Erschienen in: EnvirVis 2018 : Workshop on Visualisation in Environmental Sciences / Rink, Karsten et al. (Hrsg.).
- Goslar : The Eurographics Association, 2018. - ISBN 978-3-03868-063-5
https://dx.doi.org/10.2312/envirvis.20181138

Visual Analysis of Urban Traffic Data based on High-Resolution and High-Dimensional Environmental Sensor Data

Johannes Häußler[1], Manuel Stein[1], Daniel Seebacher[1], Halldor Janetzko[2], Tobias Schreck[3], Daniel A. Keim[1],

[1]University of Konstanz, Department of Computer and Information Science, Germany
[2]University of Zurich, Department of Geography, Switzerland
[3]Graz University of Technology, Institute for Computer Graphics and Knowledge Visualization, Austria

Abstract

Urbanization is an increasing global trend resulting in a strong increase in public and individual transportation needs. Accordingly, a major challenge for traffic and urban planners is the design of sustainable mobility concepts to maintain and increase the long-term health of humans by reducing environmental pollution. Recent developments in sensor technology allow the precise tracking of vehicle sensor information, allowing a closer and more in-depth analysis of traffic data. We propose a visual analytics system for the exploration of environmental factors in these high-resolution and high-dimensional mobility sensor data. Additionally, we introduce an interactive visual logging approach to enable experts to cope with complex interactive analysis processes and the problem of the reproducibility of results. The usefulness of our approach is demonstrated via two expert studies with two domain experts from the field of environment-related projects and urban traffic planning.

1. Introduction

Global warming and climate change are nowadays among the most pressing global problems [Car10]. These problems are also closely related not only to industrial production but also to the worldwide trend of urbanization, with a strong increase in public and individual transportation needs. With more than 30 billion tons of carbon dioxide (CO_2) emitted yearly by the population since 2006 [co2], CO_2 contributes 70 % to the anthropogenic greenhouse gases. Effects such as the *greenhouse effect* as well as presumed environmental consequences thereof (e.g., storms, floods, droughts, ice melting in the polar regions) become increasingly visible. However, the fact that some impacts of the climate change cannot be avoided but rather accommodated has led to an ongoing discussion in politics, economics, and science. Resulting climate conferences on the national and international levels aim to promote climate protection policies to regulate and reduce the CO_2 emissions globally [GVB97].

The strong human desire for mobility is another observable global trend. Traveling to any place of choice in a fast, cheap and safe manner has become a perceived basic need. This increasing demand for mobility is reflected by the worldwide CO_2 emissions where motorized individual transport is contributing an essential portion to the global CO_2 emissions such that it was the second largest sector in 2014 [Age16]. The health effects caused by the world's increasing traffic are, additionally, not only restricted to the emitted pollutants and emission gases but also to factors such as noise pollution. In a recent study, about 40 % of the population in EU countries are exposed to road traffic noise at levels exceeding 55 dB(A) [WHO]. Exposure to noise levels is also expected to have negative psychosomatic effects on affected persons in the long term.

New concepts in the field of transport planning and real-time traffic systems need to be developed, to satisfy both mobility and health demands. Resulting systems aim to take advantage of high-dimensional data sources such as vehicle-sensor data [Bur04], also enabling communication with nearby vehicles or infrastructure. Other potential data sources include videos or photos from traffic cameras as well as sensors integrated into the street itself or installed at toll stations. Recently, also *crowd-sourcing* data including anonymized information such as the vehicle's current GPS-position, its speed as well as precise values concerning other vehicle-specific properties such as fuel consumption are gaining more and more interest. These properties can be used to estimate environmental factors such as CO_2- and noise-emissions [BRS*15]. Until now, the analysis of traffic flow data focused on inspecting vehicle movement at specific points of interest such as crossings or tunnels to optimize traffic light circuits [ONL16].

The available data, however, does not only allow conclusions for prevalent traffic characteristics (e.g., travel time) but instead should enable addressing challenging tasks (further introduced in Section 3) such as:

- **T1:** Detection of current pollution situations in a city for policy actions (mid-term)
- **T2:** Regulation/adaption of traffic flows to minimize congestion and reduce pollution (short term, can probably be done in real-time)
- **T3:** Prediction of future mobility demands by urban population (mid to long-term)
- **T4:** Monitoring of individual drivers/vehicles for regulation by cities (e.g., to temporarily revoke driving permission on high-pollution days), or by insurance companies to set incentives for drivers to drive more eco-friendly (mid to long-term)

In this work, we focus on **T1** and **T2** by proposing a holistic view to the analysis of mobility data by helping experts to develop and realize sustainable mobility concepts. We consider visual analysis as an intuitive way to interact with this kind of mobility data supporting analysts to create, refine and verify hypotheses. Using data from the citizen-science platform *enviroCar* [BRS*15], we contribute a Visual Analytics system allowing analysts to leverage their background knowledge in the analysis process. Therefore, we provide a suitable set of visualizations, focusing on different aspects of the data including geographic, temporal, and multi-dimensional aspects as well as visualization techniques combining these. To this end, the system provides multiple interactive views linked via state-of-the-art techniques such as *Brushing and Linking* as well as extensive and intuitive filter mechanisms to widen the exploration possibilities. Furthermore, we enable the expert to treat environmental factors such as CO_2 emissions equally to common traffic key performance indicators (e.g., stop times). By including these environmental factors into the analysis, the expert can gain additional knowledge connecting both, the traffic and environment domain. Additionally, we contribute interactive visual logging approach to support traceability and orientation of the analysis process. Each user interaction is captured and visualized separately allowing the analyst to keep an overview of the complex interactive analysis process and to cope with the reproducibility of results. Consequently, we strengthen the trust in the system and the analysis results.

