

# Interactive Visual Analysis for Vehicle Detector Data

Yi-Cheng Chen, Yu-Shuen Wang, Wen-Chieh Lin, Wei-Xiang Huang, I-Chen Lin

National Chiao Tung University, Taiwan

---

## Abstract

*Visualization of vehicle detection (VD) data is essential because the data play an important role in traffic control and policy development. Most previous works focus on visualizing trajectories obtained from global positioning system (GPS), which are detailed but less representative. In contrast, VD data report the traffic statistic at each sensing site during a time span, including speed, flow, and occupancy of each lane, which contain comprehensive traffic information for analysis. In this work, we visualize three-year VD data of freeways in Taiwan. The visualization depicts the traffic situation at a site over time using a color-coded chart that extends from left to right over time. The charts are vertically stacked and horizontally aligned according to VD's located mileage and data time, respectively, to provide global insight. Our system allows semantic zoom, which changes the chart appearance in a continuous manner, to enable macro- and micro- scopic visualizations. Analysts can explore events that span an area with different sizes and that persist a time span with various lengths. To ensure the feasibility of our visualization, before the system design, we conducted a study with experts who work in the national freeway bureau and the institute of transportation of Taiwan. We also showed our results to the experts after the prototype system was built. The feedback shows that our VD data visualization is helpful to traffic control and policy development.*

---

## 1. Introduction

Automotive data analysis is getting important in recent years because movement between locations involves our daily life. The analysis helps traffic controllers and city planners prevent traffic jams and develop traffic networks. However, interpreting a large amount of historical traffic data is challenging because traffic statistic is location dependent, multivariate, and time-varying. Consequently, developing an intuitive visualization system that helps analysts make sense of traffic data easily is essential.

Most traffic visualization methods are presented to depict vehicle GPS trajectories. Although these visualizations can clearly show high frequent routes and detailed vehicle movements, they are difficult to convey overall traffic situations because the trajectories are contributed from users and do not exist all the time. Therefore, experts in the fields of traffic control and development usually rely on statistic data recorded from vehicle detectors (VDs) when making decisions. Specifically, VDs are distributed in roads to detect the presence of large metal objects. Each recorded data item contains an average speed, flow, and occupancy of a region within a time span. Given that the detector senses traffic situations of a local region 24 hours per day, the recorded data are very helpful for analysts to predict traffic flows, examine

the effectiveness of a traffic control and evaluate the utility rate of a road segment. Although VD data depict comprehensive traffic situations, these statistic data are difficult to interpret. Analysts have to read a small amount of data at a time and gradually build the overall concept by themselves to fully understand a traffic situation. This interpretation process is time consuming, demands skills, and prevents the experts from making urgent decisions in time.

We introduce a visualization system that represents VD data using color-coded charts. Each chart reveals the traffic situation of a local region and the changes over time. A color-coded chart can show the relation of either speed-occupancy or flow-occupancy at various time spans to satisfy the demands of different purposes. Considering traffic situations are spatially and temporally related, we vertically stack and horizontally align the charts according to their mileages and time, respectively, on the screen. Our designed color-coded charts are space efficient so that the system can display a large number of charts and enable the investigation of traffic situations that span a wide area and persist long time. To allow analysts discovering traffic situations at different scales, our system deforms the color-coded charts to enhance data clarity when macro- and micro- scopic views are transited. This semantic zoom achieves high quality vi-

sual experience. Given that most analysts attempt to further study the traffic situation when an unexpected event occurs, our system also provides a retrieval interface for users to find similar traffic patterns by example. They can examine the places, dates, and frequencies to understand the cause of formation. Overall, our system helps analysts discover historical traffic data easily, and then enable them to predict future traffic situations, confirm the effectiveness of a traffic control, and study the unexpected traffic situations.

To the best of our knowledge, the presented technique is the first visualization system that aims to provide experts with both macro- and micro- scopic views of VD data. Users can easily understand traffic situations by discovering color-coded charts. In Section 4, we show how our system depicts local and global traffic events, including traffic bottlenecks, weekly traffic peak, human mobility at a traditional holiday, and hurricane strikes. The experts who worked in traffic management and research centers verified that efficiently interpreting these data is important. Hence, they confirmed the practical usability of our prototype system and looked forward to using it in future.

## 2. Related work

Spatial temporal data often contain geography, time, and other properties [vLBA\*12]. Many data such as cell phone positions, animal migrations, and Facebook check-ins belong to this field. In this paper, we focus on traffic data that involve our daily life.

**Traffic data visualization.** There have been many algorithms presented to visualize traffic data. Many of them focus on visualizing vehicle GPS trajectories. Yuan et al. [GWY\*11] presented a TripVista system that helps analysts interactively examine traffic events and unexpected occurrences. The visualization depicts driving of different behaviors or against traffic regulations from spatial, temporal, and multi-dimensional views. Liu et al. [LGL\*11] introduced the route diversity visualization because the paths that connect two locations are not identical. Users have different options and the system helps them make the best decision. Bak et al. [BMH\*12] focused on the encounter patterns of trajectories. Users can retrieve the defined encounter patterns of bus and tram trajectories to study the reactions of vehicles being close to one another in terms of space and time.

Given that drawing a large number of trajectories would introduce visual clutter, aggregation usually is used to prevent this problem. Specifically, the method presented by Andrienko et al. [AA08] aggregates trajectories according to similarity, time, and traffic situation. As a result, analysts can clearly observe traffic flows, high frequent routes, start and end positions, and the ratio of vehicles slower than a certain threshold. After that, they extended their method [AA11] by extracting high curvature points from trajectories, followed by aggregating them as features and connecting the

features using straight lines to reduce visual clutter. Tominski et al. [TSAA12] stacked trajectories on the map and colorized these trajectories to represent different speeds. The system also allows users to specify regions of interest and investigate changes of traffic flows within the region over time. Doraiswamy et al. [DFD\*14] introduced event-guided exploration of large spatiotemporal urban data. Events are extracted, grouped, and indexed to allow for users to interactively discover patterns of interest.

