maps of traffic accident spots. However, making such maps in wide areas and with the time information had been difficult because most of them are made based on an investigation. Utilizing long-term driving records can extract wide area spatio-temporal caution spots. This paper proposes a visual interaction method for exploring caution spots from large-scale vehicle recorder data. Our method provides (i) a flexible filtering interface for driving operations using various combinations of attribute values such as velocity and acceleration, and (ii) a 3D visual environment for spatio-temporal exploration of caution spots. We demonstrate the usefulness of our novel visual exploration environment using real data given by one of the biggest transportation companies in Japan. Exploration results show our environments can extract caution spots where some accidents have actually occurred or that are on very narrow roads with bad visibility.

I. INTRODUCTION

Traffic accidents are still troubling our society. In 2013, a total of 629,021 traffic accidents occurred in Japan according to recent transportation statistics1. Many local governments in Japan have made potential risk maps of traffic accident spots to reduce the number of such tragedies. Risks are estimated using the questionnaires from drivers and pedestrians, and/or the occurrence of actual accidents. However, most such maps lack accuracy and comprehensiveness. Questionnaires can reflect vague human impressions. Collected risky locations are often unevenly distributed in the neighborhood of elementary or junior high schools. They also do not reflect time and weather information that may affect the risk of traffic accidents.

The transportation industry, which performs most of their work by automobiles, must reduce its number of accidents. The industry has started to introduce dashcams or vehicle recorder systems to retrieve information on accidents and to increase drivers’ safety awareness. Collected information helps drivers to look back on their daily driving at the end of the workday. If we collect many drivers’ records over a long time, the data will allow us to find caution spots for driving. Extracting such spots helps to create a new risk map on the basis of many facts about risky situations. The maps will be able to cover wide areas and to reflect both spatial and temporal information. Extracted caution spots are useful for drivers’ safety education. They also allow us to discover structural problems of roads and towns, leading to designing a safe city.

Some research has been conducted on spatio-temporal analysis and visualization of mobility data collected by tracking technologies such as GPS [1], [2], [3], [4], [5], [6], [7]. However, most studies have focused on analyzing traffic jams or movement patterns. As far as we know, no research has explored caution spots for driving on the basis of real vehicle recorder data using flexible filtering interfaces of driving operations and 3D spatio-temporal visual environments.

This paper proposes a novel visual exploration interface (as shown in Figure 1) to explore caution spots in wide area spatio-temporal space from vehicle recorder data including braking and handling operation logs.

Our major contributions are as follows:

• We describe our design for a visual interface to filter driving operations flexibly in accordance with various combinations of their attribute values. An exploration interface is provided to calculate the degree of caution for each driving operation using the relationships between the attribute values of the driving operation logs and accident information in a specific area (Figure 1 (a)). Also, a dynamic query interface is provided, one that uses a parallel coordinates view for filtering driving operations by the calculated degrees of caution (Figure 1 (b)).
• We present an exploration interface using 3D spatio-temporal space (Figure 1 (c-f)) to discover caution spots using a huge amount of driving operation records. The spots are pointed out using a visual filtering interface (Figure 1 (b-d)). Our system can investigate the details in caution spots by visualizing movement paths of each driving (Figure 1 (f)).

We demonstrate the usefulness of our novel visual exploration environment using a series of case studies extracted from real data given by one of the biggest transportation companies in Japan. Exploration results show how our environments can flexibly extract various kinds of caution spots such as those where some accidents actually occurred or where very narrow roads with bad visibility exist.

In what follows, we give an outline of related work in Section II. We next introduce an overview of our framework for a visual exploration of caution spots from vehicle recorder data in Section III. We offer information about our data set in Section IV. Then, in Section V and Section VI, we describe visual interfaces to filter driving operations flexibly and a 3D spatio-temporal visualization environment for exploring caution spots, respectively. We present some case studies in Section VII. This article is concluded in Section VIII.

II. RELATED WORK

Many studies have been conducted on recording driving speed and/or location information and on utilizing them in various applications [8], [9], [10], [11]. Although many studies have been done on utilizing operations during driving [12], [13], [14], most of them focused on analyzing the characteristics of drivers and on clustering them.