2. Related Work

We discuss research related to our presented approach in visual movement and traffic analysis, followed by an outline of related techniques in the field of analytical provenance.

2.1. Visual Movement and Traffic Analysis

Typical movement analysis tasks are the detection of regularly recurring patterns, outliers or the analysis of performance. here has been done much basic research in the past regarding the analysis of movement [AAB*13, AA06] and the application areas are manifold. This includes, for example, sports analysis [SJL*18] or the analysis of animal movement [SBJ*11]. The handling of trajectories involves many challenges, e.g., the generation as well as the extraction of semantic information, the analysis via data-mining algorithms as well as the extraction of useful trajectory characteristics [PSR*13]. To implement a successful analysis of movement data, typically a combination of several tools, methods, and procedures is proposed [AAW07, AA13]. Nevertheless, there are still open challenges and active research [AAC*17].

The analysis of traffic trajectories differs from other domains in various ways as trajectories are bound to the street network and traffic rules. Many application-driven approaches have been developed to address challenges of urban planning concerning traffic and mobility. For example, several systems have been proposed for the effective visual exploration of taxi trips to get an understanding of cause and effect of traffic jams [WLY*13, FPV*13]. Furthermore, visualizations of travel time and reachability in a city transportation network as well as approaches for the identification of locations with increased traffic volumes have been proposed [ZFA*14, AAH*11]. To this end, Prouzeau et al. proposed a prototype for the visual exploration of real-time traffic data on high-resolution wall displays [PBC16].

2.2. Analytical Provenance

Visual exploration requires an interactive and collaborative process between human and computer (*human-in-the-loop*) in order to allow analysts to confirm or falsify hypotheses. Novel interaction techniques are fundamental to optimize usability and thus, the complete analysis process [LIRC12]. A historical record of the exploration process allows analysts to keep track of their actions, to reproduce the whole (or specific parts) of the analysis as well as improved collaboration and presentation of the analysis. According to the iterative and exploitative nature of visual exploration, various systems have been proposed to characterize meaningful interactions [HMSA08, GZ09]. Resulting systems often also aim to not only capture user interactions but also to note and rate relevant insights to increase trust in the analysis process [SBFK16]. In the end, analysts do not only want to capture and visualize provenance data, but also reduce the complexity of the visualization by aggregating and highlighting relevant parts to the user using a modular degree-of-interest function [SLSG16].

3. Visual Analysis of Urban Traffic Data

To gain a deeper understanding of the needs of professional analysts in the domain of traffic analysis, we conducted several informal expert interviews. The invited domain expert has long-term experiences in the application, development, and sales of software solutions in the traffic area. Furthermore, he leads a company in the area of urban traffic planning. Therefore, he is highly experienced in consulting regarding the implementation of cooperative traffic systems and the analysis of traffic controls and transport systems regarding pollutants. Consequently, we consider his insights highly valuable for the design and implementation of our system. In our discussions, we found that the current analysis of traffic data is limited due to the **tasks**, **tools**, and available **data** in use by traffic experts. Traffic experts are mainly trying to determine key factors of traffic flow. To do so, analysts look for certain characteristics of traffic data such as travel time, speed, number of stops, waiting times, and other factors proposed by different manuals (e.g., the *Highway Capacity Manual* [Man00]). Depending on the task and characteristics, analysts need to obtain different views on the

(a) (b) (c) (d)

Figure 1: Visualizations of all data points within a certain area around an intersection colored based on their CO_2 emission applying the *viridis* color scheme [Gar15]. **(a)** *Dot Map* of a frequently traveled intersection. **(b)** *Dense-Pixel-Display* visualization of the same intersection sorted based on their CO_2 emission. **(c)** Data points sorted based on their speed. **(d)** Data points sorted based on their engine speed. The pixels' drawing order within the *Dense-Pixel-Displays* (b)-(d) left to right and top to bottom.

(a) (b)

Figure 2: Visualization of trips' trajectories colored by their mean CO_2 emission applying the *viridis* color scheme [Gar15]. **(a)** Exact course of all trip trajectories in the german city of Mönchengladbach between 0 am and 6 am. **(b)** Magnification of the selected region in Figure (a) highlighted by the red rectangle.

