QuteVis: Visually Studying Transportation Patterns Using Multi-Sketch Query of Joint Traffic Situations

Shamal AL-Dohuki  
Kent State University

Ye Zhao  
Kent State University

Farah Kamw  
Kent State University

David Sheets  
Kent State University

Jing Yang  
UNC Charlotte

Xinyue Ye  
New Jersey Institute of Technology

Wei Chen  
Zhejiang University

Abstract—QuteVis is a visual analytics system that supports the study of urban traffic patterns. Unlike most existing approaches which investigate traffic patterns in user specified spatial regions and temporal periods, QuteVis supports a different type of data query and analytical tasks. Multi-sketch query and visualization helps users discover those specific times and days in history, which have specified joint traffic patterns distributed on different city locations. Users can use touch input devices to define, edit, and modify multi-sketch on a city map. A set of visualizations and interactions are provided to help users browse and compare the retrieved traffic situations at multiple and spatially diverse geo-locations, and a visual system is demanded to investigate when and how such patterns happen from transport database.

In this paper, we present a visual analytics system, named QuteVis, which helps users interactively (q)uery (u)rban (t)ransport databas(e) (Qute) to conduct these tasks over large scale datasets. We have addressed several major challenges in this new VA application system. First, it is demanded that QuteVis should allow users to conveniently specify multiple conditions at separate locations on the map. We then develop multi-sketch query so that users can direct draw on a city map by hand or by mouse to define traffic attributes (e.g., traffic speed) as multiple input sketches. Reliable sketch recognition interface is designed with free style sketches. The system further allows users to edit and modify sketches with easy interactions. Second, the user sketch should be applied to different geo-structures at both large and fine scales. In a large scale, users may draw a long path that cover a sequence of streets. On the other hand, users may also sketch to select a small street segment (e.g. close to an intersection). Typical street network geometry from GIS databases does not support such operations. We then design an optimal geographical data structure to support interactive sketches on street segments, paths, and regions, at different scales.

Third, interactivity is the key to effective visual query. Thus data query response over heterogeneous data should be fast enough to enable interactive visualization. We develop a new database to efficiently store and retrieve the traffic information, spatial information (e.g., city roads and POIs), mobility information (e.g., vehicle pickups/dropoffs and trajectories), and other information (e.g., weather condition). Data retrieval speed is accelerated by geo-spatial indexing and a specific data cube which plays an indispensable role to enable smooth user interaction.

Fourth, a global “weighted similarity” is computed for multiple sketches, which is used to quickly find matching results in massive traffic data records at different time periods. Moreover, a weight is assigned to each spatial sketch which is customized to user priority. For example, a large weight may be given to a major street, while small weights are assigned to secondary streets. These similarities on multiple sketches are visualized to provide cues for users to investigate their preferred results from the joint conditions.

QuteVis system presents a set of visualizations within a sketch+visualization interface. It facilitates users to overview the distribution of joint traffic patterns from their sketch inputs. They can further investigate top matched time periods, and select any interesting time periods or days for...
drill-down analysis. Moreover, a multi-map comparison view allows them to compare traffic situations in the whole city and/or at any specific locations. In this paper, case studies and domain expert feedback are presented to illustrate its usability and efficiency.

In summary, the main contributions of this paper include:

• A different type of investigation model: “what+where⇒when+how” is supported to study urban transportation data. To the best of our knowledge, it is the first visual analytics system devoted to such tasks other than the “where+when⇒what” data access model.
• Joint traffic situations can be specified on multiple locations that can distribute separately in space, which have not been supported in existing transportation application tools.
• Multi-sketches are supported by reliable sketch recognition algorithms for different types of sketches on streets, paths and regions.
• The QuteVis system integrates sketch+visualization interface and heterogeneous data management is implemented for domain users to conduct interactive visual query and drill-down study.

2 RELATED WORK

Transportation Data Visualization and Interaction There has been extensive research on visual analytics of traffic and various movement data [2]. Typically, map-based displays and information visualization techniques were combined in visualizing the spatio-temporal data.