In addition to vehicle trajectories, some researches pay attentions to origin destination (OD) data visualization [BBBL11, GZJ\*12, WDS10]. OD data is a collection of flows, which is usually represented as a matrix, used to indicate the amount of humans/vehicles moving between pairs of geographic locations. These data could be obtained from personalized user-tagged cards for public transportation or extracted from massive GPS trajectories. The analysis of OD matrix is helpful in transportation planning [ZFAQ13], urban mobility discovering [JFG12, YZX12, FPV\*13], and city planning [WSD11]. Although OD data visualization reveals spatial structure of an urban city, which is useful to city planners, the data are too abstract to describe traffic situations due to the lack of moving routes. While only the amount of entities traveling between locations is available, interpreting such data is not sufficient to make sense of traffic events such as vehicle accidents and moving bottlenecks.

Experts who work in traffic management and research centers rely on VD statistic because the data provide comprehensive and systematic traffic statistics. A tool presented by VanDaniker [Van09] shows vehicle speed and flow of a small area affected by an incident. However, this simple approach is oblivious to occupancy and potentially misleads users into wrong information. Konyha et al. [KMG\*06] introduced an interactive system that links multiple views, such as maps, histograms, scatterplots, parallel coordinates, and function graphs, for experts to analyze vehicle statistics at various regions. The system also provides several operations for filtering uninteresting data. However, each linked view is uninformative and users have to switch attention frequently when investigating traffic data. In contrast, our system provides clear visual means in the main window and allows for both macro- and micro- scopic views. It also provides pattern retrieval for finding similar traffic patterns.

**Time series data visualization.** VD data can be considered a kind of time series data because speed, flow, and occupancy change over time. To study such data, ThemeRiver [HHN00, HHWN02] are widely used. The river extends in one dimension over time and its width in another dimension depicts temporal changes of variables. There were many extensions based on this structure presented to enable social data analysis [Wat05, WK06, CLT\*11]. To prevent visual clutter, Thakur and Rhyne [TR09] applied kite diagrams to form a data vase system, which draws time-varying quantities in parallel. Besides the parallel mode, Weber et al.

[WAM01] represented time series data by using spirals that support better identification of periodic data structures. Although the aforementioned techniques help users make sense of changes in time series data, they are not space efficient so that only a few series can be displayed on a screen. In addition, these methods are not designed according to traffic properties, which are not sufficient to provide insightful visual patterns for traffic events.

### 3. Design methodology

We present a system to visualize traffic situations of free-ways in Taiwan. Before designing the system, we conducted a study with experts who work in traffic management and research centers. Below, we show the data we used, the insights interested the experts, the tasks we attempted to fulfill, and the design we invented for traffic situation visualization.

#### 3.1. Data backgrounds

Speed, occupancy, and flow play important roles in traffic theory because they show vehicle situation of a place within a time span. Specifically, speed is used to show how fast a vehicle can travel on the road; occupancy indicates the percent of time that the detection zone of a detector is occupied by some vehicles; and flow means the total number of vehicles that pass a detector. These data are detected by vehicle detectors that are distributed in road, contain traffic statistic of a lane within a minute, and are returned over time. We average the statistic of multiple lanes in this work because the overall throughputs are considered in most cases. By distributing sufficient amount of vehicle detectors in a road network, traffic management center is able to monitor the traffic in a city. Figure 1 shows the fundamental diagrams, where data are plotted on a 2D space with different variables to show the relations of the traffic data. Clearly, speed is inversely proportional to occupancy due to many vehicles on the road. However, the relation of speed and flow is complex; both high and low speeds indicate light flow because of different occupancies. The complex relation also appears in flow and occupancy; these two variables are roughly proportional when occupancy is lower than a threshold but they become roughly inversely proportional when occupancy gets higher. Although the fundamental diagrams clearly depict variable relations, they are not sufficient to show the changes of traffic situations at different locations over time.

#### 3.2. Problem definition

Most traffic experts rely on VD data to discover traffic situations because VD data provides both macro- and microscopic views of spatiotemporal data. However, due to a large amount of numerical values, experts are difficult to get insight to the traffic statistic of a wide area and a long time span. This motivates us to develop a visualization system, which can help traffic experts make sense of historical data,

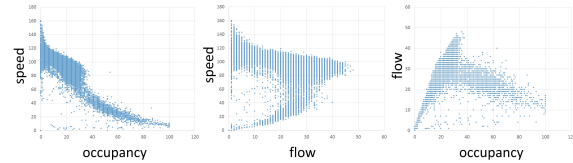


Figure 1: Fundamental diagrams are used to depict relations of speed-occupancy, speed-flow, and flow-occupancy.

and then predict or control traffic situations after data exploration. For example, a national day on Thursday and on Friday would result in very different traffic flows because the latter results in a three-day holiday. To achieve a precise traffic prediction, an expert has to observe traffic patterns in historical data, where the factors, such as the length of holidays in a row and the date of the festival in holidays, are similar. Besides prediction, data exploration helps experts achieve better traffic control. After applying strategies such as ramp meter, toll-free travel in off-peak periods, or high occupancy restrictions during peak periods, experts can observe the data w/wo the strategy to examine the effectiveness of a traffic control. They also can investigate whether side effects occur at nearby regions rather than only the region with control. Moreover, whenever experts discover a traffic pattern that has never been thought, they can retrieve similar patterns in the historical data, examine the places, dates, and frequencies of these patterns to find the cause of formation. In summary, traffic experts are interested in:

- Historical traffic situations that span a wide area and long time; particularly, the situations caused by holidays, weather, and quality of roads.
- Difference of traffic situations w/wo controls.
- Influenced area and time span of a vehicle accident under various circumstances.
- Cause of formation of various traffic situations.