Some research has been conducted on spatio-temporal analysis and visualization of mobility data collected by tracking technologies such as GPS using 2D visualization space [1], [2], [3], [4] and 3D visualization space [5], [6], [7]. However, most of them focused on analyzing traffic jams or movement patterns. Ferreira et al. provided a query model and a visual environment for exploring human activity patterns in New York City using taxi trip data [2]. Wang et al. extracted and visualized traffic jams and their propagation from data of taxi trips in Beijing [1]. Andrienko et al. extracted and characterized important places from mobility data such as GPS tracks of cars and flight trajectories and visualized them in 2D/3D spatio-temporal space [4], [5], [15].

SAFETY MAP\(^2\) independently plots locations in which traffic accidents occurred and drivers suddenly braked. However, it does not provide a mechanism for exploring caution spots using relationships between accident data and event data extracted from a probe-car system. G Map\(^3\) plots points in which rapid acceleration occurred. Although it provides a filter interface by the magnitude of acceleration, the size of the provided data and areas is very small.

As far as we know, no research has been done on the visual exploration of caution spots for driving based on a huge amount of real vehicle recorder data. Our novel method provides a visual interface to explore various kinds of caution

\(^2\)http://www.honda.co.jp/safetymap/, (in Japanese)
spots flexibly using the various combinations of attribute values of driving logs such as velocity, longitudinal acceleration, and jerk.

III. OVERVIEW

Figure 2 gives an overview of our system. Vehicle recorder data include various noises: e.g., time and latitude/longitude are apparently out of range. Therefore, we eliminate such outliers from collected data in the preprocessing step. Operation records such as braking, handling, stopping, and turning operation records originally do not include concrete time information but include driving duration time. Therefore, we convert driving duration time to concrete time in this step. The details of the generated datasets are described in Section IV.

Next, we calculate the degree of caution for each operation to extract the caution spots. To do this, we first utilize our provided visual interface to determine standards for flexibly calculating degrees of caution. The interface provides scatter plots with rotatable standard lines for various combinations of attribute values such as velocity, acceleration, and jerk to explore a high correlation coefficient between degrees of caution and the number of real accidents in the same regions. Degrees of caution for operations are calculated based on the slope of specified standard lines for degree of caution. We can calculate various kinds of degrees of caution for driving operations from various combinations of attributes. It enables us to explore various kinds of caution spots flexibly, such as narrow areas or huge junctions. To explore caution spots, we next utilize a provided dynamic query [16] interface using a parallel coordinates view [17] for filtering driving operations with the calculated degrees of caution. We present the details of the method used to explore degree-of-caution standards and our visual interface in Section V.

Degrees of caution for operations are visualized in 3D spatio-temporal visualization space. We can interactively explore caution spots in 3D visualization space using dynamic queries based on calculated degrees of caution. Moreover, we can explore what happened in the extracted spots by visualizing movement trails in detail. The details on the 3D visualization environment are shown in Section VI.

IV. DATASETS

For the experiments, we use large-scale real driving records collected by Sagawa Express Co., Ltd., one of the biggest transportation companies in Japan providing a door-to-door delivery service, in cooperation with Datatec Co., Ltd. The records consist of over 100 drivers’ worth of data for about one year (from July 21, 2014 to June 4, 2015). Data were obtained by a multifunctional vehicle recorder, developed by Datatec Co., Ltd., that has a longitudinal accelerometer, lateral accelerometer, gyro compass, and GPS.

The vehicle recorder captured driving trajectory and status every 0.5 seconds. The vehicle recorders also automatically detected four basic driving operations: braking, handling, turning, and stopping. Several values such as speed or acceleration during the operation were recorded. The details of the recorded data for each operation are shown in Table I.