Considering transportation data types, these approaches may be divided into three directions: (1) VA systems study recorded human/vehicle trajectory data: Ferreira et al. [5] allowed users to visually query taxi trips on spatio-temporal constraints. TrajectoryLenses [9] used lenses to support visual, set-based filter expressions to select trajectories. Al-Dohuki et al. [1] transformed taxi trajectories into texts consisting of street names and text labels denoting taxi speeds for users to query a taxi trajectory database using a text search engine. Our system does not work on analyzing trajectories. Instead, we allow users to query traffic information on city structures which is computed from trajectories. (2) VA methods retrieve and query traffic flow data: Sheepens et al. [14] visualized the directions of traffic flows using a particle system on top of the density map so that users can extract traffic flows using a selection widget of an area and a range of directions. (3) VA methods investigate aggregated traffic data on city structures: Wang et al. [20] explored traffic data recorded on sparsely distributed cells in a city. Pu et al. [13] visually monitored and analyzed complex traffic situations in big cities. Wang et al. [19] presented a visual reasoning approach for the data-driven transport assessment on urban roads. These approaches are designed mostly based on the data access model of “when+where queries”, by giving both spatial and temporal constraints to retrieve traffic data and then perform analysis. In contrast, our VA approach is the first to supports “what+where” queries to help users examine similar times and days that approximately have the user-specified values on multiple given streets, paths and regions.

Sketch-Based Visual Analytics Sketch-based interfaces and modeling has been studied to enable users to interact with a computer through sketching. Unlike our sketch based method that working on urban maps, many existing systems are designed to sketch over different types visual charts, glyphs, and graphics. Browne et al. [4] described the design and realization of SketchVis that leverages hand-drawn input for exploring data through simple charts. Visualization-by-Sketching [15] enabled artists and other visual experts to create accurate and expressive data visualizations by painting on top of a digital data canvas. SketchStory [10] allowed the presenter to record a sequence of charts along with direct hand-drawn annotations. Our system extends these efforts in multiple ways: (1) Joint traffic situations can be specified on multiple locations that can distribute separately in space, which have not been supported in existing transportation application tools. (2) We automatically identify user sketching of different objects including streets, paths, and regions, while the existing work usually supports one predefined sketching type; (3) We accept free drawings for multiple paths and regions on the map, and these drawings are considered as joint segments. To enhance the sketch accuracy, we optimize street network geometry by artificially generating street segments with preferred length; (4) We automatically identify user sketching of different objects including streets, paths, and regions, while the existing work usually supports one predefined sketching type; (3) We accept free drawings for multiple paths and regions on the map, and these drawings are considered as joint spatial conditions for data retrieval; (4) Finally, our system allows users to perform easy editing of the sketch results. In particular, users can add and remove street segments from existing selections, which has not been make easy in existing tools.

3 OVERVIEW

3.1 Goal and Tasks

The objective of QuteVis was formed by interviewing five experts (named as 5ExP) who have been working on urban planning and transportation analysis. 5ExP appreciated the success of existing VA systems of traffic data analytics based on their experience and our introduction. They realized the existing work mostly focus on retrieving and visualizing traffic data based on given spatio-temporal conditions. Then they proposed the following example tasks which are important but not well supported:

• Assuming a highway bridge \( H \) and a street \( M \) are two critical traffic nodes, in two separate regions of a city. What are the specific days and times if both of them have traffic jams? What are the distributions of the matched results on different periods (e.g., morning, afternoon, night), different days (e.g., Mon, Tue, etc.), and different weather conditions? and what is the traffic situation in a nearby residential district in such cases?

• These questions inspire us to develop QuteVis system which aims to help users “see” and examine the traffic patterns in large volume of historical traffic data by giving their speculated values in separate location all over the city, so that they can test their hypotheses with real datasets.

We further discussed system requirements with 5ExP, who guided our design from domain users’ perspective. A key challenge is to allow users to quickly speculate traffic behaviors and immediately get visual feedback from large scale historical time-varying traffic data together with related attributes such as weather and taxi activities. In particular, several major tasks are identified:

• T1: Flexible multiple selections and editing: Users should be able to quickly locate and select their preferred urban structures (including streets, paths or regions) over urban geographical context. They should easily set up query conditions such as the values of traffic speed or taxi pickups. Users should also be able to edit their selections and values. It would be better if hints can be given about historical values at specific locations for users to set up query conditions.
Fig. 1: *QuteVis System framework of visual query, database retrieval, and visual analytics.*

Fig. 2: *QuteVis interface: supporting visual queries and analytics of historical days and times that have similar traffic patterns speculated by users. Users can sketch on the map canvas to select streets and regions that sparsely distributed in a city.*
• T2: Fast response and interactive exploration: The system should provide immediate response for queries and the results should be visualized quickly so that users can iteratively explore the large data.
• T3: Easy perception and understanding: Visualization of the query results should be straightforward for users to discern and understand. Visualization may not focus on small differences in numerical values (e.g., speeds of 20 or 22 km/hours are not distinctive in traffic analysis).
• T4: Visual comparison: Visual comparison of the situations in multiple times/days and at different locations is an essential function which should be supported. Multiple locations diversely distributed in a city should be compared together.