According to the tasks, we carefully design the visualization interface for traffic experts to discover VD data. Specifically, the system has to satisfy the following requirements:

- Support both macro- and micro- scopic views; users are allowed to scale spatial and temporal ranges individually during data exploration.
- Prevent ambiguity. Low vehicle flow could indicate both smooth and obstructed traffic situations; they should be clarified in the visualization.
- Possibly retain geographic relations. Traffic flows of upstream and downstream cannot be distorted.
- Data of similar properties, such as weekdays and holidays, can be displayed side by side.
- Support data retrieval for experts to discover interesting traffic patterns.
- Achieve interactive performance.

### 3.3. Data preprocess

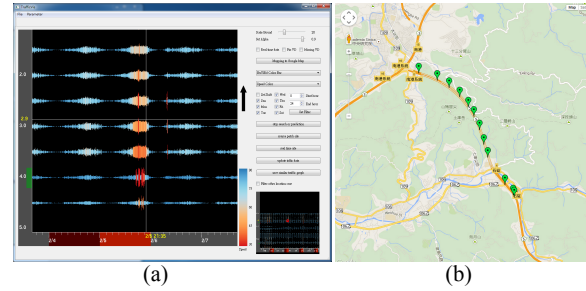
VDs are not evenly distributed on the freeway. In the areas close to an urban district, the distances between neighboring detectors are around 0.5 kilometer. In the areas close to countryside, however, the distances become around 2.0 kilometers. To prevent geographic distortions and to achieve clear visualization, we render a color-coded chart every  $d$  kilometers, where  $d = 0.5$  in our system. The speed, flow, and occupancy used to define the chart appearance are linearly interpolated from neighboring VDs. Although there are some other techniques such as Aw-Rascle-Zhang [AR00, Zha02] traffic model that can fulfill the task, we observed that linear interpolation works well because VDs are dense enough. Besides the uneven distribution of VDs, some VD data may be lost due to unexpected problems such as machinery breakdown or internet disconnection. In such cases, we let users decide whether to linearly interpolate the data because data missing often spans a large area. Even the state-of-the-art techniques cannot well handle this problem due to many interchanges on a freeway.

### 3.4. Color-coded charts and semantic zoom

**Color-coded charts.** We visualize each VD data item using a color-coded chart. The  $x$  coordinate represents a time span and the chart width at  $y$  coordinate indicates the occupancy. The color at each time span can show either speed or flow. It is determined based on the respective transfer function. We point out that the wider the vertical width, the higher occupancy. This visualization is designed according to the fact that speed and flow are important only when there are vehicles on the road, especially at local event examination. Therefore, at low occupancy, both low flows and low speeds indicate smooth traffic situations. In addition, our system visualizes speed and flow separately because traffic flows and speeds are usually analyzed in different scenarios; flows are used to measure road usage or to predict the area and time span of a traffic peak, and speeds are used to evaluate the smoothness of traffic or the effectiveness of a traffic control.

We design transfer functions for speed and flow, respectively, because these two variables have very different properties. The first variable shows how fast a vehicle can travel on the road. The second one indicates the number of vehicles that pass a region within a time span. In other words, the higher speed but the lighter flow, the better traffic quality is. Therefore, we map flow to green, yellow, and red as the flow increases. We also map speed to red, white, and blue in a similar way. Note that these two transfer functions are of different color styles, by which users can easily distinguish the variables when discovering our designed color-coded charts. Furthermore, both color styles highlight bad traffic situations in red to get attention from experts.

**User interface.** Figure 2 shows our user interface, which contains a visualization area, axes that depict spatial and

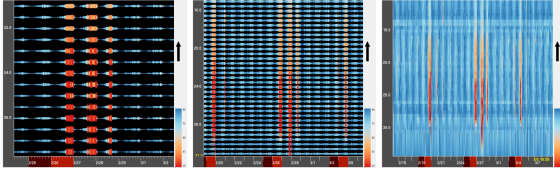


**Figure 2:** (a) Our system draws color-coded charts of a region within a time span to convey traffic situations. The vertical and horizontal axes represent VD's mileage and data time, respectively. The color-coded charts with the data interpolated are rendered with half-transparency. The black arrow indicates the vehicle moving direction. The color bar depicts the speed-color mapping. Notice that our system provides both focus (left) and context (bottom right) views for experts to maintain the overall image when they zoom-in to a small area of the data. The red rectangle in the context view indicates the region detailed in the focus view. (b) By a simple click, our system plots the locations of the visualized VDs on the map to reveal the geography and helps experts make sense of traffic situations.

temporal coordinates, vehicle traveling direction, and a transfer function. Our system renders color-coded charts on a black background. These charts are vertically stacked and horizontally aligned according to VD's mileage and data time, respectively. Considering different human mobility in weekdays and weekends, we highlight Saturday and Sunday using dark and light red, respectively, at the temporal coordinate. Our system also allows visualizing only a number of dates in a week for users to compare traffic patterns that have similar properties. For example, displaying only Sunday to know the difference of traffic situations when people go back to work, or the dates with hurricane strikes to know the human mobility under such bad weather situations.

To explore VD data at different scales and to obtain different viewpoints of a traffic event, our system allows users to zoom the data in either spatial, temporal, or both coordinates. They just need to scroll the mouse wheel when the cursor is over the spatial or temporal axes to achieve non-uniform zoom and over the charts to achieve uniform zoom. It is worth noting that, the presented visualization shows data for one road at a time. To analyze the traffic network, experts can create two or more windows by using our system and examine vehicle statistics on the intersecting roads. In addition, our system plots the locations of visualized VDs on the map to compensate for the lack of 2D geography. The link between the map and the color-coded charts help experts get insight into traffic situations.

**Semantic zoom.** The visualization of the whole VD data



**Figure 3:** Our system provides semantic zoom for users to explore traffic events at different scales. From left to right are the color-coded charts at local, meso, and global scales.