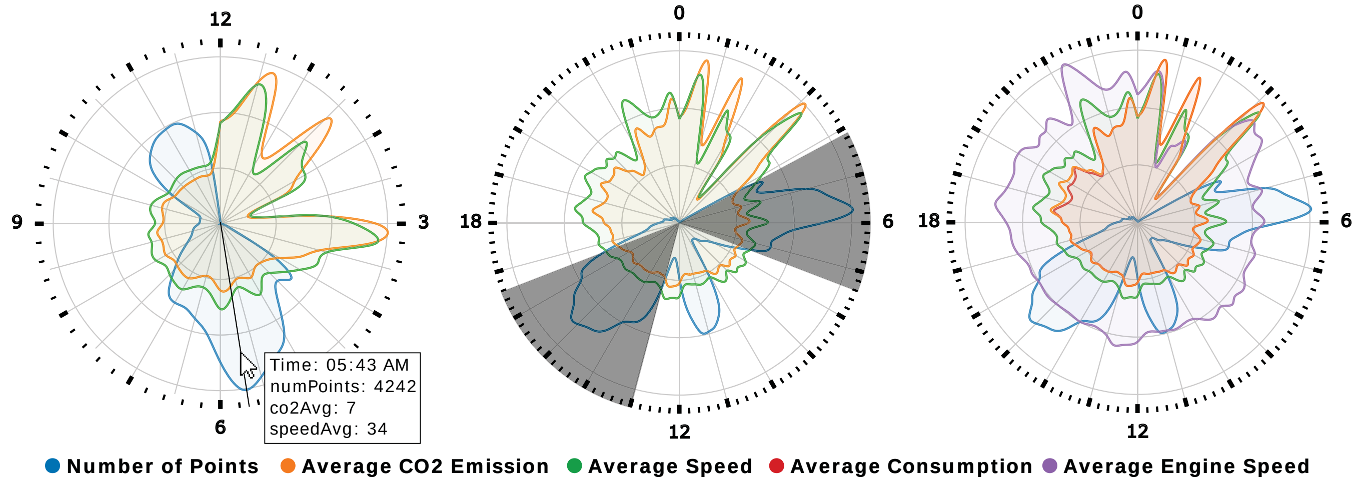
available data. The geographical scale may range from very specific geographical points to a selection of route segments, depending on the goal of the analysis. The inspection of trips through specific geographic points, for example, enables the optimization of traffic light installations [ONL16].

Environmental measures such as particulate matter or noise emissions are considered as well when analyzing traffic data, but are mostly regulated by local approaches such as prohibiting roads for specific vehicle types (environmental zones [SÜD16]) or setting a low speed limit for specific roads with the attempt to reduce the noise emission level. The issue for finding generalizable solutions lies within the lack of available mobile and high-resolution datasets. As a consequence, the used data covering the introduced tasks often comes from limited statistical data such as traffic census data from manual notes, video cameras, toll stations, sensors on traffic lights, and other sensors integrated into the traffic infrastructure. Furthermore, the expert explained that analysts do not only want to conclude common key traffic indicators such as travel speed or driving profile, but also be able to improve their understanding of environmental-related factors in their data. The available data, however, often does not provide the needed information. Experts are, additionally, facing a lack of expertise as they are nowadays trained properly in geography but often miss a special education in the field of visual data analysis. Concluding our pre-study, we formulated several design goals **T1 - T4**, as can be seen in Section 1. In our proposed interactive system, we focus on task **T1** and **T2**. Nevertheless, we assume that the other tasks can be solved by interactive visual analyses approaches as well. As an additional feature, we provide a video showing the different visualizations and interactions (`http://files.dbvis.de/haeussler/Visual_Analysis_of_Urban_Traffic_Data.mp4`).

3.1. Data

To support analysts in inspecting environmental factors in urban traffic data as well as developing generalizable solutions, we reviewed several promising data sources. The citizen-science platform *enviroCar* (`https://envirocar.org/`), for example, has grown to a large platform, providing *crowd-sourced* traffic data easily accessible via a REST-Interface. Its active community leads to continued growth of the dataset. Currently, the available data contains approximately more than 1.7 million data points in more than 5800 trips (2012 to 2017). This data is collected from the existing on-board diagnostics, using Bluetooth adapters connected via the standard ODB-II interface. Each single data record can be identified as part of a trip and contains 24 attributes reflecting sensor values of the vehicle (e.g., speed, rpm, ...) and a CO_2 estimation. Every trip is, furthermore, associated with a sensor and thus a particular car. As not every car is equipped with every sensor, the data is somewhat sparse. Specifically, 36% of the values are missing.