3.2 System Design
QuteVis thus is designed as an interactive VA system combining user sketching and visualization, which is supported by efficient data integration and management. The system framework is illustrated in Fig. 1 consisting of major functional modules.

Multi-Sketching on the map plays a key role which facilitates easy user selection and editing of multiple geo-objects at different locations on the map (for T1). Users can define their speculated patterns by (1) selecting multiple, spatially diverse locations by drawing on a city map. The selected objects (i.e., sketches) can be streets or paths, as well as arbitrarily shaped regions in a city; (2) Specifying query values at these locations about transport attributes (e.g., traffic speed), with visual hints of their historical values to guide their input. Street geometries are optimized for accurate and flexible sketching; and meanwhile the sketch processing are well implemented to enable prompt responses (for T2).

The interactive sketches are seamlessly combined with visual interface for immediate visual analysis (for T2 and T3). The data records of different times at different days to the sketched queries are visualized for users to interactively study distributions of top-matching results in different periods (e.g., morning, afternoon), different weekdays, and different weather conditions. With the suggestion from domain users, traffic data records at different times and days are categorized into not similar, little similar, very similar and extremely similar by their similarities to the given traffic pattern of multiple query sketches (for T3). This global similarity is jointly computed from individual similarities of the multiple sketches by the given weights. User can customize these weights of importance to adjust the contributions from the sketches. Moreover, QuteVis provides a convenient multi-map view as an effective tool for comparative study of top matching times or any interesting times (for T4). Users can examine transportation features on any location of the city.

3.3 QuteVis Interface

The interface is designed as a map-centric application where the map plays the role as a canvas in the middle for users to sketch on. This design gives users maximal capacity to draw and edit. Fig. 2 shows the visual interface which includes the following major views:

• A canvas over city map, Fig. 2(1), to facilitate context-aware operations on the city. In Sketch Mode, it supports the selection of multiple streets, regions, or paths, where three sketches are shown (in the following, we also refer a sketch as one selection). To help users differentiate between multiple sketches, different colors are used to encode multiple sketches on the canvas. Users can also edit the sketches interactively. In Info Mode, the map view visualizes traffic data, POIs, and other urban information, and also users can interact with the map view for their exploratory study. Users can select any region/street to study its traffic behavior and weather in a specific time. The map view can be shown in different styles and it supports smooth zooming and panning operations.

• A control panel with multiple tabs (Fig. 2(2)). The tabs include: (a) Manage Sketches: which allows users to define query conditions on multiple sketches, and to control whether the sketches are active in a query; (b) Study Distribution: which helps users investigate the distributions of similar times over weekdays, hours, and weather conditions; (c) Find Similar Times: which shows a grid heatmap view for users to find and study the similarities of available historical times of days. Those candidate times with the largest similarities are highlighted and put into a top list by default, while users can also select any arbitrary times of their interest and add them to the top list.

• A detail study panel (Fig. 2(3)) to investigate selected times in the Top List. Top List are visualized for users to study detail information of each time in the list, including the input value and actual value on each sketch, similarity categories, and the weather condition. Furthermore, Map Comparison allows users to observe and compare the traffic behaviors at different times on a multi-map view. This view is coordinated with the major map canvas. Fully functioned map operations on either the canvas map or any of the small maps are well synchronized. Therefore, users can study details of any specific area of the city.

• A top panel (Figure 2(4)) allows users to toggle between Info Mode and Sketch Mode, quickly locate a street/POI by name, re-initialize the system, and set up sketching parameters.

4 Multi-Sketch on Geo-Structures
QuteVis need to support users to easily define and edit multiple sketches on the map. Its sketch recognition algorithm, unlike general methods, is built up on the geographical or urban structures. Users want to sketch on three different types of entities including street segments, path, and region which are targeted by the recognition algorithms.

4.1 Sketch Recognition and Editing Algorithms

Users can draw on the central map canvas (Fig. 2(1)) with a mouse, a touch pen or hand drawing (if supported). In Fig. 2(1), three sketches are selected, which represent three major streets (categorized as primary in Hangzhou’s street hierarchy) in downtown Hangzhou. One sketch is a set of small street segments selected from the optimal street network of a city (Sec. 5).