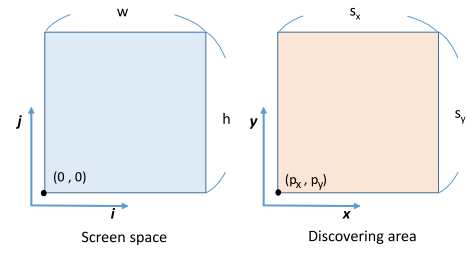
set that geographically span more than 300 kilometers and persist longer than a year would prevent analysts from data exploration (1292 VDs recording data every minute). Particularly, the vertical width of a chart would become difficult to view due to limited display space. Therefore, we present a semantic zoom for VD data visualization, in which chart appearances change according to the zoom level. That is, the vertical width gradually grows to maximum when users zoom out the visualization to the global side (Figure 3). This strategy enhances clarity of visualizing speed and flow at global scale to provide macroscopic visualization. Formally, denote by  $h$  the maximum height of a chart, by  $s_x$  and  $s_y$  the numbers of dates and VDs, respectively, we compute the vertical width of a chart at each time span using

$$T(h) = h^\alpha, \quad \text{where } \alpha = \min(s/s_{bound}, 1), \quad (1)$$

$s = \frac{s_x + s_y}{2}$  and  $s_{bound}$  is a parameter, which is experimentally set to 10 in our system. Note that this semantic zoom is continuous. Popping artifacts are not noticeable when changing the zoom level. We encourage the readers to watch our accompanying video because the smooth transition is difficult to appreciate in still images.

Our system allows semantic zoom of VD data in either spatial or temporal coordinate. This non-uniform zoom strategy enables investigating events that span an area with different sizes and that persist a time span with various lengths. For example, to study the traffic peak that happens after long holidays, analysts may want to zoom the spatial coordinate to a global scale but the temporal coordinate to a few days. On the other hand, to evaluate road usage, analysts have to zoom in to a small area but long time span. We will show the events investigated by using this non-uniform zoom later.

**Implementation details.** VD data visualization could be achieved by using either geometry based or pixel based rendering. The former draws the charts that are not clipped by screen. The latter is similar to ray tracing, which determines the color of a pixel according to the mapped VD data item. Our system adopts the pixel based approach to achieve interactive performance because of fixed computational cost when viewpoints are manipulated. For simplicity, we describe only the way of flow visualization and the speed visualization can be achieved in a similar strategy. As illustrated in Figure 4, our system allows users to specify the discov-



**Figure 4:** (Left) Our system allows users to specify the range of VD data that span  $s_y$  kilometers and  $s_x$  days during visualization. (Right) It then projects the data onto a screen with a resolution of  $w \times h$  using pixel based rendering to achieve interactive performance.

ering area by tuning parameters  $s_x$  and  $s_y$ , which represent the ranges of dates and mileages of VD, respectively. It then projects the selected data onto a screen with a resolution of  $w \times h$ . To implement this idea, we first consider the pixel positions, find out the time span and the corresponding VD data, and then determine its color. That is, we compute the color of pixel  $(i, j)$  using

$$C_{ij} = \begin{cases} M(f_{xy}) & \text{if } \frac{y}{d} - \frac{y_l}{d} < T(h_{xy}) \\ \text{background color} & \text{else} \end{cases} \quad (2)$$

where

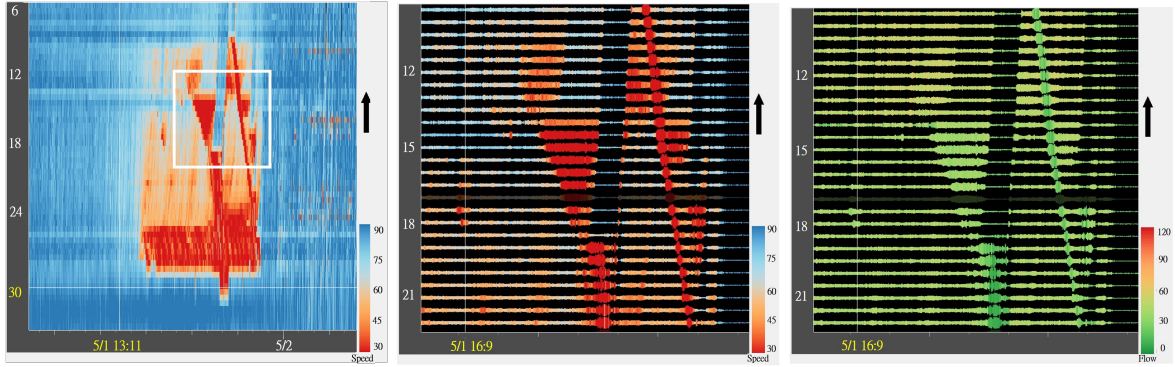
$$x = p_x + \frac{i}{w} \times s_x, \quad y = p_y + \frac{j}{h} \times s_y,$$

$(p_x, p_y)$  is the bottom left corner of the selected data,  $x_l = \lfloor x + 0.5 \rfloor$ ,  $y_l = \lfloor y + 0.5 \rfloor$ ,  $d = 0.5$  is the road length that a VD data item covers (Section 3.3),  $M$  is the color mapping function, and  $h_{xy}$  and  $f_{xy}$  denotes the occupancy and flow of VD data at  $(x_l, y_l)$ , respectively.

The main advantage of this technique is that pixels are independent to each other and their colors can be computed in parallel. This characteristic achieves stable performance when different scales of data are visualized. Note that a pixel may map to many values when VD data are observed from a global viewpoint. We average the values under this circumstance to indicate the overall situation.

### 3.5. Traffic pattern retrieval

Experts from the traffic management centers asked us to provide a graphical interface for them to retrieve interested traffic patterns. They proposed the requirement because they attempt to know where a specific event, such as a wide moving jam or a vehicle accident, usually occurs. A fast retrieval will help them isolate road segments that are not well designed or demand broadening. Moreover, when they observe some traffic patterns that have never been thought, they can retrieve similar patterns in the database and discover additional attributes to understand the phenomenon. Accordingly, we



**Figure 5:** (Left) The visualization of speed bottlenecks. To investigate the moving bottlenecks, we zoom into the region of interest (highlighted using a white rectangle) and show the color-coded charts that represent speed (middle) and flow (right).

add a retrieval-by-example function to our system; by specifying a region in the viewing window, the system points out similar patterns in the database for further investigation.

