Big Data for Mobility Tracking Knowledge Extraction in Urban Areas

SoA – Visual Analytics

The main idea of Visual Analytics (VA) is to develop knowledge, methods, technologies and practices that exploit and combine the strengths of human and electronic data processing (Keim et al. 2008a). Visualization is the means through which humans and computers cooperate using their distinct capabilities for the most effective results. Visual analytics is "the science of analytical reasoning facilitated by interactive visual interfaces" (Thomas & Cook, 2005a, p. 4), which focuses on developing human-computer methods and procedures for data analysis, knowledge building, and problem solving (Keim et al. 2010a). Visual analytics leverages methods and tools developed in other areas related to data analytics, particularly statistics, machine learning and geographic information science (Andrienko & Andrienko, 2012a). Visual analytics tools address the complex task of presenting multi-dimensional information in several displays of interactive 2D projections. The main components of a VA tool consist of a set of selectors and aggregators in a combination with visual facilities to drill down into the available information and to maintain synchronized displays.

1 Visual Analytics for Big Mobility Data[1]

1.1 Transportation Data

(Andrienko et al. 2017a) focus on (i) data, (ii) movements and people behaviour, and (iii) modelling and planning.

Data and Data transformations

The proposed data typology distinguishes spatial events (bound to a certain location and lasting for a limited time), trajectories (chronologically ordered records describing position of a moving object) spatially referenced time series (chronologically ordered sequences of values of time-variant thematic attributes associated with fixed spatial locations or stationary spatial objects). Trajectories are either quasi-continuous (when it is possible to plausibly estimate intermediate positions) or episodic (in the extreme case only the origin and destination of the trajectory are known). A set of interesting representation methods for the different types can be found in (Andrienko et al. 2017a).

[1] The content of this section is based on (Andrienko & Andrienko, 2012a), Andrienko et al. (2017) (Andrienko et al. 2017a) and (Andrienko et al. 2016c). Most of the text below was taken from these papers and only slightly adapted.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 780754.

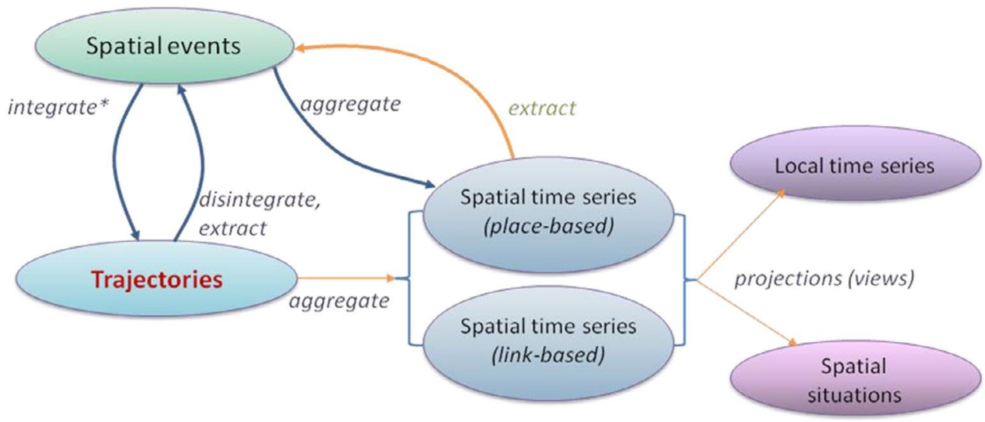

Figure 1: The principal transformations applicable to movement data, depending on the task/analysis goal (taken from (Andrienko et al. 2017a)).

A summary of possible transformations between the spatiotemporal data types is presented in Figure 1. The left part of the diagram shows the tight relationships between spatial events and trajectories. In fact, trajectories consist of spatial events: each record in a trajectory of an object represents a spatial event of the presence of this object at a specific location at some moment in time.

Other transformations may be beneficial for particular tasks. For example, (Chu et al. 2014a) transform trajectories of taxis into sequences of the names of the traversed streets and apply text mining methods for discovery of “taxi topics”, i.e., combinations of streets that have a high probability of co-occurrence in one taxi trip.