A street sketch recognition algorithm is used to map the user input on screen to the street segments, which consists of several steps:

1. When a user draws on the map (Fig. 3(a)), their touch operation is collected as a sequence of sketch points (Fig. 3(b)). These points are processed to remove obvious errors and noises.
2. A polyline is formed from the points by connecting consecutive input points. Then, a narrow band enclosing this polyline, while following its shape, is generated (Fig. 3(c)). This narrow band mimics the stroke of this sketch. The width of the stroke is defined adaptively according to the zooming level of the map, which plays an important role in defining the accuracy of the sketch selection. In implementation, this stroke band is geometrically represented as a bounding polygon of the polyline.
3. This stroke bounding polygon is used to apply a region-based spatial query on the road network (Fig. 3(d)), to retrieve all the inside street segments. For example, two segments in red and blue are selected in Fig. 3(e), which represent two separate lanes in opposite directions on the same road.
4. However, when users drawing on the map, the input stroke may collect unintended street segments, such as the blue segment in Fig. 3(e). On the other hand, sometimes a stroke may not cover a necessary segment which is needed. Then editing operations (Remove and Add) is required for users to edit any sketch by adding and removing related street segments. As shown in Fig. 3(f), users can click to remove the blue segment. Finally, the selected street segments form the result sketch. This algorithm allows users to choose a single street or a long path between two locations as well, which is flexible for traffic study.

In addition, a sketch can also be defined as a region in the city. In a region sketch recognition algorithm, the user-input points still create a polyline. Then we make this polyline as a closed curve by connecting all points, so that it forms a polygon. We then use this polygon to query the road network and all the street segments inside are selected. Similar editing operations are available to refine the sketch results. All these operations are available on traditional devices as well as touch input devices.

### 4.2 Improving Sketch with Street Segment Optimization

In the sketch algorithms, street segments are the main geographical elements which links human input to urban objects. Therefore, the scale of street segments defines the sketch selection resolution. Also, the segments should not only represent the city road network but also (1) facilitate accurate sketch selection and edit, and (2) enable high-resolution traffic data storage and management. However, an important challenge is that the “raw” road segments retrieved directly from an existing geo-database (e.g. from OpenStreetMap) cannot be directly used in QuteVis sketch operations.

There exist many problems of these raw street segments. A raw street segment may be very long that pass multiple road intersections, or it may have a complex geometric shape. In these cases, a sketch that intends to select a small street part will return a long path or a complex shape than expected. On the other hand, many street segments have numeric errors. These problems make it hard for users to interactively select desired streets or paths on map, leading to frustrated sketching experience.

Our goal is to provide satisfied sketch accuracy and experience by regenerating optimized street segments. The algorithm has two stages: First, we define new street segments to make sure their endpoints located at neighbor intersections. This is implemented by (1) finding all geometric intersections from the raw street segments; and (2) defining one new segment between each pair of the neighbor intersections. Second, in the new set of street segments, we divide a long segment into multiple small ones. The purpose is to make each street segment short enough so that people can sketch to select them with high resolution. This is very important to ensure accurate object selection and study in sketching operations. In implementation, a total number of 14,639 street segments are generated for Hangzhou, while the raw road network from OpenStreetMap has 9,764 segments.

### 4.3 Managing Multiple Sketches

After users draw multiple sketches on the canvas map as in Fig. 2(1), users need to specify an input for each sketch with the value of their preferred traffic attributes (speed, volume, pickups, and dropoffs). For weighted matching, each sketch has a default weight which can be adjusted by users. For instance, a large weight may be given to a major street, while small weights are assigned to secondary streets. The default weights are predefined by our experts based on the hierarchical levels of sketch streets (e.g., 1.0 for highway and primary roads and 0.7 for secondary streets). For each sketch a bar chart is shown on the right, which provides visual cues to users about the historical values of the corresponding attribute. For example, users become aware of the normal traffic speed or outliers speed of a sketch, so that they can easily speculate the input value.

### 5 Sketch-Based Visual Query on Traffic Data

Through sketching on map, users select multiple geo-structures at different locations. Then related traffic data is queried based on the sketches. To support immediate response to the sketch interaction, data retrieval should be accelerated. We first develop a spatial database integrating urban structures (e.g., street geometry, POI locations, and their names), traffic information (e.g., volume and speed over time), human mobility information (e.g., taxi pickups/dropoffs and trajectories), and other information (e.g., weather condition). These data is generated from raw taxi trajectory datasets. We also extract historical weather conditions of the city.