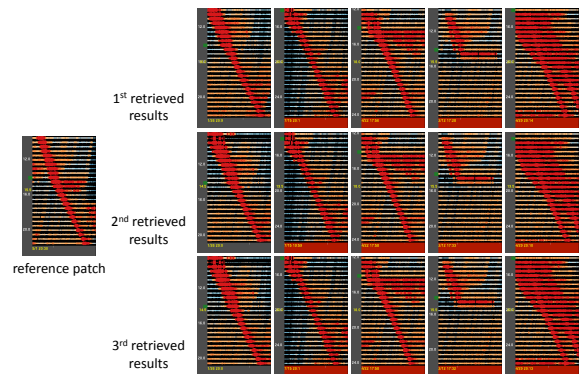
We define the distance of traffic patterns by considering speed, flow, and occupancy. Due to different ranges of these attributes, traffic data are normalized with a set of reasonable maximal values. Namely, the distance of two arbitrary patterns is given by

$$D = \left( \frac{F_s - F_t}{F_{max}} \right)^2 + \left( \frac{S_s - S_t}{S_{max}} \right)^2 + \left( \frac{O_s - O_t}{O_{max}} \right)^2, \quad (3)$$

where  $F$ ,  $S$ , and  $O$  denote flow, speed, and occupancy, respectively, the subscripts  $s$  and  $t$  indicate two patch indexes. We set  $F_{max} = 40h$ , where  $h$  is the number of lanes,  $S_{max} = 100$ , and  $O_{max} = 1$  in our system.

Due to a large database, which contains 525,600 minutes and 2,994 VDs per year, retrieving traffic patterns is expensive. To achieve interactive performance, we apply a coarse-to-fine random search algorithm to find approximated results. That is, we build a VD pyramid, randomly sample patches at the coarsest level, and iteratively refine patch positions until the searching process reaches the finest level. Typically, at the coarsest level, either spatial or temporal resolution of the reference patch is reduced to one unit. Denote by  $\ell$  the pyramid level, where  $\ell = 1$  indicates the coarsest. Based on the sampling theory indicated in [BSFG09], at level  $\ell$ , we randomly sample  $\frac{M}{m n \ell}$  patches in the database, where  $M$  and  $m$  are the resolutions of the database and the reference patch, respectively, and  $n$  is the number of patches that users attempt to retrieve ( $n = 5$  in our implementation). After computing distances to the reference patch (Equation 3), our system discards  $\frac{1}{\ell_{max}}$  of the patches that have larger distances and enters the next level, where  $\ell_{max}$  is the number of levels in the pyramid.

Considering similar patches potentially appear at the regions with neighboring time periods but different days, we



**Figure 6:** We retrieve traffic patterns similar to the reference patch three times. The patches from the most similar to the fifth similar are listed from left to right. The retrieved patterns are generally the same, except their starting time spans and mileages are a little bit different.

additionally seed a patch every week at the same time and the same day to the reference patch to ensure that more similar traffic patterns are considered. Note that we sample patches in the whole database only at the coarsest level. When the retrieving process gets into a finer level, only neighboring regions of the patches obtained from the previous level are tested. The searching area can become smaller at fine levels because rough comparisons have been done previously. The computational cost is therefore reduced. In our implementation, we set the searching area to have the resolution  $\frac{W_\ell}{2^{\ell-1}}$  and  $\frac{H_\ell}{2^{\ell-1}}$ , where  $W_\ell$  and  $H_\ell$  denote the width and height of the database at level  $\ell$ , respectively. Although the presented random search is only an approximate solution, this process generally obtains stable results when a traffic pattern is given. Figure 6 shows an example, in which moving bottleneck patterns are retrieved three times. As indicated, the retrieved patterns are generally the same. While random

sampling is used, these stable retrieving results show that our system can approximate global optimum in most cases.

#### 4. Results

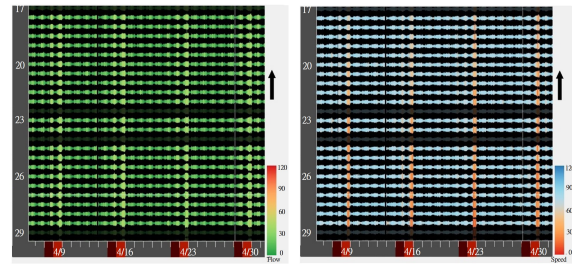
We have implemented our system and run the code on a desktop PC with Core i7 3.4 GHz CPU. The performance depends on the resolution of a view window due to the adoption of pixel-based rendering. We also apply OpenMP to enable parallel computing because pixel colors are independent to each other. This implementation performs in real time when the view window is set to  $800 \times 800$  pixels. We believe that leveraging the GPU would boost the performance, yet our current system is sufficient to provide interactive data exploration and investigation. In the case of traffic pattern retrieval, the process takes 0.8 seconds in average when retrieving the wide moving jam example (Figure 6), in which the patch spans 1.5 hours and 12 kilometers. Retrieving a large patch also is efficient. It takes 5.8 seconds when retrieving the hurricane example (Figure 8), in which the patch spans 23 hours and 371 kilometers. In other words, long waiting time is not necessary, thanks to the coarse-to-fine random search approach.

##### 4.1. Case studies

We explore the traffic data obtained from 2994 VDs of eight freeways and from 2011 to 2013. Below we show the events occurred at different spatial and temporal scales, and interest experts from traffic management and research centers.

