“THEME ARTICLE”, “FEATURE ARTICLE”, or “COLUMN” goes here: The theme topic or column/department name goes after the colon.

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

Jing Yang
UNC Charlotte

Xinyue Ye
New Jersey Institute of Technology

Wei Chen
Zhejiang University

QuteVis uses multi-sketch query and visualization to discover specific times and days in history with specified joint traffic patterns at different city locations. Users can use touch input devices to define, edit, and modify multiple sketches on a city map. A set of visualizations and interactions are provided to help users browse and compare retrieved traffic situations and discover potential influential factors. QuteVis is built upon a transport database that integrates heterogeneous data sources with an optimized spatial indexing and weighted similarity computation. An evaluation with real-world data and domain experts demonstrates that QuteVis is useful in urban transportation applications in modern cities.

A large number of transportation datasets are collected and utilized in urban computing applications. Efficient and user-friendly tools for visually querying and examining these datasets are important in modern transportation studies and practices. Most existing visual analytics (VA) systems support “*where+when→what*” queries to retrieve and visually examine traffic attributes such as speed and volume at given time periods. In these queries, users define spatial constraints by drawing a spatial region, brushing a road, or giving a road name. They also define temporal conditions by selecting dates, times, or weekdays. However,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

there are no existing VA systems devoted to the "*where+what→when+how*" queries. One example of such a query is to find specific time periods when both a highway bridge and a city street have traffic jams. For constructing such a query, users need to specify traffic situations, the what, at ***multiple and spatially diverse*** geo-locations, the where. This type of queries is demanded by domain experts to investigate when and how patterns of interest happen in locations of interest. There is a dire need for a VA system that allows users to interactively construct this type of query, retrieve query results from transport databases, and visually examine the query results.

In this paper, we present a visual analytics system named QuteVis. It helps users interactively (q)uery (u)rban (t)ransport databas(e) (Qute) to address this need.

First, QuteVis allows users to specify multiple conditions at separate locations on the map with multi-sketch query. Users can directly draw sketches on a city map and specify desired traffic attributes (e.g., traffic speed) on them as query conditions. A reliable sketch recognition algorithm is proposed and implemented in QuteVis so that freestyle drawing is automatically matched to street geometry. The system further allows users to edit sketches with easy interactions.

Second, QuteVis allows users to sketch on geo-structures of different granularities. Users may draw a long path that covers a sequence of streets. They may also select a small street portion at a critical intersection. Typical street network geometry from GIS databases cannot support such operations well. We propose a street segment optimization algorithm to improve sketch accuracy and implement it in QuteVis.

Third, to achieve good interactivity, QuteVis is built upon a new database to store and retrieve traffic information efficiently. Interactivity is very important for effective visual queries. Query response time should be short enough to enable interactive visualization. The database accelerates the data retrieval speed using geospatial indexing and a specific data cube which plays an indispensable role to enable smooth user interactions.

Fourth, QuteVis computes a global "weighted similarity" among multiple sketches, which is used to quickly find matching results in massive traffic data records at different time periods. A weight is assigned to each spatial sketch which can be adjusted by users. The similarities on multiple sketches are drawn as visual cues for users to investigate their preferred results.

Fifth, QuteVis presents a set of visualizations within a sketch+visualization interface. They provide users an overview of the joint traffic patterns. Users can further investigate top matched time periods and select any interesting time periods for drill-down analysis. Moreover, a multi-map comparison view allows users to compare traffic situations at different locations.

In summary, the main contributions of this paper include:

- QuteVis supports "*where+what→when+how*" queries, which is an important investigation model to study urban transportation data. To the best of our knowledge, QuteVis is the first VA system devoted to such a model.
- QuteVis allows users to specify joint traffic situations in multiple separate locations, which have not been supported in existing transportation application tools.
- Reliable sketch recognition algorithms are proposed to support different types of sketches on streets, paths, and regions.
- QuteVis integrates a sketch+visualization interface and heterogeneous data management for domain users to conduct interactive visual queries and drill-down studies.

RELATED WORK

Transportation Data Visualization

There has been extensive research on visual analytics of transportation data[2]. Typically, map-based displays and information visualization techniques are combined in visualizing the spatio-temporal data. Existing approaches can be divided into three categories according to the types of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

transportation data they visualize: (1) VA systems studying human/vehicle trajectory data: Ferreira *et al*.[5] allowed users to visually query taxi trips with spatiotemporal constraints. TrajectoryLenses[9] used lenses to support visual, set-based filter expressions to select trajectories. AL-Dohuki *et al*.[1] transformed taxi trajectory into texts so that users could query them using a text search engine. (2) VA methods studying traffic flow data: Sheepens *et al*.[14] visualized the directions of traffic flows using a particle system on top of a density map. Our system does not work on trajectories or traffic flows directly. Instead, it allows users to query traffic information on city structures. (3) VA methods investigating aggregated traffic data on city structures: Pu *et al*.[13] visually monitored and analyzed complex traffic situations in big cities. Wang *et al*.[19] presented a visual reasoning approach for data-driven transport assessment on urban roads. Most of these approaches were designed to access traffic data based on *when+where* queries. In contrast, our approach is the first to support *what+where* queries. It helps users discover times and days when multiple streets, paths, and regions have user-specified traffic values.

Sketch-Based Visual Analytics

Sketch-based interfaces allow users to interact with a computer through sketching. Sketches in QuteVis are defined on urban maps, while many existing systems supported sketching over other types of visual elements, such as charts, glyphs, and graphics. SketchVis[4] leveraged 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 a presenter to record a sequence of charts along with example icons before a presentation, and to invoke them with simple sketch gestures in real-time. SketchSliders[17] provided a mobile sketching interface for creating sliders to interact with multi-dimensional datasets on a wall display. These methods may be adapted in our system to allow users to sketch over traffic attributes.