Andrienko et al. (2017a; 2017b) present an overview of visual representations and interactive techniques for detailed exploration of following classes of data: (i) individual movements, (ii) sets of taken routes, (iii) movement dynamics along a particular route, (iv) sets of origin-destination pairs, (v) collective movement over a territory, (vi) events (including extraction of events), (vii) contextualized movement (e.g. by integrating with meteorological data), (viii) impacts and risks (e.g. exposure to pollutants which is also a form of contextualizing).

Movement and people behavior

VA provides tools to investigate the use of transportation means by people. The existing techniques analyze the spatial and temporal patterns and trends, reveal behavioral differences between user groups, and relate the use of transport to the spatial and temporal context and people’s activities. Human mobility behaviors over public transit systems are commonly explored to identify commute patterns and reveal behavioral differences. For example, (Wood et al. 2011a) and (Beecham & Wood, 2014a) visualize and analyze the dynamic patterns of a bicycle hire scheme in London.

(Laharotte et al. 2015a) used Bluetooth detectors in Brisbane to create B-OD matrices to describe the dynamics of a subpopulation of vehicles to characterize urban networks. van der Hurk et al. (Hurk et al. 2015a) present a methodology for extracting passenger routes based on smart card data from the Netherlands Rail System. (Kieu et al. 2015a) explored the use of smart card data for passenger segmentation.

(Kruger et al. 2013a) develop an interaction technique, TrajectoryLenses. Complex filter expressions are supported by the metaphor of an exploration lens, which can be placed on an interactive map to analyze geospatial regions for the number of trajectories, covered time, or vehicle performance. Another work by Kruger et al. (2015a) enriches the trajectories of the scooter users with semantic information concerning the visited places to infer users’ activities and travel purposes. Semantic insights of points of interest are discovered from social media services. The uncertainties in time and space, which result from noisy, imprecise, and missing data, are visually analyzed by the geographic map view and a temporal view of OD patterns.

Mass mobility

The works described in this subsection deal with analyzing people's collective mobility behavior, i.e., mass movements. This includes routine daily and weekly patterns as well as anomalies due to extraordinary events.

Von Landesberger et al. (Landesberger et al. 2016a) present an approach to explore daily and weekly temporal patterns of collective mobility, where the source data are episodic trajectories of people reconstructed from georeferenced tweets or mobile phone use records. The trajectories are aggregated into flows between territory compartments by hourly intervals within the weekly time cycle.

(Beecham et al. 2014b) present a technique for automatically identifying com- muting behavior based on a spatial analysis of cyclists' journeys. They use visual analytics to compare the output of various workplace identification methods to explore data transformations and present insights to analysts in order to develop origin-destination theories of commute patterns.

(Ma et al. 2016a) also develop methods for studying urban flow. This work uses cell phone location records to approximate trajectories across a city, and flow volumes, links, and communities of users are visualized to help analysts identify typical patterns of movement within the city.

Similarly, work by Yang et al. (Yang et al. 2016a) focuses on identifying human mobility hotspots based on mobile phone location data from Shenzhen, China.

Yang et al. applies kernel density estimation and clusters identified hotspots based on the temporal signatures to identify spatial locations with high travel demand.

Work by (Chae et al. 2014a) develops a visual analytics framework for exploring public behavior before, during, and after disaster events. This work utilizes geographically referenced Tweets to create movement trajectories during disasters to identify evacuation flows. Interactions allow users to drill down into the data to also look at the underlying discourse occurring around the movements. Infrastructure data, disaster data (such as hurricane tracks), and Twitter data are all provided as map overlays in order to enable decision support and analysis.