To derive meaningful measures and statistics for traffic experts, we preprocessed the data of *enviroCar* according to the KDD pipeline. As an essential part of the preprocessing step, we removed noisy data points detected by a rule-based approach. For example, if a vehicle stands still for a certain amount of time without moving and, furthermore, the engine is turned off such that no more vehicle sensor data is recorded, the noisy data points can be removed, and the trips can be segmented. Other occurrences of noise are due to GPS inaccuracies. Additionally, we enriched the available data by integrating datasets from OSM (`https:`



(a) (b) (c)

Figure 3: Visualization using a clock metaphor, displaying the temporal development of several data attributes. **(a)** Clock with a 12-hour layout, displaying the number of data points, the average CO_2 emission, and speed between 12 am and 12 pm. **(b)** Clock with 24-hour layout and a selection of two potential maximum values. **(c)** Additional superimposed attributes. During rush hours, although there is a high number of data points available, the mean of the CO_2 follows a steady trend, while there are several peaks past midnight.

`//www.openstreetmap.org)` such as the road network, to allow the derivation of semantically meaningful aggregations.

3.2. System Design

To support an exploration of mobility data regarding environmental factors, our proposed system needs to provide a set of interactive visualizations, linked via *Brushing and Linking*. Analysts need to be able to easily get an overview of the data and focus on certain areas or time intervals. To achieve this, we follow the visual information seeking mantra of Ben Shneiderman which is “Overview first, zoom and filter, details on demand” [Shn03]. Therefore, our system provides various interaction techniques to generate different views of the data by either aggregating or filtering the data or by changing the visual appearance of the visualization.

The central component of our system is an interactive map as for analysts, the most important aspect of nearly every task is described by the geographic context of the data. The map can be individually enhanced with different visualization layers, while each layer is individually designed with respect to the data and corresponding task. These layers include, for example, trajectory, road network, and animation overlays. Additional focus and context techniques such as an integrated minimap with a zoomed view of hovered (highlighted) elements enable keeping the current context while focusing on specific user-selected elements, e.g., trajectories through a specific crossing. In the following, we describe our proposed visual analysis methods for solving **T1** and **T2** in detail.

3.2.1. T1 – Detection of Current Pollution Situations

According to our expert, for determining the current pollution situation, it is important to focus on the characteristics of individual data points close to specific infrastructures instead of whole trips, according to our expert. Important infrastructure, for example, includes whole districts as well as road network characteristics such as crossings. Furthermore, areas where infrastructural changes have been implemented are of additional interest.

An overview of the available trip data is provided by a trajectory layer where the trips’ trajectories are drawn onto the map. Additionally, travel start and destination points are encoded as blue/red dots. An example of the city Mönchengladbach (Germany) is depicted in Figure 2. By coloring each trajectory according to a data attribute (e.g., CO_2), we provide a rough overview of the trips’ characteristic. In Figure 2 (a), for example, one can easily spot parts of freeways surrounding Mönchengladbach due to the darker colors. Nevertheless, some parts of the inner city are colored darker than others and may serve as a first lead towards further investigations within this area. To get a less cluttered overview of the current pollution situation close to traffic infrastructure, the system also provides a view of the road network. The amount of points encoded by a particular road segment is visualized by its width which allows an intuition about busy roads. Additionally, color is used to encode an aggregation of a data attribute of interest such as CO_2.

Our system, furthermore, provides spatial filtering by switching into a *draw-mode* and selecting areas of interest allowing to exclude unwanted roads. For example, an expert may want to focus on a city center while excluding all freeways surrounding it. To enable experts to drill down to inspect all data points within a particular area, the system provides a pixel-based visualization. Several examples are depicted in Figure 1, where all data points within a previously filtered area are shown. As shown in Figure 1 (a), a dotmap is not sufficient due to overlapping elements and visual clutter. Therefore, our system provides a *Dense-Pixel-Display* [Kei00],

Figure 4: Single frame of an animation of several trips within a user-defined area visualized by a dot-map. The temporal context is represented as an interactive temporal legend.

where each pixel represents exactly one data point and the color of the pixel encodes its value (Figure 1 (b)-(d)). Here, we intentionally decided to neglect the geographical position of the data points within this focused view allowing us to additionally encode the position of each pixel. In Figure 1 (b), color and position encode the CO_2 value of each data point while in contrast for Figure 1 (c) and Figure 1 (d) the position of each pixel is determined by speed and engine speed respectively. Comparing Figures 1 (c) and (d) indicates that there are correlations between *speed*/*engine speed* and CO_2 emission, while the *engine speed* seems to have a stronger correlation.

3.2.2. T2 – Regulation of Traffic Flows

The nature of traffic flows implies a high sensitivity to time in general. Accordingly, reoccurring patterns and anomalies concerning day time such as the hour of day, day of week, etc. are of special interest to traffic experts. To provide an intuitive overview of the temporal distribution of the available traffic data such as the number of data points, average speed, and CO_2 value, we developed a radial visualization (*Clock View*) based on a clock metaphor, as can be seen in Figure 3. Due to its clock-like structure, the *Clock View* enables analysts to quickly detect intervals of interest (e.g., due to trends or peaks). The shown data dimensions of the Clock View can be interactively adjusted. The *Clock View* visualizes and orders the corresponding data values aggregated by the average on a minute per day base. We, additionally, incorporated various interaction techniques to enable analysts to filter the dataset by intervals of interest, e.g., by certain points in time and/or time intervals. By default, the clock layout shows all 24 hours of the day such that 12 AM is at the very bottom while 12 PM is at the very top. Although it is a very detailed layout and reflects the periodicity of the hours of a day, it does not reflect the exact metaphor of an analogous clock. We, consequently, provide a second layout, where either PM, AM, or an aggregated version of AM/PM is shown such that only twelve hours are displayed.