**Data Fact:** In our prototype, we use the datasets from Hangzhou city in China, which has about 9.1 million residents as the capital of Zhejiang province. Traces of 8,120 taxis in three months, Dec. 2011 to Feb. 2012, are utilized. Each trace includes a set of raw GPS sample points and associated attributes (ID, speed, time, etc.). The large volume of the raw data sets are 46.1 GB for Dec., 42.8 GB for Jan., and 44.2 GB for Feb. The city’s road network including the geometries of street segments is retrieved from OpenStreetMap. Moreover, the historical weather data of the Hangzhou city is downloaded from an online portal (forecast.io) at each hour of these days. The weather attributes include summary (cloudy, rain, etc.), and other properties such as pressure, visibility, humidity, temperature, etc. They are saved in an independent table in the database.

#### 5.1 Accelerating Sketch Response Speed

All the sample points from taxi traces are indexed by their spatio-temporal attributes. For the time dimension, we assign each point to time windows using two-hour periods of each day (hourly or other finer resolution can be used too). It is also assigned to a specific weekday. Each GPS point is also map-matched to a specific street segment. Consequently, the raw data is stored in a database table whose records are in the form of \((p, h, d, w, S_i)\), where \(p\) is the GPS point, \(h\) is the hour, \(d\) is the day, \(w\) is the weekday, and \(S_i\) is the corresponding street segment where \(p\) resides on. Next we build a traffic data cube by aggregating these raw data records to compute and store the traffic attributes such as speed, taxi
pickups, taxi dropoffs, and taxi flow volume, on each of these street segments. Unlike existing work such as Nanocube [11], our data aggregation and caching is conducted on street segments instead of spatial grid cells (of a tree subdivision). When users specify a sketch, a spatial indexing structure based on R+ tree is used to quickly find the street segments. Then, the traffic attributes over any time periods of any day (or weekday) can be directly retrieved from the data cube. Based on the traffic data cube and spatial indexing, QuteVis well supports immediate sketch response and interactive visualization.

5.2 Retrieving Similar Data from Sketch

Given urban transport database $U$, a sketch defines a set of user-specified spatial constraints $Q$. The similarities of all data records in $U$ to $Q$ are ranked to retrieve similar ones to the sketch.

To implement the sketch query, a global similarity value $S$ is defined for each data record $r$. It is a value between 0 to 1 (or percentage 0% to 100%), which is computed according to the given $Q$ as:

$$S(r|Q) = \frac{\sum_i \omega_i d_i}{\sum_i \omega_i}$$  \hspace{1cm} (1)

Here $\omega_i$ is the weight given to the sketch $i$. $d_i$ is the similarity between the actual value, $S_i(a_m)$, and the given value, $c_i(a_m)$, of attribute $a_m$. It is computed as:

$$d_i = 1 - \frac{|c_i(a_m) - S_i(a_m)|}{\max(c_i(a_m), S_i(a_m))}$$  \hspace{1cm} (2)

where the rightmost term is the percentage difference of the two values. $d_i$ is the similarity of one spatial constraint (i.e., one sketch). In practice, users can give one constraint by selecting one path, one intersection, or one region, which includes multiple small street segments. The average of actual values over these segments defines $S_i$.

The weighted global similarities provide hints about how the traffic attributes in different times close to the given pattern. The top matched pairs results with large similarities can form recommendations to users. On the other hand, each $d_i$ is also useful for users to identify the similarity of each sketch. In our visual design, we visualize both $S$ and their constituent factors $d_i$.

Similarity Categorization: At the beginning of our design, the numeric similarity values are directly color-encoded in visualization charts. However, 5ExP who worked with us on system design are not satisfied with the complex visual representations from various similarity values. They suggest that in real scenarios, most users would like to quickly identify the distributions of similar times to their query. Meaningful and qualitative visualizations are preferred as quantitative color mapping requires more effort to read and understand. Based on their suggestion, we predefined categorical metrics of similarities as:

- 0% ≤ $S$ < 50%  \hspace{0.5cm}  \text{"Not Similar"}
- 50% ≤ $S$ < 70%  \hspace{0.5cm}  \text{"Little Similar"}
- 70% ≤ $S$ < 90%  \hspace{0.5cm}  \text{"Very Similar"}
- 90% ≤ $S$ ≤ 100%  \hspace{0.5cm}  \text{"Extremely Similar"}

The category descriptions are heuristic and users can set them up by their preference in system configuration.