People's activities and interests

In order to understand the current use of transportation systems and plan for expansion and development, it is helpful to understand the reasons why people travel, i.e., the activities and interests related to traveling. Recent work by Andrienko et al. (Andrienko et al. 2016a) presents a procedure for obtaining data similar to personal daily mobility diaries. Such a diary reports what places were visited by a person during a day, at what times, and for what purposes. The presented procedure aims at extracting similar information from long term sequences of spatio-temporal positions of people, which may come from georeferenced tweets or from mobile phone use records. From these sequences, the proposed procedure extracts repeatedly visited personal and public places along with the times these places were visited within the daily and weekly cycles. An interactive interface involving techniques for multi-criteria evaluation and ranking supports assignment of probable meanings ('home', 'work', 'eating', 'shopping', etc.) to subsets of places based on visit times and information about the land use or point of interest categories at these places. The analysis is done in a privacy-respectful manner without accessing individual data.

Modelling and planning

This section reviews research in visual analytics concerned with traffic modeling and transportation planning. This includes the derivation of models from data, applications of traffic forecasting and simulation models, transportation scheduling, and the exploration of decision options.

There is a series of works showing how predictive models of vehicle traffic can be derived from historical data consisting of a large number of vehicle trajectories (Andrienko & Andrienko,2013b), (Andrienko et al. 2015a), (Andrienko et al. 2016b). The approach is based on spatial abstraction and aggregation of the trajectory data into collective movements (flows) of the vehicles between territory compartments. The authors discovered that the dependencies between the traffic intensities and mean velocities in an abstracted transportation network at different levels of abstraction (Figure 2) have the same shapes as in the fundamental diagram of the traffic flow described in traffic theory (Gazis,2002a). While the fundamental diagram refers to links of a physical street network, it turns out that similar relationships also exist in abstracted networks. These dependencies can be represented by formal models (Figure 3).