Besides the *Clock View*, we developed an *Aggregated Animation* layer as a map overlay to provide a spatio-temporal context visualization of traffic flows. Enabling users to control the animation according to their needs and cognitive abilities, we integrate several interaction possibilities (e.g., for animation speed, temporal hints, and navigation) directly into the map. Due to the temporal spread recorded trip data, an animation in chronological order is not suitable in our case. Therefore, we start the animation of all trips at the same point in time. During the animation, we aim to communicate the temporal context to a degree that the user is able to preserve a steady temporal orientation. This is realized via an interactive temporal legend of the animation time. A smooth animation of the trips ($\approx$ 25 fps) is further realized by linearly interpolating two temporally neighbored data points, which are usually 15 seconds apart from each other. Figure 4 depicts a single frame of an animation of several trips within a fixed, geographical, user-defined area where each frame shows a dot-map. The color of each point corresponds to the current CO_2 emission of the particular data point and might change drastically, depending on the current driving behavior. To further reduce the amount of overlapping data points, the user is enabled to geographically navigate (e.g., via *Zoom & Pan*) in addition to the temporal navigation.

4. Visual Interactive Logging

Performing a visual analysis of urban traffic data is a complex task that requires multiple visualizations on the different aspects of the data. Various visualizations provide numerous interaction techniques for performing visual data analysis such as navigation, zooming for the map and filtering for the clock-based glyph visualization. In our case, we used *Brushing and Linking* to keep all visualizations consistent. However, performing interactions in various visualizations and chaining together filtering, aggregation and navigation steps can quickly become overwhelming and may lead to a lost overview of the analysis process. It can further lead to frustration with the system and also a loss of trust in the findings. Another important aspect is the issue of reproducibility of results arising from a systems' inherent complexity. To cope with both of these problems, we need to enable the analyst to maintain an overview of the previous analysis process. One option would be a plain history approach. By logging which interactions were performed and by providing a sorted list of these interactions, the analyst should be able to get an overview of the analysis process. However, in a real scenario, an analysis is usually not a straight-forward process. As, for example, described in the knowledge generation model of Sacha et al. [SSS*14], the process consists of various loops to explore data, to verify hypothesis and to gain knowledge. Thus, we need to provide a more extensive overview of the analysis process.

In our as well as in most general cases, the analysis process consists of a set of system states V, which can be changed by performing a set of possible interactions I. The initial state of a system is indicated by V_0, which is transformed into state V_1 by performing interaction I_1. The index variables serve as an indicator for the temporal ordering of states and interactions, with $t(V_0) < t(V_i) < t(V_n)$ and $t(I_1) < t(I_i) < t(I_n)$. A simple, straight-forward analysis pro-

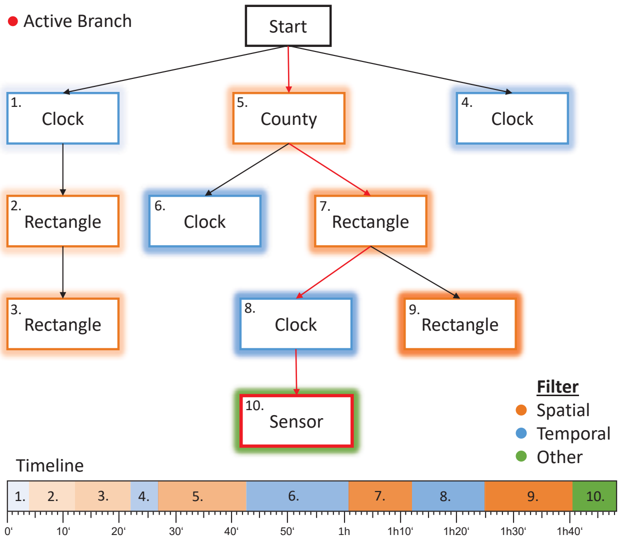


Figure 5: Interactive visualization prototype of the logged analysis process. Each node of the hierarchical tree represents one filter operation. The colors of the nodes' outline indicate the data type used for filtering. The edges and node highlighted in red show the current analysis branch. Additionally, a timeline of the analysis process is provided, showing the individual states of the system, as well as their duration, during the analysis process.

cess can then be illustrated as a linear chain of interactions $I_1, \ldots, I_n$, which lets the user visit and inspect the different system states $V_1, \ldots, V_n$. However, usually, the analysis process is more complex. For instance, users could undo previous interactions or perform completely new exploration processes. Thus, we imagine the analysis process as a graph, in particular, a tree. A subtree corresponds to one chain of thought such as a rush hour analysis for a particular intersection, and the individual branches of the subtree correspond to various filter and aggregation strategies. Individual interactions within a branch do not have to be consecutive, since a user can try different strategies or jump back to a previous state to pursue a different train of thought, but they are still in chronological order. If analysts finish their analysis process, either because they have not found anything or they have accepted or rejected a previously assumed hypothesis, they can get an overview of their performed interactions and visited system states by following the path from the root node to the current leaf node. This path allows the user to review the analysis process and rate the performed interactions, but also allows for the reproducibility of the performed analysis.