**Moving bottleneck.** Traffic bottleneck easily appears when there are too many vehicles on the road. In Figure 5 (left), the region in red indicates a traffic obstruction that spanned 18 kilometers and persisted around 8 hours. Analysts can zoom into the data and observe obstruction details, as shown in Figure 5 (middle and right). Specifically, an oblique red line (starts from 9<sup>th</sup> kilometer) on the stacked color-coded charts indicates a moving bottleneck. Given that vehicles at the head of this bottleneck could accelerate but those at the tail would queue, the bottleneck moved upstream over time, which was the direction opposite to that of the traffic flow. In addition, when the vehicle flows at downstream were slow but at the upstream were fast, this bottleneck would grow and appear as an inverse triangle (starts from 14<sup>th</sup> kilometer). For the low occupancy at the right side of this inverse triangle, the cause to this phenomenon was not clear. We suspected there should be an official roadblock.

**Weekly traffic peak.** Figure 7 shows that visual patterns were very different at weekdays and weekends. This phenomenon was especially obvious at regions that were close to urban districts. In this example, we visualize the freeway segment that connected Taipei, the capital of Taiwan, and its backyard, Ilan. The travel direction was from Ilan to Taipei. Obviously, visual patterns at weekends indicate high flow and occupancy but low speed. These visual patterns show



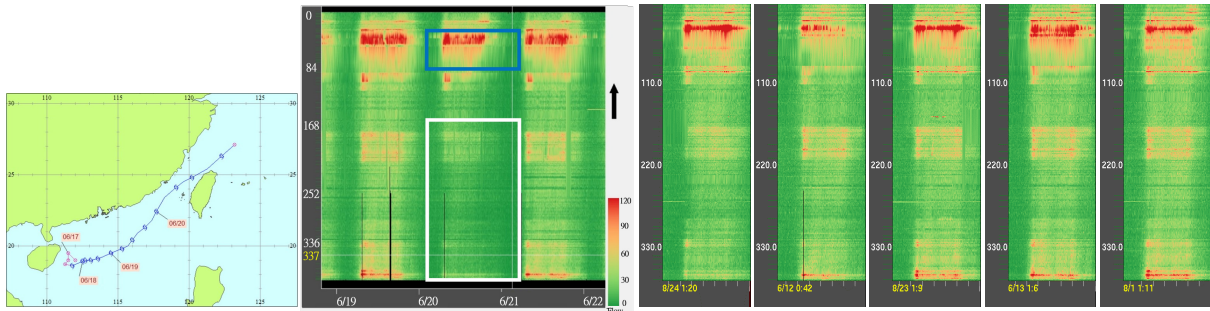
**Figure 7:** Left and right images show the flow and speed of VD data from Ilan to Taipei. The visual patterns have different appearances at weekdays and weekends because of different travel purposes.

that more vehicles traveled from Ilan to Taipei. The highest peak appeared at the daytime of Sunday because citizens had to go back to work on Monday. We point out that the flow and occupancy at Sunday were not visibly larger than those at Saturday but the speed at Sunday was apparently slower. This phenomenon was due to unstable traffic situation when the number of vehicles was larger than a threshold.

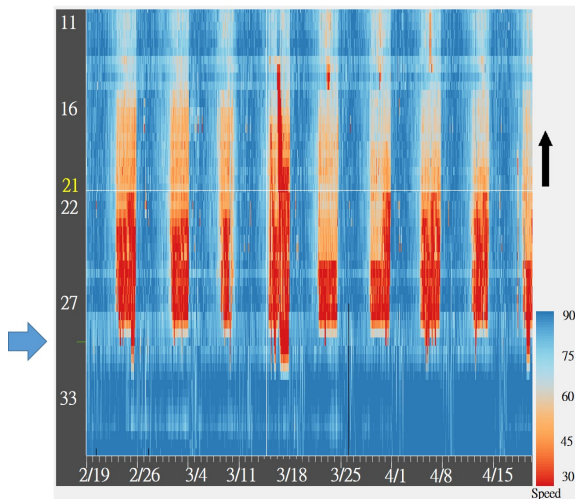
**Hurricane.** There were hurricanes passing through Taiwan in summer. Torrential rain and strong winds usually hit this island during the visiting. Figure 8 (left) illustrates the moving path of a hurricane. This hurricane was close to the southern part of Taiwan in the morning of June 20. Its influence to the south area was large compared to that in the north area. Therefore, class and work suspensions were issued by only governments in south but not in north, which also made the traffic flows different in these two areas. As can be seen in Figure 8 (middle), at the middle and southern parts of the freeway, both the traffic flows heading north and south on June 20 were different to those on June 19 and June 21, respectively, because most citizens stayed at home during the hurricane strike. In contrast, also at June 20, the traffic flow at north had no difference compared to that of its neighboring weekdays.

We retrieved traffic patterns similar to that of June 20 and show the results in Figure 8 (right). After checking the dates of these patterns, we found that hurricanes struck the island at all of these dates. Our retrieval by example allows the experts to find traffic events that happen at irregular time.

**Tunnel.** Hsuehshan tunnel was the fifth longest road tunnel in the world. The range of this tunnel was 15.2-28.1 kilometers, with the total length of 12.942 kilometers. This long tunnel brought a very special traffic phenomenon. In Figure 9, we visualize vehicle speed of consecutive Sundays because they had similar properties. Clearly, vehicle speeds in the afternoons were usually slow in the tunnel. The speeds at the mileage of 22-28 kilometers usually were the slowest, became faster at the mileage of 22-15 kilometers, and raised to the normal level after the mileage of 15 kilometers. We



**Figure 8:** (Left) The path of a hurricane in June. (Middle) The traffic flows heading north. Because the class and work suspension were not issued by the government at north at June 20, the traffic flows at north remain the same compared to those at weekdays, as highlighted by blue rectangles. However, most citizens in the middle and southern parts of Taiwan stayed at home during the hurricane strike so that traffic flows became much lighter, as highlighted in white rectangles. (Right) We retrieved traffic patterns similar to that of Jun 20 and found that hurricanes struck the island at all these dates.