6 Visual Analytics Functions

To better describe the visual representations and functions, we assume that an urban analyst, named Amy, makes three sketches on the canvas as shown in Fig. 2(1). Then she defines the slow traffic speeds (15 km/h) for these sketches. She wants to study how such patterns occur in history traffic database of Hangzhou. A set of coordinated views in QuteVis help her perform VA tasks:

Study Distribution View: First, Amy discovers the distributions of historic times on the Study Distribution view (Fig. 4). It shows three different types of distributions. The colors in the bars represent the categories of similarity. Here a set of popular color palettes are made available for Amy to choose from a menu. In each chart, Amy can click on each of the icons of these categories to make them visible or invisible. In Fig. 4, Amy finds that her input traffic situation on the three streets does not appear often as shown only as blue bars.

On the top bar chart, the x-axis represents the times (in two-hour time windows) of a day, which is 12am-2am, 2am-4am, and so on. The y-axis shows the count of data records that falls into these time windows. The Hangzhou database has 91 data records for each time window, that is, 91 days in the three months. Amy can find the distribution of similarities in each time window. In her case, those times when the three streets have a low speed 15 km/h only appear during the daytime. The afternoon rush hours (4pm-8pm) are the most frequent times for such situation happens. The morning rush hours are not comparable to the afternoon, which hints a mobility pattern of residents.

On the middle bar chart, the distribution over weekdays from Sunday to Saturday is perceived by Amy. Each weekday has 156 different data records in the database (12 time windows per day in 13 weeks). Amy finds that her input traffic situation on the three streets does not appear often on Saturday and Sunday, while Sunday has the fewest counts. Meanwhile, Friday is the weekday that the three streets suffer most from low speed.

Finally, in the bottom, Amy reads the distribution of data records on different weather conditions. For example, she can find the highest bars represents cloudy and foggy (and polluted) days.

Grid Heatmap View with Similarity Entropy: Amy then wants to find details of the matching times in different days. As shown in Fig. 2(2), a grid of dots on the right panel shows a heatmap of similarities. Each dot presents one time window in a specific day in a specific month. Its size represents the similarity value, $S$ in Eqn. 1, of this time to the query. The top ten matched times are highlighted with a green boundary so that Amy can immediately find what times are in the Top List of matching.

In Eqn. 1, $S$ is indeed computed from a set of $d_i$, each of which is the similarity of the actual value compared to the given query value on one sketch. Two dots may have the same size, but they have different $d_i$ values over different sketches. For example, one dot may have small differences on Sketch1 and Sketch2, and the other dot has a larger difference on Sketch1 but no difference on Sketch2. On the heatmap, we use the color of these dots to give users some hint of this diversity. In particular, we map the entropy, $H$, of $d_i$ values to the dot color, which is computed as:
Fig. 4: Study the distribution of query results.

\[ H = \sum_{i \in L} p(d_i) \log p(d_i) \text{ where } p(d_i) = \frac{d_i}{\sum_{i \in L} d_i}, \]  

Here \( L \) is the number of sketches (constraints). As shown in Fig. 2(2), two top matched dots are indicated by two arrows in red. Their corresponding times are 6pm at Feb. 10 and 4pm at Feb. 21, both of which have a large similarity considering Amy’s query. However, the visualization shows that the entropy \( H \) is lower in the former time than the latter time. Amy may pay special attention to study the dot for Feb. 21 to find the differences. In particular, Amy can hover the mouse over the dots to find such details. Please note users can freely change the color scheme with their preference. She can also click on any dot to add it to the Top List. The details of Top List is given in a list view on the left panel.

**Top List View:** After Amy overviews top matches and possibly selects her interesting times, she can further investigate their details in the Top List view. As shown in Fig. 2(3), this view shows a set of cards of all items in the list. She can read one card about the specific time, the date, and the weather. On the right, she can find the differences between the input values and the actual values of each active sketch. Amy can further click one card to show the traffic information on the central map canvas.

**Multi-Map View:** In Fig. 5(a), the multi-map view is shown in the left panel. Each map shows the traffic information of one specific time in Top List. The colored lines indicate the traffic information on the roads. Roads with smooth traffic flows are marked in green, roads where congestion is moderate are in orange, while those that are congested are shown in red. Amy can select different attributes to be shown on these maps. She can click one map to duplicate (enlarge) it in the central map canvas. Zooming, panning and other map operations can be performed on any of these maps, including the central canvas. All the maps are coordinated so that Amy can compare any area and street in different times. She can also give a street name or POI name, in the panel of Fig. 2(4), to quickly locate them in these maps.

7 Evaluation

7.1 Evaluation Team

We have shown that experts in group 5ExP helped us on system requirements and provided us feedback during our design and implementation steps. Their suggestions were integrated into the prototype. Here, we show evaluation after QuteVis is fully completed based on several cases for the Hangzhou datasets, and domain experts feedback after testing our system.