Our interactive visualization prototype of the logged analysis process, as shown in Figure 5, consists of two main components, a hierarchical tree visualization as well as a timeline. The tree visualization represents each filtering interaction and resulting system state preserving a logical structure of the complete analysis process. Additionally, we maintain the linear, time-oriented order by visualizing a timeline beneath the tree. The timeline is divided into multiple rectangular parts where each rectangle represents one state of the system and accordingly one node of the tree. The width of a rectangle refers to the duration of the particular system state to indicate how long a user stayed within a specific analysis step. Each node of the tree, as well as its associated rectangle within the timeline, are given a certain color representing the type of interaction (e.g., temporal or spatial filtering, ...). The coloring enables the user to review the interactive analysis process and inspect the characteristics that drove the analysis and finally lead to a particular path (subtree). To enable a clearer association between a node of the tree and the corresponding rectangle within the timeline, we employ the visual variable brightness and encode the system state's creation timestamp. Thus, the color of a node and its corresponding rectangle within the timeline are characterized by time and category. Consequently, nodes which are lying further in the past are outlined brighter than more recent nodes. We additionally integrate various ways of interaction, for example, by selecting a node within a tree or a rectangle within the timeline, the user can navigate to a particular system state. Besides traditional *undo* and *redo* operations, this navigational feature enables the domain expert to revisit certain steps of the analysis or to continue the analysis process from a specific system state.

5. Evaluation

We invited the expert from our pre-study (Section 3) as well as an additional expert without prior knowledge of our developments to evaluate our proposed system via several expert studies. The newly invited expert is employed at the State Office for the Environment, Measurements and Nature Conservation of the Federal State of Baden-Württemberg (*LUBW*) where he coordinates several environment-related projects. At the beginning of the study, we wanted to verify that our chosen design corresponds with the experts' expectations of a system for the visual analysis of urban traffic data. The participants were given a blank geographic map only showing the basic infrastructure of an urban area in order to test such a proof of concept. Each expert was then asked to enhance the map with visual elements to describe the issue of pollutant emissions.

After the experts were introduced to our developed system, they were given pre-defined tasks based on **T1** and **T2** to provide them with a comprehensive impression of the diverse functionalities and interaction possibilities. Afterward, a semi-structured interview was performed to gather additional insights about the general user-experience. Additionally, we asked specific questions regarding the diverse visualizations of our developed system. During the tasks, we encouraged the participants to express *ad-hoc* comments via the *thinking aloud* method [BR00]. Additionally, screen-, audio-recordings, as well as manual notes made by the examiner, provided the main data gathered during the evaluation.

The results of our qualitative evaluation are very promising. Regarding our proof of concept, both experts used color to map various attributes such as CO_2-emission on the road network. Additionally, they used visual variables, such as size, to visualize the traffic volume. One expert especially mentioned that he likes the possibility to visualize data directly on the road network if he has the opportunity to display additional information such as signaling systems or speed limits. Still, both experts mentioned that a purely map-based approach is not sufficient to visualize all environmental factors. We take the statements of both experts as confirmation that the vision of domain experts corresponds to our chosen design. During the following familiarization with our system and the subsequent semi-structured interview, we collected feedback from the experts. We provide an overview of the expert feedback concerning the various components of our system as follows: Both experts state that they consider the initial overview of the trip data very important. One expert introduced use cases in which a comparison between diesel and gas vehicles is useful and, therefore, considers the visualization of the sensors and the possibility of filtering according to sensor-specific properties to be valuable. The other expert had the opinion that the possibility to get details on demand allowed him to infer interesting rides or interesting geographical areas. Furthermore, both experts reported they had a positive experience performing the visual analysis using the *Aggregated Animation* layer introduced in Section 3.2.2. One expert especially reported that with the help of the animation, he was able to develop a better understanding of the movement of the trips. This allowed him to identify patterns that he could not see with static visualizations. The other expert added that he sees the animation as being of particular use in communicating environmentally relevant facts on certain routes. For example, when presenting knowledge to decision-makers at administrative and political level.