**Figure 9:** We visualize the vehicle speed of consecutive Sundays in Hsuehshan tunnel. The visual patterns show that speed was slow in the tunnel. One possible reason for this phenomenon was defensive driving style in the tunnel. Another reason could be vehicles entering the freeway from the interchange highlighted using a blue arrow.

discussed with the experts from the freeway bureau and suspect the phenomenon was due to defensive driving in the tunnel. Given that drivers were not allowed to change lanes in this closed space, the vehicle speed decreased rapidly when there are slow vehicles at the head of a flow. These defensive drivers potentially raised the speed when they became comfortable to the tunnel and totally recovered when they left the closed space. However, this traffic phenomenon might also be due to the interchange close to the tunnel entrance. The experts were interested in this new finding. They will further

study the phenomenon to understand the cause of formation and apply strategies to increase vehicle speeds.

### 5. User study and feedback

We conducted a user study by interviewing experts after our prototype system was built. Specifically, two experts are from the National Freeway Bureau of Taiwan, two from the Institute of Transportation in the Ministry of Transportation and Communication, and one from the Department of Transportation Engineering and Logistics Management. This study was qualitative rather than quantitative because VD data interpretation requires domain knowledge. In this study, we explained the concepts of our visualization, including the interface, the design of color-coded charts and the idea of semantic zoom. Then, we let these experts freely interact with the system and explore the recorded VD data by themselves. Once they got familiar with our design, we showed them the traffic events that we had observed before. Generally, the training procedure took about 20 minutes. They learned very quickly, just like they had used the system before. Finally, we asked for the feedback related to our visualization designs and confirmed the cause of formation of these events. The interviews lasted about two hours in average.

**Data collection and interpolation.** The experts from the traffic management center of national freeway bureau mentioned that VD data are usually used to study traffic events so as to enable prediction and precise control. Other data such as GPS trajectories are utilized to estimate travel time between two locations. They were also proud of the high availability of their recorded VD data, which was about 90%. In the case of data interpolation, they did not think the process was necessary because the interpolation was based on only a mathematic model. However, the real traffic situation may change due to weather, aggressive and defensive driv-



ing styles, and interchanges. These factors are difficult to estimate. So they would rather keep the missing data to remain missing. The experts from the institute of transportation also had the same opinion although they used to interpolate missing data before analyzing them using a computational model. They think data interpolation is not always necessary in the visual analysis of traffic patterns.

**Visualization design.** All the experts expressed that they did not see any visualization tools designed for spatiotemporal VD data. Currently, they can only render roads with different colors to indicate the traffic situation of a time span. However, if they attempt to understand a spatiotemporal event, they have to read a sequence of images or a limited-size numerical table, which is tedious and time consuming. Therefore, they confirmed the intuition of our color-coded chart design because it can clearly show speed, occupancy, and flow of VD data at every minute and at every location. The design fits the way they think and interpret traffic data. That was the reason they could quickly find interesting events from the data after our training. Generally, the experts believe that the separate visualization of vehicle flows and speeds is reasonable because of different objectives. For example, flows are usually studied before the prediction of traffic situations; experts attempt to know how many people will move and when they will move before the holidays. On the other hand, speeds are often studied for evaluating the effectiveness of traffic controls because the goal of traffic management is keeping freeways unobstructed. In addition, they appreciated the use of vertical chart width to represent occupancy because the visual pattern automatically fades out when the traffic situation is smooth. Given that traffic flow is a very special variable, where a low value may represent either obstruction or smoothness, our visualization design can clearly distinguish the situations. Overall, the experts were very happy to see this visualization tool because they can interpret traffic data efficiently. Our visualization enables them to predict traffic situations at multiple regions with various time spans and enhance the effectiveness of traffic controls.

**Suggestions.** The first version of our system aligns VD data according to physical time so that charts that represent consecutive days were adjacent. However, the experts advised that traffic situations are closely related to working days and holidays. It would be very useful if users can freely align the dates that have relations. For example, by gathering all traffic flows at Sunday, we can easily compare the difference of human mobility when the holiday is over. Therefore, we took their suggestions and added this new operation into our system (Figure 9). They also found that traffic flows at Monday mornings and Friday afternoons were similar but in opposite directions by using this new interface.

Given by the ability of traffic pattern retrieval, the experts also suggested us developing a streaming system that can predict traffic situations over time. Specifically, the system retrieves current traffic patterns in the historical data set and

predicts future situations according to the data temporally adjacent to the retrieved patterns. This strategy is reasonable because every why has a wherefore. We will take the suggestion and develop the system in future.

### 5.1. Limitations

Our system visualizes VD data using color-coded charts. Although the charts are vertically stacked according to their mileages, still one dimension of the geographic information is lost. Users may need to switch the map and the visualization window when discovering an event that involves with a complex traffic network. Solving this problem is difficult because VD data are of multi-dimension (geography, occupancy, flow, and speed) but our display devices are only planar. We plan to further reduce the distortion in future. Moreover, our system linearly interpolates the VD data that were lost due to machinery breakdown or internet disconnection. The strategy easily causes stretching artifacts when the range of missing data is large. It also does not support extrapolation and leaves the data being empty because the artifacts would become even worse under this circumstance.

## 6. Conclusion

We introduce a visualization system that allows users to discover and investigate VD data interactively. Compared to previous methods that focus on vehicle GPS trajectories, our visualization provides analysts with insightful visual patterns to interpret traffic data before they make crucial decisions. Our designed color-coded charts depict the relations of speed-occupancy and flow-occupancy at a region over time. The space efficient design allows the system to stack and align the data for investigating traffic events that often span in spatial and temporal coordinates. To create high quality visual experience and to enhance data clarity at different scales, our system transits the chart appearances seamlessly. This semantic zoom shows occupancy to make clear local traffic situations and deforms charts to rectangles to highlight global changes during data exploration.

We have presented our prototype system to experts who worked in traffic management and research centers. They studied VD data in their daily work for many years and confirmed the practical usability of our system. Due to no visualization tools that could help them conveniently interpret spatiotemporal traffic data before, they were very happy to use our system and keep working with us in future.