V. FLEXIBLE FILTER INTERFACE

A. Exploration of Standards for Degrees of Caution

We provide a mechanism for exploring appropriate standards to determine degrees of caution for operation records on the basis of combinations of their attribute values. Many types
TABLE I
EXPERIMENTAL DATASET OF DRIVING RECORDS COLLECTED BY A DOOR-TO-DOOR DELIVERY SERVICE COMPANY

<table>
<thead>
<tr>
<th>Record type</th>
<th>Number of records (plots)</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periodical driving status</td>
<td>445 million</td>
<td>date, time, driving duration, latitude/longitude, speed (by GPS), speed (by tachometer), etc.</td>
</tr>
<tr>
<td>Braking operation</td>
<td>2.1 million</td>
<td>date, time, driving duration, latitude/longitude, speed (V), longitudinal acceleration (Gx), and jerk (the derivative of acceleration with respect to time)</td>
</tr>
<tr>
<td>Handling operation</td>
<td>3 million</td>
<td>date, time, driving duration, latitude/longitude, speed, yaw velocity (Yr), yaw angular acceleration, and lateral acceleration (Gy)</td>
</tr>
<tr>
<td>Stopping operation</td>
<td>2.4 million</td>
<td>date, time, driving duration, latitude/longitude, speed, longitudinal acceleration, and stopping duration</td>
</tr>
<tr>
<td>Turning operation</td>
<td>1.3 million</td>
<td>{V, Gx} before corner, {V, Composed G, Yr fluctuation} during the turn, and {V, Composed G} after corner</td>
</tr>
</tbody>
</table>

Fig. 4. Example of Standards Exploration View for handling operations including four attributes, speed, yaw velocity, yaw angular acceleration, and lateral acceleration.

1) Count the number of real accidents per area (100-meter grid in this case)
2) Extract operation records for every same area and plot them into a 2D scatter plot using two arbitrary attributes of operation for axes.
3) Draw a standard line having an arbitrary angle in the scatter plot, calculate a degree of caution for each operation record on the basis of the standard line, then total them up for every area.
4) Calculate a correlation coefficient between the number of accidents and sum of degrees of caution for every area.
5) Repeat steps 3&4 to obtain the appropriate standard line having the higher correlation coefficient.
6) Try other combinations of attributes in step 2 to get other kinds of standards for degrees of caution.

For this purpose, we utilize five years’ worth of accident place records totaling around 500; all of the locations are plotted on the map provided by Bunkyo ward, which is one of the local governments in Tokyo\(^5\). We counted the number of accidents for every 100-meter grid to calculate the correlation coefficient with the total of the degrees of caution for operations in the same grid.

Figure 3 and 4 show the scatter plot matrix based visual exploration interface for degree-of-caution standards, called the standards exploration view (SEV). Each scatter plot in the SEV has two axes selected from attributes of operations. In the case of braking operations, speed, longitudinal acceleration, and jerk can be selected (Figure 3). In the case of handling operations, speed, yaw velocity, yaw angular acceleration, and lateral acceleration can be selected (Figure 4). SEV plots operation records included in the same regions as those on the map provided by Bunkyo ward. Each plot has a grid id. In this case, we use the operation records for about one month (from July 21, 2014 to August 20, 2014).

SEV measures the distance of each operation plot from a standard line as a degree of caution (Figure 3 (a)). The degree of caution for each plot changes in accordance with the slope of the standard line. Users can interactively rotate the standard line to determine the standard to enable a high correlation coefficient between the sum of the degrees of caution and

the number of accidents in the same 100-meter grids. The calculated correlation coefficient is shown in the center of each scatter plot (Figure 3 (b)).

By using an SEV interface, we can explore various kinds of standards from various attributes. We can select two or more standards lines to calculate degrees of caution for filtering operation records (Section V-B) and for visualizing them in 3D spatio-temporal space (Section VI). Selected standards are highlighted in Figure 3 and 4.

B. Filtering Operations based on Degrees of Caution

We can explore caution spots by filtering driving operations, such as braking, handling, stopping, and/or turning operations based on calculated degrees of caution using a PCV [17] for dynamic query [16] as shown in Figure 5. We can also filter operations using attribute values. Axes in the PCV represent selected degree-of-caution standards specified in the previous section and attributes such as speed, longitudinal acceleration, and jerk as shown in Figure 5 (b). Line graphs in the PCV represent operations. Users can interactively move the selection handles to define the selection ranges for axes.