The visual interactive logging, as described in Section 4, is also seen beneficial from our experts. According to the experts, our approach promotes collaborative work and the exchange of analysis processes between experts. In certain circumstances, such as very complex analyses with many branched steps, they would also like to use this component as a note-taking environment or to present findings. Consequently, both experts are convinced that the system can be used to perform manifold urban traffic data analysis. They particularly appreciate the usability of the presented system and emphasize that it is not necessary to know the data in detail before working with the system. Furthermore, according to our experts, no previous specialized knowledge is required. One expert sees a special analysis potential in using the system to verify the effectiveness and efficiency of already implemented projects in the transport sector and to check whether further improvements are necessary or possible. For instance, by comparing periods before and after changes in the road layout and how these changes affect the surrounding road network. Another possible application would be to improve communication between traffic engineers and administrative levels by using the system to present important findings. Eventually, both experts have expressed the need to filter the data based on fixed time periods and to visualize the characteristics of a weekday or a certain date. They mention that an indication of how many trips are assigned to a visual element is of high interest. They would welcome an encoding of the number of trips, or if there would be a possibility to switch between the number of trips and the number of data points. One expert mentioned that the relationship between traffic and environment will become even more important in the future.

6. Discussion and Conclusion

We presented an interactive visual system for the exploratory analysis of environmental aspects in urban traffic data. Accordingly, our proposed system implements various interactive visualizations to allow dynamic and versatile data exploration. The results of our qualitative evaluation are very positive and, furthermore, allow us to formulate future challenges for our research in this field.

The used *enviroCar* dataset only reflects a small sample of the whole traffic. Thus, it is nearly impossible to conclude to the whole traffic using this dataset. We tried to overcome this issue by aggregating data points at specific locations to make sure that experts can at least get intuitions of the traffic flow. In future work, we plan to extend our analysis by integrating additional data sources from external data providers such as *HERE* (`https://www.here.com/`) or *Google Maps* (`https://maps.google.com/`). Combining diverse data sources may reveal specific insights concerning traffic flows or traffic infrastructure and allow semantically enriched and more detailed analyses. For example, one can imagine a scenario where analysts can derive actual pollutant emissions (*enviroCar*) of the overall traffic volume (*HERE*) within specific areas. Furthermore, we used the CO_2 emission as an exemplary environmental characteristic to introduce the different visualizations. If a precise estimation of noise emissions is possible in the future, analyses of this environmental characteristic would be of great interest as well.

In future work, we plan to create additional visual representations to give more insights into the data. A graph visualization, for example, where the size of every node represents the average waiting time at crossings and the thickness of every edge represents the average speed could provide an intuitive overview about possible movement patterns in the data. Additionally, we would like to improve the existing visualizations. This includes, for example, interactively or automatically changing the scaling technique of the animation time of the animation layer to avoid cognitive overload, e.g., depending on the rate of change. The implementation of a space-filling curve for the dense-pixel-display visualization. On the machine learning side, we plan to use relevant data characteristics of trips as well as individual data points as training data for the semi-automatic classification of interesting situations in the available data.

Furthermore, we presented an interactive visual approach for the semantic traceability of the analysis process. User interactions are captured and visualized as a tree. Interacting with this tree enables user to restore previous workflows. We consider the current design of the interactive visual logging as work-in-progress. We plan to build on top of this system to allow analysts to create more compact and structured representations of their workflows by removing irrelevant nodes. Additionally, we want to provide a visual abstraction of this visualization by semantically grouping interactions based on their underlying intentions. Another planned extension for the visual interactive logging is a presentation mode. This would make it possible to extract relevant nodes from the tree as a structured presentation, for example, to show insights to a broader audience and foster collaborative working.

Eventually, making use of crowd-sourcing data with platforms such as *enviroCar* could help to protect the environment as well. It could be very interesting to cluster locations which often serve as starting or ending points of a trip. This could be used to support ride sharing among the users of each platform.

References

[AA06] Andrienko N., Andrienko G.: *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer Science & Business Media, 2006. 2

[AA13] Andrienko N., Andrienko G.: Visual Analytics of Movement: An Overview of Methods, Tools and Procedures. *Information Visualization 12*, 1 (2013), 3–24. 2

[AAB*13] Andrienko G., Andrienko N., Bak P., Keim D., Wrobel S.: *Visual Analytics of Movement*. Springer Science & Business Media, 2013. 2

[AAC*17] Andrienko G., Andrienko N., Chen W., Maciejewski R., Zhao Y.: Visual Analytics of Mobility and Transportation: State of the Art and Further Research Directions. *IEEE Transactions on Intelligent Transportation Systems* (2017). 2

[AAH*11] Andrienko G., Andrienko N., Hurter C., Rinzivillo S., Wrobel S.: From Movement Tracks through Events to Places: Extracting and Characterizing Significant Places from Mobility Data. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on* (2011), IEEE, pp. 161–170. 2

[AAW07] Andrienko G., Andrienko N., Wrobel S.: Visual Analytics Tools for Analysis of Movement Data. *ACM SIGKDD Explorations Newsletter 9*, 2 (2007), 38–46. 2

[Age16] Agency I. E.: *CO2 Emissions from Fuel Combustion 2016*. Organization for Economic Co-operation and Development (OECD), 2016. 1

[BR00] Boren T., Ramey J.: Thinking Aloud: Reconciling Theory and Practice. *IEEE Transactions on Professional Communication 43*, 3 (2000), 261–278. 6