Sketch-based queries were utilized in the GIS domain, where users sketch over maps to select geographical objects. SpaceSketch[7] provided interactive tools of sketching paths and regions for spatiotemporal data visualization. Sketch-based queries were supported to find locomotion patterns of human trajectories[16]. Blaser *et al*.[3] highlighted a set of interaction methods and sketch interpretation algorithms that were necessary for pen-based querying of geographic information systems. Users of nuSketch[6] sketched on maps when reasoning about a hostile battlespace. Spatial data characteristics were visually studied along a line drawn across a map[18]. Malik *et al*.[12] allowed users to sketch time periods of interest. HotSketch[8] allowed police officers to sketch a path on a map by specifying control points of a polyline to query point-based crime data. Our method is different from these approaches in that (1) our sketch recognition algorithm identifies and matches sketch inputs in the units of optimized street segments, (2) user's free drawing is automatically matched to street segments to enhance sketch accuracy and smoothness, (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, and (4) our system allows users to perform easy editing of the sketch results by adding and removing street segments from existing selections, which has not been implemented in existing GIS sketching tools.

OVERVIEW

Goal and Tasks

The objective of QuteVis was set by interviewing a group of five experts (named as 5ExP) working on urban planning and transportation analysis. They commented that most existing work is focused on retrieving and visualizing traffic data based on given spatiotemporal conditions. Then they proposed the following example tasks that were 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 when both of them have traffic jams? What are the temporal distributions of the results on different periods (e.g., morning, afternoon, night)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

and different days (e.g., Mon, Tue, etc.)? What are the weather conditions of the results? What are the traffic situations in a nearby residential district in those days and times?

These questions inspired us to develop QuteVis to help users “see” and examine the traffic patterns in a large volume of historical traffic data, by giving their speculated values in separate locations 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 was 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 were identified for a VA system:

- T1: **Flexible multiple selections and editing**: Users should be able to quickly locate and select their preferred urban structures (including streets, paths or regions) within an urban geographical context. They should be able to easily set up query conditions such as 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 traffic values at specific locations for users to set up query conditions.
- T2: **Fast response and interactive exploration**: The system should provide immediate response for queries, and the results should be visualized quickly for users to iteratively explore the large data.
- T3: **Easy perception and understanding**: Visualization of query results should be easy to understand. Visualization should not be focused 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 situations in multiple times/days and at different locations should be supported. Multiple locations diversely distributed in a city should be compared together.

System Design

QuteVis is an interactive VA system integrating user sketching and visualization, which is supported by efficient data management. Its framework and major functional modules are illustrated in Fig. 1.

Multi-sketching on a map plays a key role to facilitate easy user selection and editing (for T1). Users can define their speculated patterns by (1) selecting *multiple, spatially diverse* locations on a city map. The selected objects (i.e., sketches) can be streets, paths, or arbitrarily shaped regions; (2) specifying query values at selected locations about transport attributes (e.g., traffic speed), with visual hints of their historical values. Street geometries are optimized for accurate and flexible sketching. Meanwhile, sketch processing is implemented to enable prompt responses (for T2).

The interactive sketching is seamlessly integrated into a visual interface for visual analysis (for T2 and T3). Traffic data records of different times at different days are visualized to study the distributions of top-matching results at different periods (e.g., morning, afternoon), different weekdays, and different weather conditions. According to a suggestion from domain users, the matched traffic data records are categorized into *not similar*, *little similar*, *very similar,* and *extremely similar* by their similarities to user-specified traffic patterns (for T3). The global similarity is jointly computed from individual similarities of multiple sketches, where users can define their weights in the computation. Moreover, QuteVis provides a multi-map view as an effective tool for comparative studies (for T4). Users can examine transportation features on any location of a city.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

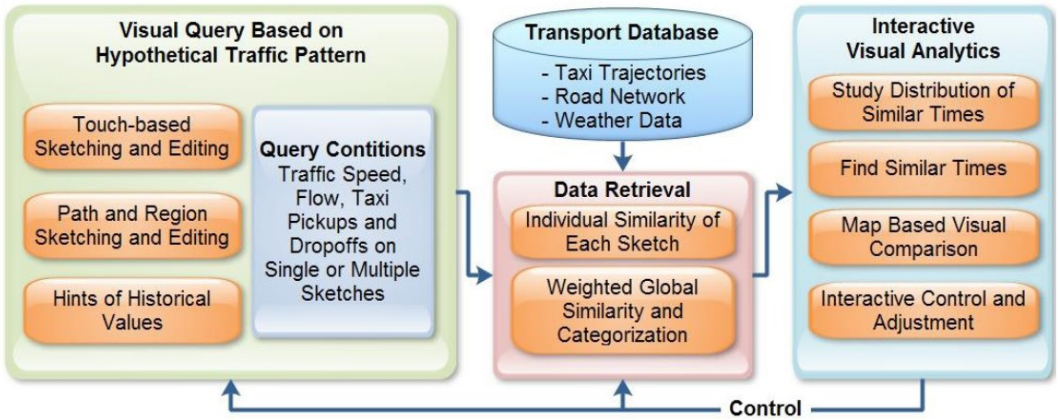


Fig. 1: QuteVis System Framework. QuteVis is an interactive visual analytics system integrating user sketching and visualization. It has efficient data integration and management.

QuteVis Interface

The interface has a map-centric design. As shown in Fig. 2, it includes the following major views:

- A canvas over a city map (see Fig. 2(a)) to facilitate context-aware operations in the city. In the **Sketch Mode**, it supports the selection of multiple streets, regions, or paths. Three *sketches* are drawn in Fig. 2(a). In the following text, we also refer to a *sketch* as a selection. Colors are used to encode different sketches on the canvas to help users differentiate between selections. Users can also edit the sketches interactively. In the **Info Mode**, the map view visualizes traffic data, POIs, and other urban information. Users can select any region/street to study its traffic behavior and weather in a specific time. The map view supports different map styles and smooth zooming and panning operations.