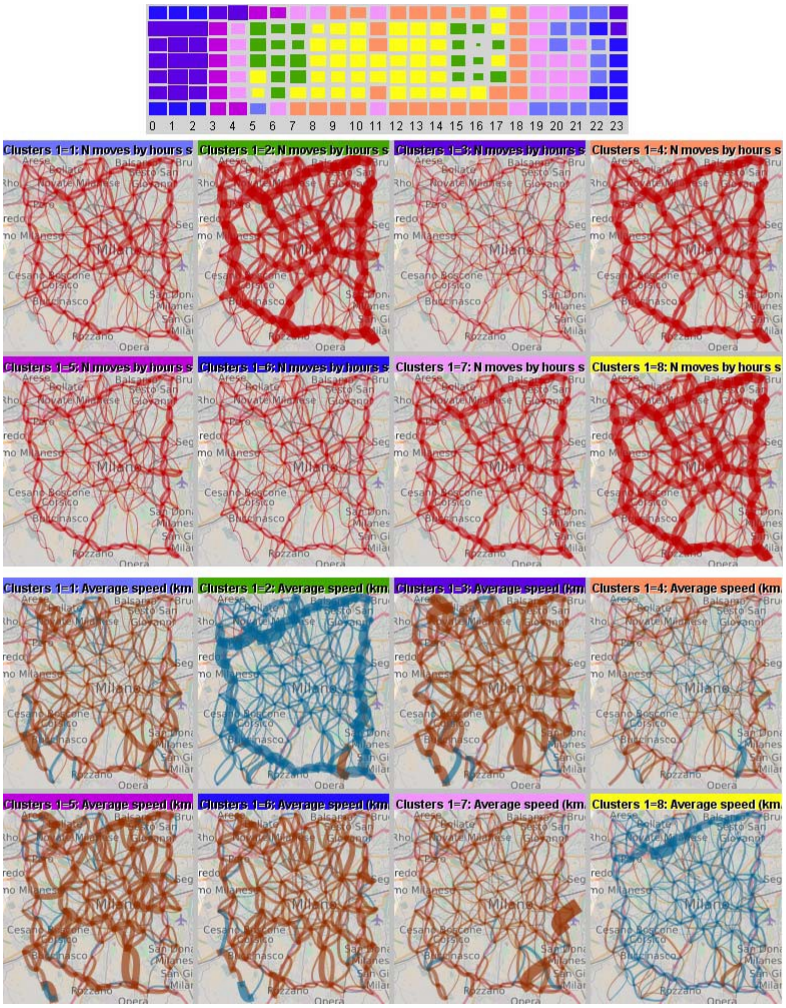

Figure 2: Hourly time intervals over a week have been clustered by the similarity of the spatial situations in terms of the flow magnitudes and average speeds. In a time matrix at the top, the rows correspond to the days from Sunday to Saturday and columns to the day hours. The time intervals are represented by rectangles coloured according to the cluster membership; the sizes show the closeness to the cluster centres. Below, representative spatial situations for the clusters are shown by flow maps. In the upper set of 8 maps, the widths of the flow symbols are proportional to the mean flow magnitudes. The lower set of 8 maps represents how the mean speeds in the clusters differ from the median mean speed attained on the links. Positive and negative differences are encoded by proportional widths of flow symbols coloured in brown and blue, respectively (Andrienko et al. 2013a).

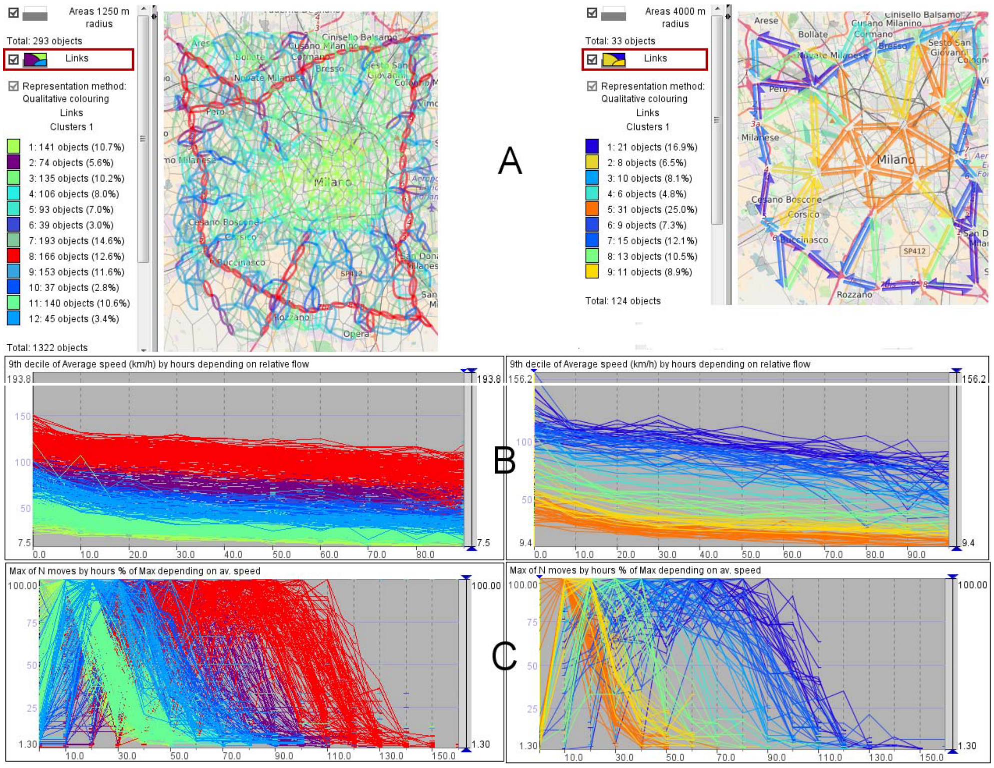

Figure 3: Dependencies between the traffic flow intensities (hourly volumes) and mean velocities on the links of an abstracted transportation network at different levels of abstraction. A: Abstracted networks with the cell radii of about 1250 m (left) and 4000 m (right). The links are clustered and coloured according to the similarity of the volume-speed dependencies. B: The dependencies of the mean velocity (vertical dimension) on the traffic flows (horizontal dimension): the velocities decrease as the flows increase. C: The dependencies of the flows (vertical dimensions) on the velocities (horizontal dimension): maximal flows can be achieved for certain velocities and decrease for both lower and higher velocities (Andrienko et al. 2015a).