[BRS*15] Bröring A., Remke A., Stasch C., Autermann C., Rieke M., Möllers J.: enviroCar: A Citizen Science Platform for Analyzing and Mapping Crowd-Sourced Car Sensor Data. *Transactions in GIS 19*, 3 (2015), 362–376. 1, 2

[Bur04] Burelle C.: On-Board Diagnostics II (OBDII) and Light-Duty Vehicle Emission Related Inspection and Maintenance (I/M) Programs. *Transportation Systems Branch, Environment Canada* (2004). 1

[Car10] Carlowicz M.: World of Change: Global Temperatures: Feature Articles. 1

[co2] CO2 Emissions (kt). `http://data.worldbank.org/indicator/EN.ATM.CO2E.KT?end=2013&start=2006`. 1

[FPV*13] Ferreira N., Poco J., Vo H. T., Freire J., Silva C. T.: Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2149–2158. 2

[Gar15] Garnier S.: Viridis: Matplotlib Default Color Map. *R package version 0.2* (2015). 3

[GVB97] Grubb M., Vrolijk C., Brack D.: *The Kyoto Protocol: A Guide and Assessment*. Royal Institute of International Affairs Energy and Environmental Programme., 1997. 1

[GZ09] Gotz D., Zhou M. X.: Characterizing users' visual analytic activity for insight provenance. *Information Visualization 8*, 1 (2009), 42–55. 2

[HMSA08] Heer J., Mackinlay J., Stolte C., Agrawala M.: Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008). 2

[Kei00] Keim D. A.: Designing Pixel-oriented Visualization Techniques: Theory and Applications. *IEEE Transactions on Visualization and Computer Graphics 6*, 1 (2000), 59–78. 4

[LIRC12] Lee B., Isenberg P., Riche N. H., Carpendale S.: Beyond Mouse and Keyboard: Expanding Design Considerations for Information Visualization Interactions. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2689–2698. 2

[Man00] Manual H. C.: Highway Capacity Manual. *Washington, DC* (2000). 2

[ONL16] Online R.: Autofahrer machen Grüne Welle selbst, 2016. URL: `http://www.rp-online.de/nrw/staedte/moenchengladbach/autofahrer-machen-gruene-welle-selbst-aid-1.6000854`. 1, 3

[PBC16] Prouzeau A., Bezerianos A., Chapuis O.: Towards Road Traffic Management with Forecasting on Wall Displays. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces* (2016), ACM, pp. 119–128. 2

[PSR*13] Parent C., Spaccapietra S., Renso C., Andrienko G., Andrienko N., Bogorny V., Damiani M. L., Gkoulalas-Divanis A., Macedo J., Pelekis N., et al.: Semantic Trajectories Modeling and Analysis. *ACM Computing Surveys (CSUR) 45*, 4 (2013), 42. 2

[SBFK16] Sacha D., Boesecke I., Fuchs J., Keim D. A.: Analytic Behavior and Trust Building in Visual Analytics. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers* (2016), Eurographics Association, pp. 143–147. 2

[SBJ*11] Spretke D., Bak P., Janetzko H., Kranstauber B., Mansmann F., Davidson S.: Exploration through enrichment: A visual analytics approach for animal movement. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2011), ACM, pp. 421–424. 2

[Shn03] Shneiderman B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The Craft of Information Visualization*. Elsevier, 2003, pp. 364–371. 4

[SJL*18] Stein M., Janetzko H., Lamprecht A., Breitkreutz T., Zimmermann P., Goldlücke B., Schreck T., Andrienko G., Grossniklaus M., Keim D. A.: Bring it to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis. *IEEE transactions on visualization and computer graphics 24*, 1 (2018), 13–22. 2

[SLSG16] Stitz H., Luger S., Streit M., Gehlenborg N.: AVOCADO: visualization of workflow-derived data provenance for reproducible biomedical research. *Comput. Graph. Forum 35*, 3 (2016), 481–490. URL: `https://doi.org/10.1111/cgf.12924`, `doi:10.1111/cgf.12924`. 2

[SSS*14] Sacha D., Stoffel A., Stoffel F., Kwon B. C., Ellis G., Keim D. A.: Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visual Analytics Science and Technology 2014) 20*, 12 (Dec. 2014), 1604 – 1613. `doi:10.1109/TVCG.2014.2346481`. 5

[SÜD16] Süd T.: Emissions Stickers, 2016. URL: `http://www.tuev-sued.de/auto_fahrzeuge/feinstaub-plakette/feinstaubplakette_ausland/england`. 3

[WHO] WHO: Data and Statistics. URL: `http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data`. 1

[WLY*13] Wang Z., Lu M., Yuan X., Zhang J., Van De Wetering H.: Visual Traffic Jam Analysis based on Trajectory Data. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2159–2168. 2

[ZFA*14] Zeng W., Fu C. W., Arisona S. M., Erath A., Qu H.: Visualizing Mobility of Public Transportation System. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec 2014), 1833–1842. `doi:10.1109/TVCG.2014.2346893`. 2