### Acknowledgments

We thank the anonymous reviewers for their constructive comments. We are also grateful to Chen-Yi Huang for producing the demo video. This work was supported in part by the Ministry of Science and Technology, Taiwan (102-2221-E-009-083-MY3, 102-2221-E-009-082-MY3, 101-2628-E-009-021-MY3, and 102-2221-E-009-081-MY3)

## References

- [AA08] ANDRIENKO G., ANDRIENKO N.: Spatio-temporal aggregation for visual analysis of movements. In *IEEE Symposium on Visual Analytics Science and Technology (VAST)* (2008), pp. 51–58. [2](#)
- [AA11] ADRIENKO N., ADRIENKO G.: Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics* 17, 2 (2011), 205–219. [2](#)
- [AR00] AW A., RASCLE M.: Resurrection of "second order" models of traffic flow. *SIAM Journal on Applied Mathematics* 60, 3 (2000), pp. 916–938. [4](#)
- [BBBL11] BOYANDIN I., BERTINI E., BAK P., LALANNE D.: Flowstrates: An approach for visual exploration of temporal origin-destination data. *Computer Graphics Forum* 30 (2011), 971–980. [2](#)
- [BMH\*12] BAK P., MARDER M., HARARY S., YAELI A., SHIP H. J.: Scalable detection of spatiotemporal encounters in historical movement data. *Computer Graphics Forum* 31, 3pt1 (2012), 915–924. [2](#)
- [BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A., GOLDMAN D. B.: PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 28, 3 (Aug. 2009). [6](#)
- [CLT\*11] CUI W., LIU S., TAN L., SHI C., SONG Y., GAO Z., QU H., TONG X.: Textflow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2412–2421. [2](#)
- [DFD\*14] DORAISWAMY H., FERREIRA N., DAMOULAS T., FREIRE J., SILVA C.: Using topological analysis to support event-guided exploration in urban data. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2634–2643. [2](#)
- [FPV\*13] FERREIRA N., POCO J., VO H. T., FREIRE J., SILVA C. T.: 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 (2013), 2149–2158. [2](#)
- [GWY\*11] GUO H., WANG Z., YU B., ZHAO H., YUAN X.: Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *Pacific Visualization Symposium (PacificVis)* (2011), pp. 163–170. [2](#)
- [GZJ\*12] GUO D., ZHU X., JIN H., GAO P., ANDRIS C.: Discovering spatial patterns in origin-destination mobility data. *T. GIS* 16, 3 (2012), 411–429. [2](#)
- [HHN00] HAVRE S., HETZLER B., NOWELL L.: Themeriver: visualizing theme changes over time. In *IEEE Symposium on Information Visualization (InfoVis)* (2000), pp. 115–123. [2](#)
- [HHWN02] HAVRE S., HETZLER E., WHITNEY P., NOWELL L.: Themeriver: visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics* 8, 1 (2002), 9–20. [2](#)
- [JFG12] JIANG S., FERREIRA JR. J., GONZALEZ M. C.: Discovering urban spatial-temporal structure from human activity patterns. In *ACM SIGKDD International Workshop on Urban Computing* (2012), UrbComp, pp. 95–102. [2](#)
- [KMG\*06] KONYHA Z., MATKOVIC K., GRACANIN D., JELOVIC M., HAUSER H.: Interactive visual analysis of families of function graphs. *IEEE Transactions on Visualization and Computer Graphics* 12, 6 (2006), 1373–1385. [2](#)
- [LGL\*11] LIU H., GAO Y., LU L., LIU S., QU H., NI L.: Visual analysis of route diversity. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2011), pp. 171–180. [2](#)
- [TR09] THAKUR S., RHYNE T.-M.: Data vases: 2d and 3d plots for visualizing multiple time series. In *International Symposium on Advances in Visual Computing: Part II* (2009), ISVC '09, pp. 929–938. [2](#)
- [TSAA12] TOMINSKI C., SCHUMANN H., ANDRIENKO G., ANDRIENKO N.: Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2565–2574. [2](#)
- [Van09] VANDANIKER M.: Visualizing real-time and archived traffic incident data. In *IEEE International Conference on Information Reuse Integration* (Aug 2009), pp. 206–211. doi: [10.1109/TRI.2009.5211552](https://doi.org/10.1109/TRI.2009.5211552). [2](#)
- [VLBA\*12] VON LANDESBERGER T., BREMM S., ANDRIENKO N., ANDRIENKO G., TEKUSOVA M.: Visual analytics methods for categoric spatio-temporal data. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2012), pp. 183–192. [2](#)
- [WAM01] WEBER M., ALEXA M., MULLER W.: Visualizing time-series on spirals. In *IEEE Symposium on Information Visualization (INFOVIS)* (2001), pp. 7–13. [3](#)
- [Wat05] WATTENBERG M.: Baby names, visualization, and social data analysis. In *IEEE Symposium on Information Visualization (INFOVIS)* (2005), pp. 1–7. [2](#)
- [WDS10] WOOD J., DYKES J., SLINGSBY A.: Visualisation of origins, destinations and flows with od maps. *The Cartographic Journal* 47, 2 (2010), 117–129. [2](#)
- [WK06] WATTENBERG M., KRIS J.: Designing for social data analysis. *IEEE Transactions on Visualization and Computer Graphics* 12, 4 (2006), 549–557. [2](#)
- [WSD11] WOOD J., SLINGSBY A., DYKES J.: Visualizing the dynamics of london's bicycle-hire scheme. *The Cartographic Journal* 46, 4 (2011), 239–251. [2](#)
- [YZX12] YUAN J., ZHENG Y., XIE X.: Discovering regions of different functions in a city using human mobility and pois. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2012), KDD, ACM, pp. 186–194. [2](#)
- [ZFAQ13] ZENG W., FU C.-W., ARISONA S. M., QU H.: Visualizing interchange patterns in massive movement data. *Computer Graphics Forum* 32, 3 (2013), 271–280. [2](#)
- [Zha02] ZHANG H.: A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research Part B: Methodological* 36, 3 (2002), 275 – 290. [4](#)