Historical traffic data can be used not only for predicting future movements under various conditions but also for spatial planning applications. For example, the system SmartAdP (Liu et al. 2017a) uses interactive visual tools to find suitable locations for billboard placement using taxi trajectories.

The task of transportation scheduling is addressed by (Andrienko, 2008a). An example application is planning of evacuation of different groups of people, such as general population, schoolchildren, and hospital patients, from a disaster-affected area. The proposed system consists of a scheduling algorithm and a set of visual displays and interactive tools for exploring scheduling outcomes. The displays al- low the user to detect problems, such as delays, understand their reasons, and find appropriate corrective measures.

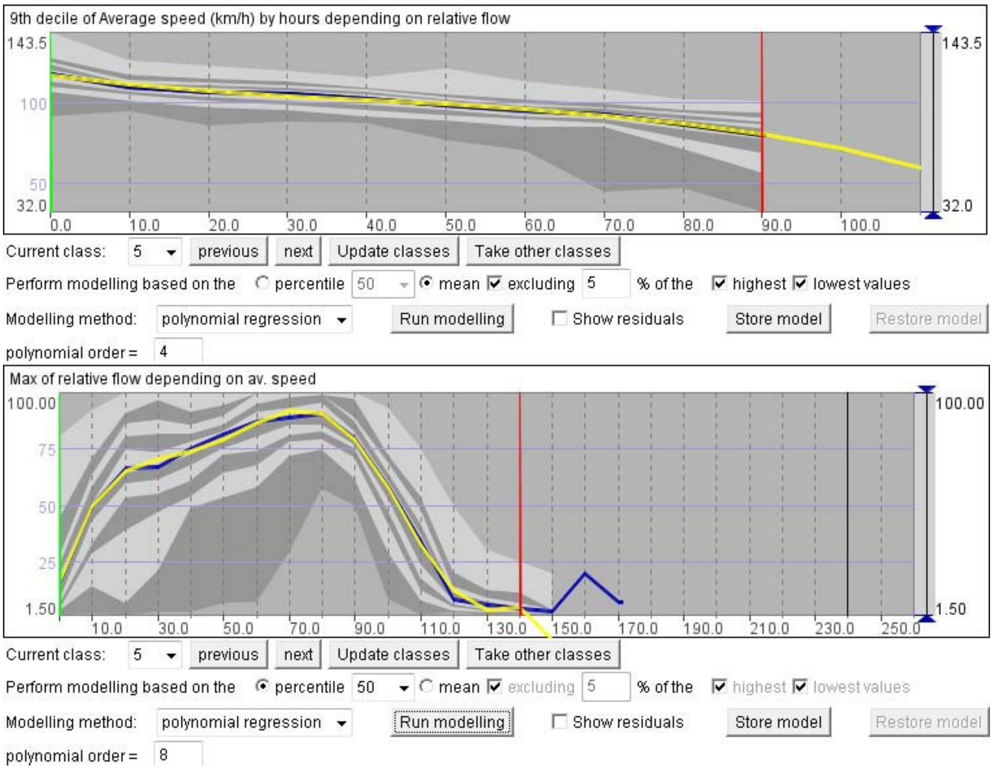

Figure 4: For one of the clusters of links of an abstracted transportation network, the dependencies flow velocity (top) and velocity flow (bottom) are being represented by polynomial regression models (Andrienko & Andrienko, 2013b).

1.2 Assessing data quality

Abstracting from the various specific technologies for collecting movement data, we identify several major methods of position recording (Andrienko,2008c): (i) Location-based, (ii) Time-based, (iii) Change-based, (iv) Event-based and (v) Combinations of these basic approaches. In particular, GPS tracking devices may combine time-based and change-based recording: the positions may be measured at regular time intervals, but recorded only when significant changes of position, speed or direction occur.

As in (Andrienko et al. 2008c) two classes of properties are distinguished:

1. Data structure related properties
   - Mover set properties
   - Spatial properties
   - Temporal properties
2. Data collection procedure related properties
   - position exactness,
   - positioning accuracy,
   - missing positions and
   - meanings of the position absence

Data quality problems are classified using their kind, the data to which the problem applies, the extent of the problem in the dataset and the extent of the problem in the respective value domain. A problem labeling system is introduced in Andrienko et al. (2016) and presented here using a slightly adapted (by adding punctuation) notation.