C. Discussion

Other approaches can be used to calculate the degrees of caution or filter driving operations.

Utilizing multiple linear regression to estimate optimal degrees of caution is one of the simplest approaches. We can estimate the number of accidents as a degree of caution from the values of driving operation attributes. However, our preliminary trial showed very low accuracy. For example, the coefficient of determination ($R^2$) was 0.01048 when we used the speed, $G_x$, and jerk of braking operations as predictor variables. We also needed to consider various combinations of attribute values as predictor variables to deal with various kinds of dangerous situations such as huge junctions with high traffic volume or narrow alleys with poor visibility.

VI. 3D Spatio-temporal Visual Environment

We provide a visual exploration environment for finding caution spots and time based on degrees of caution. Figure 1 (c) - (f) show our proposed exploration environment.

We utilize a 3D spatio-temporal visualization space (sometimes called a space time cube [18]) to project degrees of caution for operations. In our space time cube implementation, the base represents geography, and the height represents time, with later times at the bottom. Each operation record is represented as a sphere and mapped on the 3D space in accordance with its longitude, latitude, and time. The size and color of each sphere are defined by its degrees of caution or attribute values. Users can arbitrarily define their mappings. The examples shown in this paper utilize the following mappings:

- size of plots for braking operations: degrees of caution defined by Figure 3 (i)
- color of plots for braking operations: degrees of caution defined by Figure 3 (ii)
- size of plots for handling operations: degrees of caution defined by Figure 4 (i)
- color of plots for handling operations: degrees of caution defined by Figure 4 (ii)

Users can also define the standard-line as they like.
Users can interactively zoom in/out of, rotate, and pan the 3D space to observe plots from various viewpoints. Users can also pan and zoom in/out of a map on the base. In this implementation, we utilize BingMap\textsuperscript{7} to get appropriate road maps (or satellite photo maps) in accordance with manipulations.

A. 3D Design Considerations

We have mainly two types of visualization methods for spatio-temporal movement data for people or objects such as cars involving three variables: two spatial variables and a time variable. The first one utilizes 2D visualization space that simply plots the data on a 2D geographic map and sometimes changes the time by using a time slider or utilizes multiple maps with different time stamps to represent temporal changes in the target objects [1], [2], [3], [4]. Another one utilizes 3D visualization space adding the third axis to represent time that combines space and time in a single display [5], [6], [7], [6].

Amini et al. presented experimental comparisons of 2D and 3D visualizations of movement data involving commonly performed tasks [19]. Their survey showed a high possibility of 3D visualization for deeper probing of complex analytical tasks and for finding cluster objects and times because it simultaneously presents all the required information. It also shows 2D visualization requires relying significantly on scrubbing the timeline.

Our exploration purpose is to find caution spots reflecting the position of driving operations and time information. Target objects, time, and spaces to be searched are not fixed. An overview of plots in a wide area spatio-temporal space is required. Therefore, we adopt 3D visualization space for this work.

B. Additional Lines and Shadows

3D visualization sometimes causes occlusion problems. Illustrative Shadows [20] integrates 3D graphic and schematic depictions using the shadow metaphor. Such 3D shadows on a geographical plane and/or time planes and additional lines to these planes help us to understand the position of floating plots in the 3D space and to provide additional information.

Examples in Figure 1 (c), (d), (e), and (f) show shadows of plots and movement paths and also show additional lines. We can interactively change the mode to show shadows on or additional lines to the planes.

C. Aggregation View Mode

Visualized plots still look messy or crowded even after filtering out many plots. Aggregated plot representations help us to find caution spots.

We adopt hierarchical clustering to generate an aggregation of visible driving operations\textsuperscript{8}. We utilize the group average method for clustering using Euclidean distance in 3D space. Longitude, latitude, and time (converted to unix time) values for each plot are normalized between 0 to 1. We compress values for time in half to emphasize spatial differences.

Users can interactively cut a dendrogram by specifying the distance thresholds or the number of clusters, as shown in Figure 6.

Figure 7 shows three types of view modes, (a) plot view mode, (b) hybrid view mode, and (c) aggregation view mode.