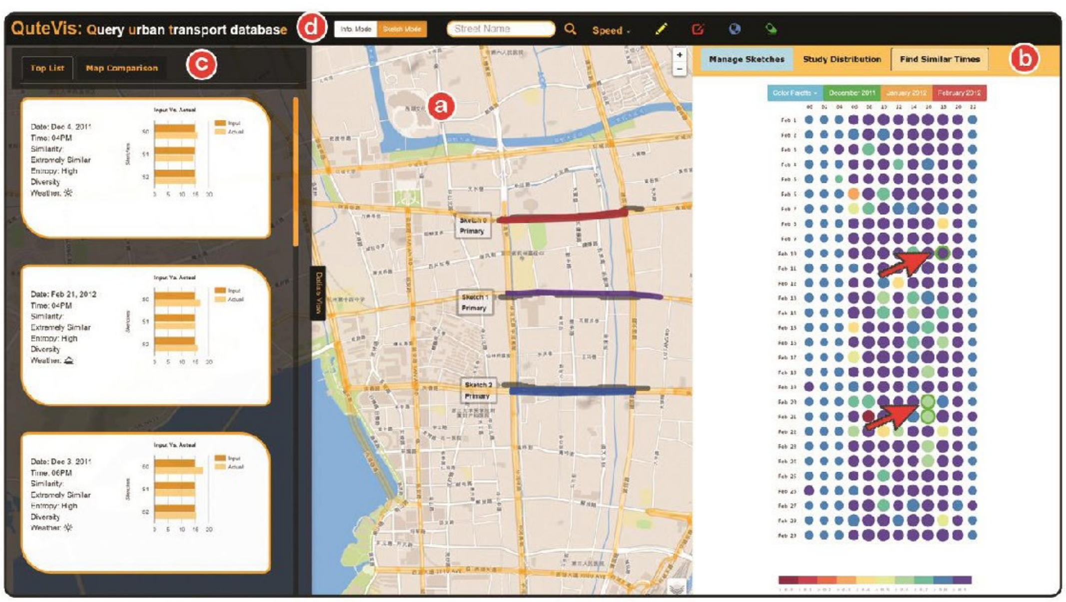


Fig. 2: QuteVis interface consisting of (a) a map canvas, (b) a control panel, (c) a detail study panel, and (d) a top panel. It supports visual queries and analytics of historical days and times that have similar traffic patterns as speculated by users.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

- A control panel with multiple tabs (Fig. 2(b)). The tabs include: (1) **Manage Sketches**: Under this tab, users can define query conditions on multiple locations and control whether the sketches are active in the control panel. (2) **Study Distribution**: Under this tab, users can investigate the distributions of the matched patterns over weekdays, hours, and different weather conditions from the control panel. (3) **Find Similar Times**: Under this tab, the control panel shows a grid heatmap view for users to find and study the similarities of available historical times to the query. Those candidate times with the largest similarities are highlighted and put into a Top List, while users can select any arbitrary times of interest and add them to the Top List.
- A detail study panel (Fig. 2(c)) to investigate traffic information at selected times in the Top List. The user input value and the actual traffic value on each sketch, the similarity categories, and the weather condition are visualized. Furthermore, **Map Comparison** allows users to examine and compare the traffic behaviors at different times on a multi-map view. This view of multiple small maps 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.
- A top panel (Fig. 2(d)) 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.

MULTI-SKETCH ON GEO-STRUCTURES

QuteVis allows users to define and edit multiple sketches on a map easily. Its sketch recognition algorithm, unlike general sketching methods, is built upon geographical or urban structures, including street segments, paths, and regions.

Sketch Recognition and Editing Algorithms

Users can draw on the map canvas with a mouse, a touch pen or by hand drawing (if supported by devices). In Fig. 2(a), three sketches represent three streets categorized as primary roads in downtown Hangzhou city in China.

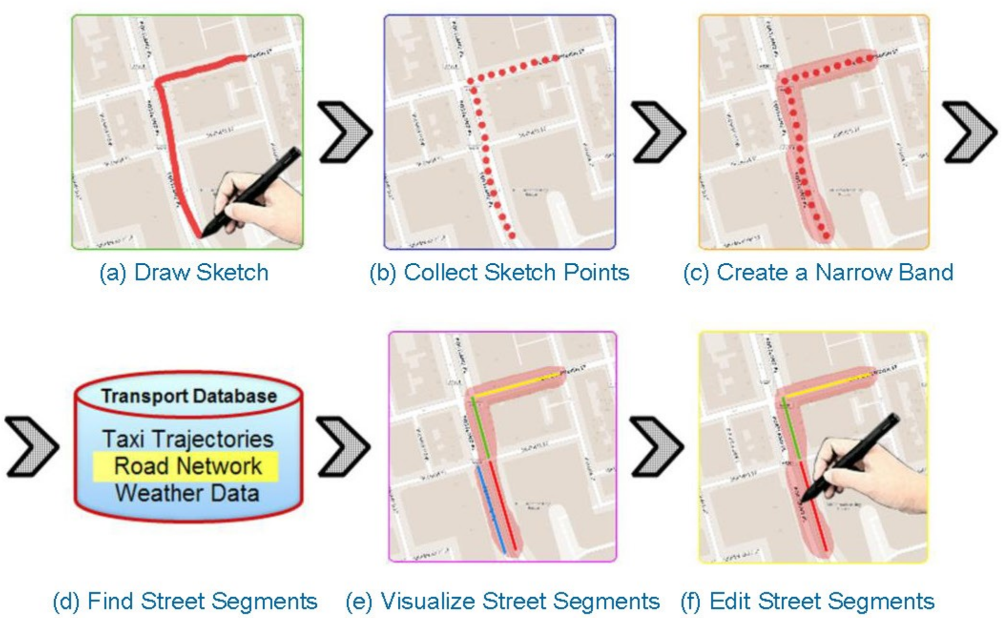


Fig. 3: Illustration of the process of path selection and editing.

A **street sketch recognition algorithm** is used to map user inputs to corresponding street segments. It consists of the following steps:

1. When a user draws on the map (Fig. 3(a)), a sequence of sketch points are collected (Fig. 3(b)). These points are processed to remove obvious errors and noise.
2. A polyline is formed by connecting consecutive input points. Then, a narrow band enclosing this polyline is generated (Fig. 3(c)). This narrow band mimics the stroke of the sketch. The width of the stroke is defined adaptively according to the zooming level of the map, which plays a vital role in determining the accuracy of the sketch selection. In the implementation, this stroke band is geometrically represented as a bounding polygon of the polyline.
3. The stroke bounding polygon is used to generate a region-based spatial query on the road network (Fig. 3(d)). The query retrieves all street segments inside the polygon. 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. This spatial query of street segments is the key for accurate sketch recognition since it can automatically match the free drawing to the street segment geometry.
4. The input stroke may collect unintended street segments, such as the blue segment in Fig. 3(e). Sometimes a stroke may not cover a segment needed. To address this problem, editing operations are required for users to adjust any sketch by adding and removing related street segments. As shown in Fig. 3(f), users can click the blue segment to remove it.