Problem kind: **M** = missing data, **A** = accuracy problem, P = precision deficiency

Data components: **Mv** = mover identification, **S** = spatial position, **T** = tem- poral reference, **At** = thematic attributes

Extent in the trajectory: **TrE** = elementary (in some elements), **TrI** = inter- mediate (in particular subsequences), **TrO** = overall (in whole trajectory)

Error occurrence can be formulated using formulas similar to M: TrO ∪ MvO ∪ SO ∪ TO which allows for systematic enumeration of error types and the VA tools required to detect and remove problems. Detailed examples are given in Andrienko et al. (2016).

Each kind of problem mentioned in section 3.3 can have several causes and emanations (forms of occurrence) and hence may require multiple types of VA detection tools. Several techniques to validate the data quality w.r.t. several criteria are listed in Andrienko et al. (2016). Such criteria may be problem or dataset specific and hence not expected in advance (e.g. the positional shift of particular traces in space). An inherent property (see main characteristics) of VA is the ad hoc interactive way of the data handling procedures. Therefore, a generic toolbox needs to be created based on the anticipated use cases and prior experience.

The presented examples show cases where the problem was discovered and identified using visual analytics: in big data this will no longer be possible and new methods need to be found based on criteria involving properties of the trajectory, the set of trajectories and the environment. Problem detection by interactive use of VA tools will be much less probable due to the large data size; hence the importance of prior enumeration of potential errors and the availability of a tool to define dataset specific validation criteria.

1.3 VA in Transportation Science

In Andrienko et al. (2017), the authors observe that the contribution of transportation scientists in VA projects is rather limited although VA typically requires the help of domain experts. *"Unfortunately, such work has been limited in the transportation domain (Ferreira et al. 2013a), (Fredrikson et al. 1999a), even though visual analytics researchers have intensively worked with transportation relevant data and developed a variety of methods and tools that could be useful for transportation domain researchers and practitioners. . . . The consequences of the insufficient communication are two-fold. On the one hand, visual analytics researchers have only limited understanding of the problems, needs, and constraints of the transportation domain, which may decrease the potential utility and usability of the methods they develop. On the other hand, the transportation community has quite limited awareness of what visual analytics can offer.* In the conclusion the authors observe: *"Both the visual analytics community and transportation community has produced a large body of exploratory research work in analyzing transportation-related data. However, the knowledge acquired and methods developed often lack collaboration between the two communities."*

Clearly the problems solved by the VA community do not sufficiently coincide with the problems encountered by the transportation research community. Potential causes are:

- In (big) data the travel/movement purpose is often missing and needs to be de- rived from the movement and environment properties which requires prediction models because traveling individuals have particular beliefs, attitudes and goals defining their behavior. VA tools can be used to describe the resulting travel behavior but delivering predictive models is difficult.
- Data are extremely expensive (and out of reach of researchers and practitioners in the transportation domain). Pre-processed data can be purchased from the data generators/owners (telecom and ICT companies) in some cases but few information is made available about the pre-processing methods (aggregation, pseudonymisation, etc). Few affordable data are available to researchers in charge to answer specific research

questions related to a given area. On the other hand VA tools are used by industrial companies who have data available[2].

This section lists some of the classes of currently open research questions, challenges and opportunities to unite VA with the transportation field needs.

Thereto, we need to keep in mind that travel demand is a derived demand. In most cases, the trip is not a goal by itself but a mean to achieve a goal. In activity-based modelling it is assumed that for a given period of time the goals and intentions of individuals determine their time use and displacements. In this case there is a hidden model (much more complex than a HMM) that controls the observed behaviour and that in many cases cannot be derived from mobility big data. Understanding the behavioural model is a prerequisite for the evaluation of intended travel demand measures (TDM). (Andrienko & Andrienko,2017b) explain the integration of state diagrams describing travel behaviour in VA.