\textsuperscript{8}We utilize hierarchical clustering in this implementation, but hierarchical clustering does not scale enough. We plan to apply DBSCAN to generate an aggregation view.

\textsuperscript{7}http://www.microsoft.com/maps/
Hybrid view mode shows both aggregated nodes as clusters and operation plots. In the aggregation view mode and hybrid view mode, the size of the aggregated node is defined by the sum of the degrees of caution in each cluster. The position of each aggregated node is specified by the center of operations in each cluster. The transparency of each aggregated node represents the number of operations included in each cluster. The colors of aggregated nodes and plots for operations are used for distinguishing clusters.

D. Detail Exploration of Movements

We can interactively select plots by specifying the range of longitude, latitude, and time as shown in Figure 1 (f). Moreover, we can visualize movement trajectories of driving as lines. Red lines represent 10-second trajectories before braking or handling operations occurred, and black lines represent 10-second trajectories after operation occurred. We can understand the moving directions of drivers from the colors of lines, as shown in Figure 9.

VII. Case Studies

This section demonstrates the usefulness of our system using extractions of various kinds of caution spots. In the experiments, we visualize driving operations consisting of about 80 drivers assigned to Bunkyo ward over a period of about one month (from July 21, 2014 to August 20, 2014).

Figure 8 shows the comparison between actual accident places, which are plotted on the map provided by Bunkyo ward (previously mentioned in Section V), and extracted caution spots are represented as aggregated nodes. These results are generated from handling operations through processes shown in Figure 1 (a), (b), (c), (d), and (e). We find that we can extract almost the same spots using our system. It is difficult to extract such spots through filtering only using attribute values on the PCV without using degrees of caution because the distribution of operations are tilted, as shown in Figure 4.

Figure 9 shows exploration examples with two interesting caution spots extracted. These results are generated from braking operations. In the bottom of Figure 9 (c), we can find many caution spots on the big road. We can also find caution spots on the very narrow roads. Figure 9 (d) and (e) visualize the explored caution spots in detail. Although drivers normally step on the brake before the corner, example (d) is in a place that is just after the corner of the narrow road. Using Google street view, we noticed the presence of entrances of a parking lot and an apartment building in places with low visibility. In the case of example (e), we can find the caution spot in the middle of a narrow road with no crossings. By zooming
the map in, we can find a gate of the apartment building in the front of extracted spot. We found that drivers had trouble recognizing the gate when they moved from north to south.

Many caution spots were extracted, especially in university campuses. One reason for that is the campuses have many closing gates and bumps on the road. Figure 10 (a) includes two caution spots around one of the school’s gates in the University of Tokyo. One spot is on the route from inside the campus to the gate. The other spot is on the route from the gate to inside the campus.

The example in Figure 10 (b) shows one explored caution spot including too many braking operations. The spot is in a place just before an intersection connecting to three one-way roads. In this spot, drivers coming from the south have no choice but to turn left.

Although these examples only used operation records from Bunkyo ward, we can reuse the same parameters such as standard lines or filtering handles to explore caution spots in wider regions.

**VIII. CONCLUSION**

In this paper, we proposed a novel visual exploration system that enables us to explore various types of caution spots flexibly for driving from a wide area of spatio-temporal space using a massive amount of vehicle recorder data. We also demonstrated the usefulness of our novel visual exploration environment with some case studies. Our system enables investigating the differences between a normal date and special date when huge events affecting wide areas such as marathons or disasters occur. It also enables us to compare situations before, during, and after construction of roads. Our system’s ability to explore caution spots is useful for drivers to improve their safety awareness and their understanding of urban development to reduce the possibility of their experiencing traffic accidents.

Some extracted caution spots in our study depended on the delivery time and routes of individual drivers. Thus, we need more long-term data to get more general results. We plan to utilize more long-term data, weather data, elevation data, and photos taken by dashcams to gain deeper insight into caution spots.

**ACKNOWLEDGMENT**

We thank Sagawa Express Co., Ltd. and Datatec Co., Ltd. for providing and collecting the valuable vehicle recorder data. This work was supported by the Research and Development on Real World Big Data Integration and Analysis program of the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

**REFERENCES**