The algorithm allows users to choose a single street or a long path between two locations, which is flexible for traffic studies. 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 turn this polyline to a closed curve by connecting the start and end points. Thus, a polygon is formed to query the road network and all the street segments inside it are selected. Similar editing operations are available to refine the sketch results.

Improving Sketching with Street Segment Optimization

In the sketch recognition algorithms, the scale of street segments defines the sketch selection resolution. Street segments are also the basic units storing traffic data records for user selection. Therefore, they should not only represent the city road network but also (1) facilitate accurate sketch selection and editing and (2) enable high-resolution traffic data storage and management. However, the “raw” road segments retrieved directly from an existing geo-database (e.g., from OpenStreetMap) cannot be directly used in the sketch operations. It is a significant challenge that needs to be addressed.

The raw street segments have many problems. First, a raw street segment may be very long and pass multiple road intersections or have a complex geometric shape. Thus, a sketch for selecting a short street part may retrieve a long path or a path with a complex shape. Second, many raw street segments have numeric errors. These problems make it hard for users to interactively select desired streets or paths on the map, leading to frustrating sketching experience.

Our goal is to provide satisfying sketching accuracy and experience by regenerating optimized street segments. Our algorithm has two stages. First, we define new street segments to make sure that their endpoints are 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. In computer graphics, T-junction detection algorithms perform similar operations on images[20], while our intersection detection algorithm is applied directly on raw street geometries. Second, in the new set of street segments, we divide a long segment into multiple shorter ones. The purpose is to make each street segment short enough so that people can sketch to select them with high accuracy. This is very important for ensuring accurate object selection in visual studies. In our implementation, a total number of 14,639 street segments are generated for Hangzhou, while the raw road network from OpenStreetMap has 9,764 segments.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

Managing Multiple Sketches

For each sketch, users need to give an input value of a traffic attribute (e.g., speed, volume, the numbers of taxi pickups and drop-offs). For weighted matching in the traffic database, 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 streets (e.g., 1.0 for highway and primary roads and 0.7 for secondary streets). For each sketch on the map, 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 will gain knowledge about the normal and abnormal traffic speeds on a street, so that they can easily speculate an input value to study the street.

SKETCH-BASED VISUAL QUERY ON TRAFFIC DATA

A spatial database is developed which integrates 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/drop-offs and trajectories), and weather conditions. The raw data is generated from taxi trajectory datasets and external data sources.

Data Fact: In our prototype, we use a taxi trajectory dataset of Hangzhou city including taxi traces of 8,120 taxis in three months, Dec. 2011 to Feb. 2012. The volume of the data is 46.1 GB for Dec., 42.8 GB for Jan., and 44.2 GB for Feb. The city's road network geometry is retrieved from OpenStreetMap. Moreover, the historical weather data is downloaded from forecast.io.

Accelerating Sketch Response

All the GPS points from taxi traces are indexed by their spatiotemporal attributes. We assign each point to a weekday, a day, and a two-hour period in that day. Hourly or other finer resolutions can also be used. Each GPS point is map-matched to a specific street segment. Consequently, the raw data is stored in a data 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 on each street segment. The data cube stores the traffic attributes such as speed, taxi pickups, taxi drop-offs, and taxi flow volume. Unlike existing work such as Nanocube,[11] our data aggregation and caching are conducted on street segments instead of spatial grid cells. A spatial indexing structure based on the R+ tree is used to find the street segments of a sketch quickly. Then, the traffic attributes over any time periods of any day (or weekday) can be directly retrieved from the data cube. Using the traffic data cube and spatial indexing, QuteVis well supports immediate sketch response and interactive visualization.

Retrieving Similar Data from Sketch

Given an urban transport database U and a set of user-specified spatial constraints Q from user sketching, the **similarities** of all data records in U to Q are ranked to retrieve similar records. A global similarity value S is defined for each data record r. It is a value between 0 and 1, which is computed according to the given Q as:

$$S(r|Q) = \frac{\sum_{i \in L} \omega_i d_i}{\sum_{i \in L} \omega_i} \quad (1)$$

Here ω_i is the weight assigned to the sketch i. d_i is the similarity between the actual value in U, $S_i(a_m)$, and the user input value, $c_i(a_m)$, of attribute a_m. It is computed as:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

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

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

The weighted global similarities indicate how close the historical traffic attributes at different times are to the given pattern. The top matched results with large similarities will be recommended to users. d_i is also useful for users to identify the similarity of each spatial sketch. In our visual design, we visualize both S and their constituent factors d_i.

Similarity Categorization: 5ExP suggested that in real scenarios, most users would like to learn the temporal distributions of similar records to their queries efficiently. In practices, ***meaningful and qualitative*** visualizations are usually preferred than quantitative color mapping which requires more effort to read and understand. Therefore, we predefine categorical metrics of similarities as:

- *0% ≤ S < 50% → "Not Similar"*
- *50% ≤ S < 70% → "Little Similar"*
- *70% ≤ S < 90% → "Very Similar"*
- *90% ≤ S ≤ 100% → "Extremely Similar"*

The category descriptions are heuristic. Users can set up them in system configuration.

VISUAL ANALYTICS FUNCTIONS

A set of coordinated views in QuteVis help users perform VA tasks with the data investigation model of "*where+what→when+how*" . We introduce these views with the following scenario. A domain user, Amy, makes three sketches on the canvas, *the where*, as shown in Fig. 2(a). Then she defines slow traffic speeds (15 km/h), *the what*, for these sketches. She wants to study how such patterns occur in the historical traffic database of Hangzhou, the *when+how*.

Study Distribution View: First, Amy examines the temporal distributions of similar records on the Study Distribution view (Fig. 4). It shows distributions of records with different similarities over time. The colors in the bars represent the categories of similarity. Here a set of popular color palettes are made available for users to choose from a menu. Users can click on the icon of a category to make records in this category visible or invisible. In Fig. 4, Amy examines distributions of *extremely similar* records shown as blue bars.