We asked a team of users (named as CTeam) to help us on case studies using QuteVis. This team had users who are familiar with Hangzhou to provide meaningful use cases. CTeam included one active urban geography researcher who also worked as an urban transportation planner in Hangzhou city for several years. CTeam also included two local residents who live in Hangzhou for more than 20 years.
Moreover, to further validate QuteVis system, we interviewed a group of eight domain experts (5 males and 3 females) whose ages range from 27 to 60. They included the 5ExP experts. Five of them are expert in the areas of urban planning and transportation, GIS, regional economy, and geography, while three of the five worked as an urban planner before. One of them had 25 years of experience as a transportation planner working with private firms and public agencies. The other three domain experts are the researchers and scholars with earned PhD degree in the areas of GIS, urban planning, and geography. This group of experts (named as FTeam) used QuteVis prototype after we finished implementation and provided feedback for several questions.

7.2 Case Studies

Three of use cases from CTeam are shown below. To better illustrate them, we continue to use the name Amy as the subject for description.

Traffic Speed on Three Major Roads: While introducing the system interface in Sec. 3.3, we have shown one case where Amy defines Pattern 1 involving slow traffic speeds on three major downtown streets (Fig. 2). Fig. 5(a) further shows the multi-map view, where three maps are used by Amy to compare the traffic situations at 4pm-6pm of Jan. 22, Dec. 4, and Feb. 21, respectively. The first one (Jan. 22) is enlarged in the central canvas. For example, four red arrows point to the same street, Zhonghe road in these maps. The traffic on this road at Jan. 22 is smooth shown in green. But at the other two days, this road has slow traffic speeds shown in red/yellow. Amy discovers the reason: Jan. 22 is in the holiday period of Chinese New Year in 2012, when many residents stayed at home or went back to their home towns.

In Fig. 5(b), she further zooms in to an area on the east bank of the West Lake. She sketches on the canvas map to select one specific street (highlighted in purple) she wants to study. In the Info Mode, an information visualization window pops up to show the traffic behavior of this road. Three lines on the two line charts display the varying traffic speed (red), taxi pickups (blue), and dropoffs (yellow) in a whole day, respectively, together with the weather conditions. For Dec. 4, Amy finds that the numbers of pickups/dropoffs decrease at 4pm-6pm to the lowest values during the day, when the three sketch roads have slow speeds as she speculates. This shows to her that the taxi activities on this road are potentially affected by the three major roads in downtown Hangzhou.

Traffic Speed and Taxi Activities of a Hospital Region: Fig. 6 shows Amy studying Pattern 2 which involves traffic speed of a major road and taxi activities in a residential area close to it. In Fig. 6(1), Sketch0 shows a primary street segment, Qingchun road, which connects the west downtown area to a primary street, East Loop road, of the city. Sketch1 defines a residential region which mainly includes two big hospitals of the city. Amy specifies a normal speed, 30 km/h, on Qingchun road, and a value of taxi dropoffs as 120, which indicates a lot of passengers arriving this region in a 2-hour window. Then, Amy observes the distribution charts of the query results. As shown in Fig. 6(2), this queried pattern happens mostly during the daytime from 8am to 2pm. For comparative study, Amy adjusts her query by changing the input speed on Qingchun road to a slow value at 12 km/h. Fig. 6(3) shows that this situation happens only during morning hours 8am-10am. Amy realizes that the counts of records having such a situation are much smaller (e.g., 18 at 8am-10am) than the counts of her first query in Fig. 6(2) where the counts are around 50-60 for any 2-hour periods during 8am to 2pm. It means when Qingchun road has a slow speed, taxi dropoffs can rarely have a large value. In comparison, when Qingchun road has fast speed, taxi dropoffs happen a lot in the region. By this comparison, Amy may conclude that the taxi activities in the hospitals are affected by the speed of Qingchun road. This finding confirms Amy’s suspicion at the beginning. She can further use map views to study several specific times for traffic behaviors on other surrounding locations.