On the top bar chart, the x-axis represents the times (in two-hour time windows) of a day. The y-axis shows the count of records. Amy can find the counts of extremely similar records in each time window. Here, the afternoon rush hours (4pm-8pm) are the most frequent times the given traffic situation happens. The morning rush hours are not comparable to the afternoon, which indicates a mobility pattern of residents.

On the middle bar chart, the distribution over weekdays is perceived. Amy finds that her input traffic situations on the three streets do not often appear on Saturday and Sunday, while Sunday has the fewest count. Meanwhile, Friday is the weekday that the three streets suffer most from low speed.

Finally, in the bottom, Amy examines the distribution of data records on different weather conditions. She finds that the high bars represent cloudy and foggy (and polluted) days.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

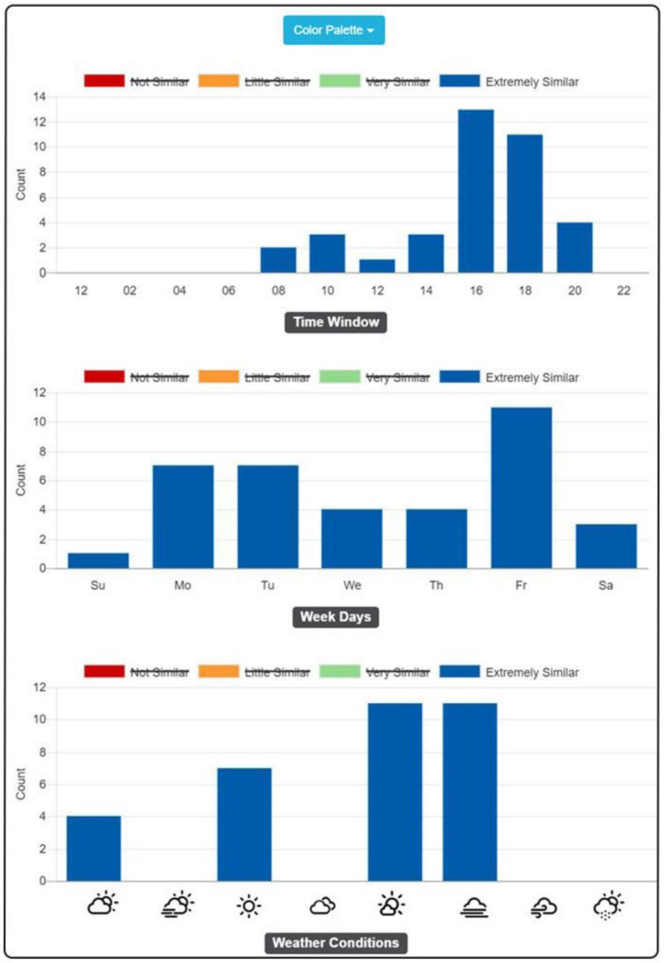


Fig. 4: Study the distribution of query results over daytimes, weekdays, and weather conditions.

Grid Heatmap View with Similarity Entropy: Amy then wants to learn details of the matching times in different days. As shown in Fig. 2(b), a grid of dots on the right panel shows a heatmap of similarities. Each dot presents a time window in a specific day of 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 matches.

In Eqn. 1, S is computed from a set of d_i, each of which is the similarity of the actual historical 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 in Sketch1 and Sketch2, and the other dot has a larger difference in Sketch1 but no difference in Sketch2. On the heatmap, the color of these dots reflects such diversity. In particular, we map the entropy, H, of d_i values to the dot color, which is computed as:

$$H = \sum_{i \in L} p(d_i) \log p(d_i) \quad where \quad p(d_i) = \frac{d_i}{\sum_{i \in L}(d_i)} \tag{3}$$

Here L is the number of sketches. In Fig. 2(b), two top matched dots are indicated by two arrows in red. Their corresponding times are 6 pm at Feb. 10 and 4 pm at Feb. 21, both of which have a large similarity. However, the visualization shows that the entropy H is lower in the former time than the latter time. Amy pays special attention to the dot for Feb. 21 to find the difference by hovering the mouse on the dot. She also clicks on the dot to add it to the Top List.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

Top List View: After Amy examines top matches and selects interesting times, she further investigates their details in the Top List view. As shown in Fig. 2(c), this view shows a set of cards of all items in the list. She reads one card about the specific time, the date, and the weather. On the right, she finds the differences between the user input values and the actual historical values of each active sketch. Amy further clicks 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 small map shows the traffic information of one specific time in the Top List. The colored lines indicate the traffic information on the roads. Roads with smooth traffic flows are marked in green, while the congested roads are shown in red. Amy clicks on a map to duplicate and enlarge this view in the central map canvas. Zooming, panning and other map operations can be performed on any of these maps. All the maps are coordinated so that Amy can compare any area and street at different times. She can also give a street name or POI name, in the panel of Fig. 2(d), to quickly locate them in these maps.

EVALUATION

Evaluation Team

After QuteVis was fully implemented, we invited a team (named CTeam) of 23 users (15 males and 8 females) to help us with case studies. CTeam had users who were familiar with Hangzhou city in China to provide meaningful use cases. Among these users, there was an active urban geography researcher who worked as an urban transportation planner in Hangzhou city for several years and two local residents who lived in Hangzhou for more than 20 years.

Moreover, to further validate the QuteVis system, we interviewed a group (named FTeam) of eight domain experts (5 males and 3 females) whose ages ranged from 27 to 60. Five of them were experts 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 were the researchers and scholars with earned Ph.D. degree in the areas of GIS, urban planning, and geography. FTeam used QuteVis and provided subjective feedback in an interview.

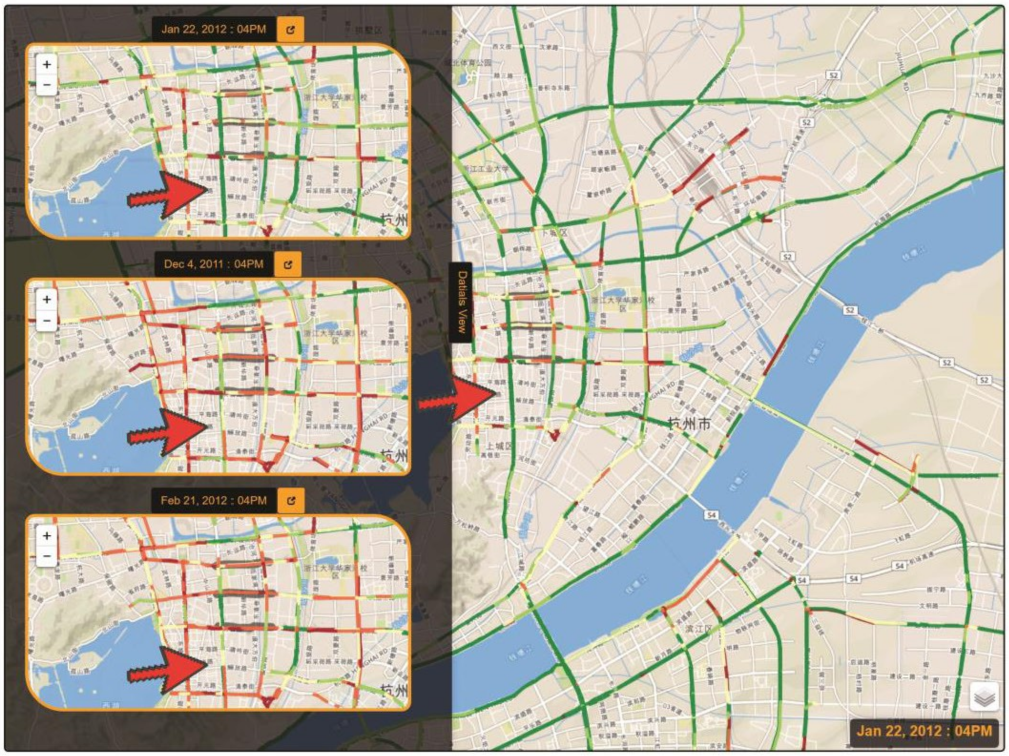
Case Studies

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

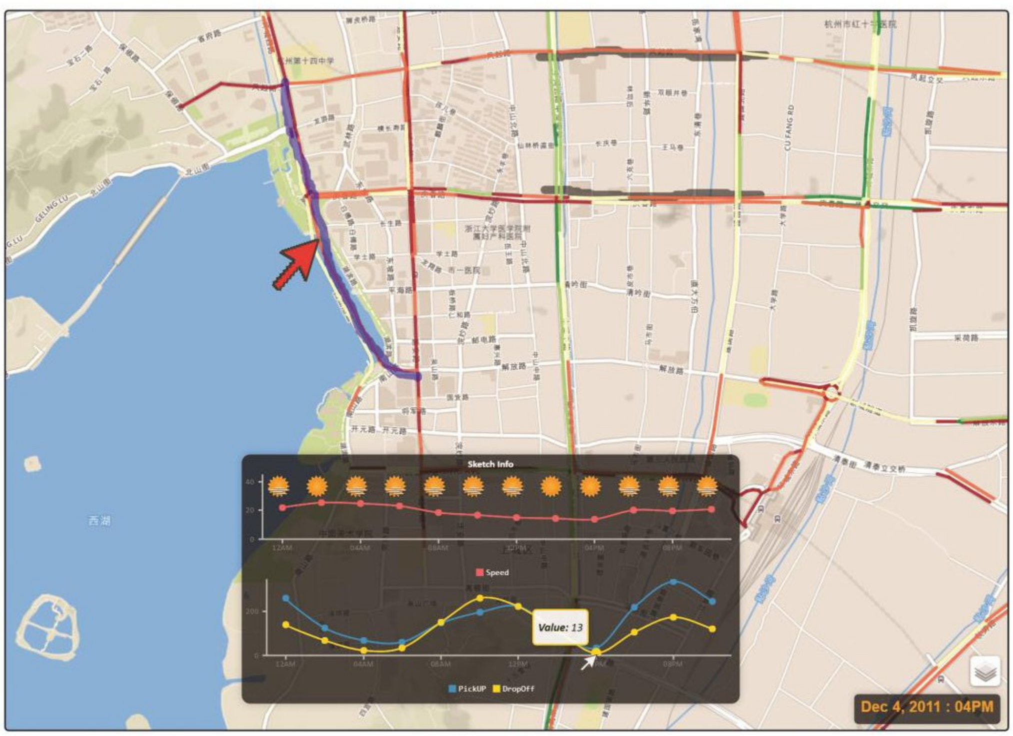
Traffic Speed on Three Major Roads: Amy defines slow traffic speeds, *the what*, on three major downtown streets, *the where* (Fig. 2). Fig. 5(a) further shows the multi-map view, where three maps are used to compare the traffic situations, *the how*, at 4pm-6pm of Jan. 22, Dec. 4, and Feb. 21, *the when*, 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 was smooth and shown in green. But at the other two days, this road had slow traffic speeds shown in red/yellow. Amy can discover the reason: Jan. 22 was in the holiday period of Chinese New Year in 2012 when many residents stayed at home or went back to their hometowns.

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). In the Info Mode, an information visualization window is popped up to show the traffic behavior of this road. Three lines on the two lines charts display the varying traffic speed (red), taxi pickups (blue), and drop-offs (yellow) in a whole day, respectively, together with the weather conditions. For Dec. 4, the numbers of pickups/drop-offs decreased at 4pm-6pm to the lowest values during the day, when the three sketch roads had slow speeds. The taxi activities on this road were potentially affected by the three major roads in downtown Hangzhou.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications



(a)



(b)

Fig. 5: Investigate traffic behaviors at different times: (a) the multi-map view that helps users to compare and study different traffic situations; (b) a pop-up window to display the traffic behavior of the specified road.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

Fig. 6: Study traffic pattern involving the speed of a major road and taxi activities in a hospital region. (a) User specified two sketches; (b) The queried pattern happened mostly from 8am to 2pm; (c) After changing query values, the pattern happened only during morning hours 8am-10am.