Taxi Activities of a Tourism Region: Fig. 7(1) shows another task to investigate Pattern 3 involving taxi pickups and dropoffs in a tourism area. Sketch0 and Sketch1 include the same region, West Lake Scenic Area, which is one of the most famous tourist sites in China. After sketching, Amy finds that the taxi pickups and dropoffs have similar histograms from Fig. 7(2). She wants to find a situation when the number of pickups is low at 40, and the number of dropoffs is high at 100. Fig. 7(3) shows that such a pattern is frequent in the morning 8am-10am. It is meaningful as the morning time is when visitors arrive for their tour. Meanwhile, this pattern happens for all weekdays. This is also reasonable for this famous place, whose visitors come from all over China and do not show big weekday/weekend differences. Amy further study specific days of interest through other visualizations.
7.3 Domain User Study and Feedback

After forming FTeam, we conducted a preliminary study with these domain experts. Our key goal here is to justify the usability of QuteVis, with its sketch+visualization interface, and to gain knowledge about the limits and to identify future directions. Formal user study of all system functions will be our future work after deploying the software to more real world urban transportation planners and analysts.

First, we explained the system to our subjects (i.e., FTeam members) through a case study with a detailed description about the functions and interface. Then, we allowed each subject to use and explore the system on a touch screen for about 20 minutes. After the preparation, we asked them to draw two sketches on the map, and then answer the following questions for four different VA tasks (shown in black):

- **Q1: Find similar traffic times**: Which weekday has the largest similarities?
- **Q2: Identify associated attributes**: Which weather condition has the largest similarity?
- **Q3: Ranking**: What are the top 3 times and dates having your given conditions?
- **Q4: Comparison** in the top 3 times, which one has the worst traffic conditions?

These questions were designed because (1) they facilitated test use of major QuteVis supported VA tasks, (2) they required the subjects to use all QuteVis visual functions including multi-sketch and all visual charts and map views to conduct these tasks. In particular, Q1 and Q2 are linked to the distribution charts and the grid heatmap, Q3 should use the Top List for ranking, and Q4 involved the multiple map comparison.

The time used by each subject on each task and their correctness were collected to facilitate quantitative measure of their performance. All of the experts completed the tasks in 4 to 5.5 minutes with an average of 4.2 minutes. 95% of the subjects achieved correct answers compared to the ground truth. Based on these tasks, all subjects agreed that the system is useful and efficient in such urban study jobs based on their domain knowledge and experience.

The FTeam experts were further asked to fill a feedback form about the system with respect to three aspects:

- **A1**: The usefulness of our system to be used in transportation planning and traffic management studies.
- **A2**: The convenience of our sketch functions for supporting users to query and investigate times and days that have “similar” traffic patterns to user input.
- **A3**: The effectiveness of our sketching, visualization and interaction functions to show, filter, and compare the results.

All experts agreed on its potential use in the field of urban planning, transportation, business marketing and education. Some specific comments wrote: “The input of focal speed is very useful for planners because any given segment has a corresponding speed limit (can be used here) or a safe speed according to the user”. “The system reflects congestion of traffic in the city that helps the investor to plan investment locations”. “The similar traffic patterns are useful to compare two different areas, for example downtown vs suburban areas”. One expert mentioned “I want to call it a platform instead of a tool because domain users and researchers can get many hints of how to use the trajectory data along with the urban streets in a more intuitive way. In other words, the current default setting can allow the user to become familiar with the data from multiple perspectives of transportation such as speed, streets layout, urban structure, and day/time in a comparative context.” Then, they unanimously agreed that the system interface is intuitive, very friendly and easy to use. They were more excited about the tools that allow them to
do multi-sketch on the map especially by using a touchscreen. One expert wrote “The sketch function plays the major role in this system, not only because it is convenient to use given so many touching screen facilities, but also it can be useful in different contexts, from novice users (e.g. education environment) to domain expert (e.g. urban planners).” They also liked the multi-map view to compare and study the traffic situations in different time windows.

Meanwhile, the experts pointed out some drawbacks and gave us valuable suggestions to improve the system, such as “it will be great to have other transportation data such as public bus and subway data”, which will be one of our future works. One expert suggested to add an interactive tutorial to the system to teach first time users. Another suggestion was to add more descriptions to the icons in the top panel. They also suggested to add a view to study the real-time traffic situation on the city if data is available. They would like to see this tool to be combined with traffic predictions, which is our future work to combine the system with commercial map service APIs.

8 Discussion and Future Work

In this paper, we have developed a visual analytics approach, QuteVis, for urban planners, traffic analyzers, and other practitioners to visually query a transport database with their speculated patterns. They can sketch on the map for flexible inputs and visually study the times and days such patterns may occur.

This system has its limitations to be improved in the future work. First, the correlation and cause-effect relationship among different locations cannot be directly discovered over the interface. We will introduce data mining tools to visually detect such information and provide profound interaction with users. Second, statistical functions may be included for users to analyze the query results, such as statistical significance testing and A/B testing. Third, more advanced sketch operations should be designed to support the above two directions.

References