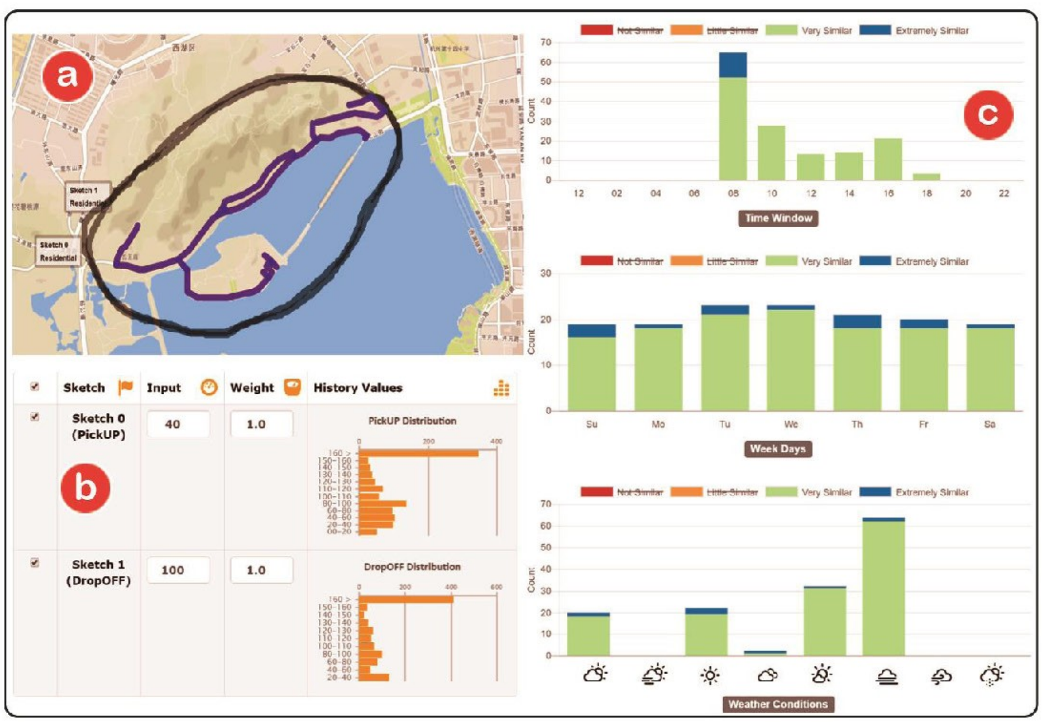


Fig. 7: Find a traffic pattern involving taxi activities in a tourism area. (a) User specified two sketches for taxi pickups and drop-offs in a tourism area; (b) Taxi pickups and drop-offs had similar histograms; (c) A pattern was frequent in the morning 8am-10am and happened for all weekdays.

Traffic Speed and Taxi Activities of a Hospital Region: Fig. 6 shows Amy studying a major road and taxi activities in a residential area close to it. In Fig. 6(a), Sketch0 shows a primary street segment, Qingchun road, which connects the west downtown area to a primary street, East Loop road. Sketch1 defines a residential region which mainly includes two big hospitals in the city. Amy specifies a normal speed, 30 km/h, on Qingchun road, and a value of taxi drop-offs as 120, which indicates a lot of passengers arriving in this region. Then, Amy observes the distribu-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

tion charts of the query results. As shown in Fig. 6(b), this queried pattern happened mostly during the daytimes from 8 am to 2 pm. 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(c) shows that this situation happened only during morning hours 8am-10am. The counts of records having such a situation were much smaller (e.g., 18 at 8am-10am) than the counts of her first query in Fig. 6(b), at around 50-60 for any 2-hour periods during 8 am to 2 pm. It means when Qingchun road had a slow speed, taxi drop-offs could rarely have a large value. In comparison, when Qingchun road had fast speed, taxi drop-offs happened a lot in the region. By this comparison, Amy may conclude those taxi activities in the hospitals were affected by Qingchun road. She can further use map views to study several specific times for traffic behaviors in other surrounding locations.

Taxi Activities of a Tourism Region: Fig. 7(a) shows another task to investigate taxi pickups and drop-offs in a tourism area. Sketch0 and Sketch1 are in 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 drop-offs had similar histograms from Fig. 7(b). She wants to find a situation when the number of pickups was low at 40, and the number of drop-offs was high at 100. Fig. 7(c) shows that such a pattern was frequent in the morning 8am-10am. Meanwhile, this pattern happened for all weekdays as for this famous place, visitors came from all over China without big weekday/weekend differences.

Domain User Study and Feedback

We conducted a preliminary study with the domain experts in FTeam. Our key goal was to justify the usability of QuteVis, with its sketch+visualization interface, and to gain knowledge about the limits and to identify future directions.

First, we explained the system to our subjects (i.e., FTeam members) with a detailed description of 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:

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

These questions were designed because (1) they included major VA tasks supported by QuteVis, (2) they required the subjects to use all QuteVis visual functions including multi-sketch and all visual charts and map views. In particular, Q1 and Q2 were linked to the distribution charts and the grid heatmap, Q3 required the use of the Top List for ranking, and Q4 involved the multiple map comparison. Based on these tasks, all subjects gained knowledge about QuteVis and provided their evaluation.

After these tasks, we conducted a subjective feedback interview with the FTeam experts. They were asked to fill a survey form about the system with respect to the following three aspects:

- **A1:** The usefulness of QuteVis system in transportation planning and traffic management studies.
- **A2:** The convenience of the 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.

Based on their answers in the survey, all experts agreed that the system was useful and efficient and mentioned its potential use in the field of urban planning, transportation, business marketing, and education. Some specific comments were: "The input of focal speed is very useful for planners because any given segment has a corresponding speed limit or a safe speed according to the user", "The system reflects congestion of traffic in the city that helps the investor to plan

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

investment locations", "The similar traffic patterns are useful to compare two different areas, for example, downtown vs. suburban areas". One expert mentioned, "The current 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." They were more excited about the tools that allow them to do multi-sketch on the map especially by using a touchscreen. One 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 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 adding statistical functions and other transportation data for bus and subway. One expert suggested adding an interactive tutorial to the system to teach first-time users. Another suggestion was adding more descriptions to the icons in the top panel. They also suggested adding a view to study the real-time traffic situation in the city. They would like to see this tool to be combined with traffic predictions, which is our future work to integrate the system with commercial map service APIs.

CONCLUSION AND FUTURE WORK

We have developed a visual analytics approach for urban planners, traffic analyzers, and other practitioners to visually query a transport database with their speculated traffic 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 future work. First, the correlation and the cause-effect relationship between different locations cannot be directly discovered over the interface. We will introduce data mining tools to detect such information and visually provide the information to users with profound interactions. Second, statistical functions should be included for users to analyze the query results, such as statistical significance testing and A/B testing. Third, more advanced sketch operations need to be designed to support the above two directions.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive comments. This work was supported in part by the U.S. NSF under Grant 1535031, Grant 1535081, and Grant 1739491. The work of W. Chen was supported by NSFC under Grant U1609217 and Grant 61772456.

REFERENCES

1. S. Al-Dohuki, Y. Wu, F. Kamw, J. Yang, X. Li, Y. Zhao, X. Ye, W. Chen, C. Ma, and F. Wang. Semantictraj: A new approach to interacting with massive taxi trajectories. IEEE Transactions on Visualization and Computer Graphics, 23(1):11–20, Jan 2017.
2. G. Andrienko, N. Andrienko, W. Chen, R. Maciejewski, and Y. Zhao. Visual analytics of movement and transportation: State of the art and further research directions. IEEE Transactions on Intelligent Transportation System, 18(8), 2232 - 2249, Aug 2017.
3. A. D. Blaser and M. J. Egenhofer. A visual tool for querying geographic databases. In Proceedings of the Working Conference on Advanced Visual Interfaces, 211–216, 2000.
4. J. Browne, B. Lee, S. Carpendale, N. Riche, and T. Sherwood. Data analysis on interactive whiteboards through sketch-based interaction. In Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, 154–157, 2011.

5. N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban data: A study of New York City taxi trips. IEEE Transactions on Visualization and Computer Graphics, 19(12):2149–2158, Dec. 2013.
6. K. D. Forbus, J. Usher, and V. Chapman. Qualitative spatial reasoning about sketch maps. AI Mag., 25(3):61–72, Sept. 2004.
7. A. Godwin and J. Stasko. Drawing data on maps: Sketch-based spatiotemporal visualization. In Poster in IEEE Visualization, 2015.
8. A. Godwin and J. Stasko. Hotsketch: Drawing police patrol routes among spatiotemporal crime hotspots. Proceedings of the International Conf. on Systems Sciences, pp. 1372–1380, 2017.
9. R. Krueger, D. Thom, M. Woerner, H. Bosch, and T. Ertl. TrajectoryLenses - A Set-based Filtering and Exploration Technique for Long-term Trajectory Data. Computer Graphics Forum, 2013.
10. B. Lee, R. H. Kazi, and G. Smith. Sketchstory: Telling more engaging stories with data through freeform sketching. IEEE Transactions on Visualization and Computer Graphics, 19(12):2416–2425, Dec 2013.
11. L. Lins, J. T. Klosowski, and C. Scheidegger. Nanocubes for real-time exploration of spatiotemporal datasets. IEEE Transactions on Visualization and Computer Graphics, 19(12):2456–2465, Dec. 2013.
12. A. Malik, R. Maciejewski, N. Elmqvist, Y. Jang, D. S. Ebert, and W. Huang. A correlative analysis process in a visual analytics environment. In 2012 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 33–42, Oct 2012.
13. J. Pu, S. Liu, Y. Ding, H. Qu, and L. Ni. T-watcher: A new visual analytic system for effective traffic surveillance. In Proceedings of Mobile Data Management, 127–136, 2013.
14. R. Scheepens, C. Hurter, H. V. D. Wetering, and J. J. V. Wijk. Visualization, selection, and analysis of traffic flows. IEEE Transactions on Visualization and Computer Graphics, 22(1):379–388, Jan 2016.
15. D. Schroeder and D. F. Keefe. Visualization-by-sketching: An artist's interface for creating multivariate time-varying data visualizations. IEEE Transactions on Visualization and Computer Graphics, 22(1):877–885, Jan 2016.
16. G. C. D. Silva, T. Yamasaki, and K. Aizawa. Sketch-based spatial queries for retrieving human locomotion patterns from continuously archived GPS data. IEEE Transactions on Multimedia, 11, 2009.
17. T. Tsandilas, A. Bezerianos, and T. Jacob. Sketchsliders: Sketching widgets for visual exploration on wall displays. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, Seoul, Republic of Korea, April 18-23, 2015, pp. 3255–3264, 2015.
18. C. Turkay, A. Slingsby, H. Hauser, J. Wood, and J. Dykes. Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data. IEEE Transactions on Visualization & Computer Graphics, 20(12):2033–2042, 2014.
19. F. Wang, W. Chen, F. Wu, Y. Zhao, H. Hong, T. Gu, L. Wang, R. Liang, and H. Bao. Visual reasoning approach for data-driven transport assessment on urban road. In IEEE Conference on Visual Analytics Science and Technology, pp. 103–112. IEEE, Oct. 2014.
20. Q. Zeng, W. Chen, H. Wang, C. Tu, D. Cohen-Or, D. Lischinski, B. Chen, Hallucinating Stereoscopy from a Single Image, Computer Graphics Forum, 34(2):1-12, May, 2015.

ABOUT THE AUTHORS

Shamal AL-Dohuki is a Ph.D. candidate in the CS Department, Kent State University. His research interests include urban data management and visualization, the visual query of trajectory data, and semantic data query and analytics.

Ye Zhao is a professor in the CS Department, Kent State University. His research interests include visual analytics of urban data, medical image processing, and information visualization.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2911230, IEEE Computer Graphics and Applications

Farah Kamw is a Ph.D. candidate in the CS Department, Kent State University. Her research interests include visual analytics of urban trajectory data, urban accessibility, and urban navigation.

Jing Yang is a professor in the CS Department, the University of North Carolina at Charlotte. Her research interests include visual analytics of multidimensional data, time-oriented data, text data, and urban transportation data.

Xinyue Ye is an associate professor in the Department of Informatics, New Jersey Institute of Technology. His research interest includes spatial social network analysis, urban computing, and GIS.

Wei Chen is a professor in CAD&CG, Zhejiang University. His research interests include visualization, visual analytics, and bio-medical image computing.