Research Questions

Problems solved by transportation planners can be classified as follows by the size of the set of travellers involved.

1. problems involving a small set of affected individuals:
    a. Operations research (OR) like solutions to particular problems (short term, static environment): the aim is to provide advice to customers of transportation services. It requires streaming data (parking, expected congestion, incident detection, …)
    b. Traffic safety research is interested in hot spots, their physical properties (road infrastructure design) and the usage profile (timing, vehicle properties, movement properties). Incidents are often related to detailed interactions between one actor and the environment (requiring detailed description of local situations) or between multiple actors (requiring detailed description of their behaviour in a short period preceding the (near) accident).
2. problems involving a large set of affected people
    a. short term (one week to one year): problems due to planned sudden local changes in a static environment. Traffic guidance measures can be put in place based on flow prediction models. VA can be used to show the recent (1 hour) history and the near future prediction in order to generate advice to the transportation system operation management.
    b. long term: TDM’s effect prediction (long term, evolving environment, evolving actors): advice to designers of transportation systems (urban planners, regional development managers, public service, ...). As mentioned above, activity-based modelling is used to this end. Following travel generators can be distinguished:
        i. households: members execute frequently recurring activities (for several purposes) as well as exceptional ones. All activities aim to the achievement of a private goal/purpose (e.g. by maximizing utility de- rived from activity execution).
        ii. business/commercial: trips are executed on behalf of or indirectly triggered by households and aiming to achieve of a goal stated by a third party (example: parcel delivery)

Activity based modelling nearly exclusively focuses on travel generated by households. It is important to estimate the number and properties of business/commercial since modelling travel demand induced by companies turns out to be problematic due to lack of data. Big data and VA may contribute to solve the problem by classifying daily movements of vehicles from revealed

[2] e.g. https://www.be-mobile.com/products/flowcheck/.

trajectories w.r.t the distance driven, the number of trips in a tour (number of visited places), clustering visited locations etc. Classes may coincide with taxi-like pattern, sales/service-person like pattern, packet-delivery like pattern. After finding patterns their share in the flow may be estimated. Particular attention is to be paid to bias in the available datasets.

Data availability challenges

1. In order to solve some of the above questions, person traces instead of car traces are required. These may be collected by smartphones but induce more privacy problems than car traces and are more difficult to process (because of trip and mode detection).
2. In order to capture behavioural aspects of travel, longitudinal (as opposed to cross-sectional) data and privacy preserving measures are required. These requirements may be mutually incompatible (see the issue of mover identity errors mentioned earlier.
3. Streaming data may originate from parking use or as floating car data; however, no such data will be available in the project (none are known to be available for research)

RAM problem

VA tools need to be adapted to the use of large datasets because they typically use RAM based datasets. In (Andrienko & and Andrienko,2012a) the authors propose to use a stepwise approach by which data aggregations are created in advance and loaded into RAM for processing by VA tools.

Dedicated procedures need to be designed for data quality assessment (see also section 3.3 because preprocessing by aggregation-based methods may be insufficient

2 Visual Analytics for Complex Event Recognition

Few research works focus specifically on Complex Event Recognition in Visual Analytics: in fact, CER research covers various formal languages and models which are hard to be studied in a same unique framework. However, main ingredients in CER are multivariate time series and event sequences. In these domains, the visualization of data streams has attracted increased interest the last decade. The research work reviewed in the following focus mainly on visualization of multivariate time series and of simple event patterns, without elaborate first-order logic or entanglement of events as CER can deal with. Nevertheless, these techniques are of interest to T&K, providing basic bricks to understand the challenge of CER visualization. The final part of the review is dedicated to dimension reduction techniques allowing to summarize in a few dimensions the original data lying in a high dimensional space. These techniques have been successfully adapted to multivariate time series data and it will be interesting in the T&K project to investigate the hybridation of usual visual analytics techniques and dimension reduction approach, especially in the context of representational learning.

Number of tools have been developed to visualize raw sequences of events. For instance, LifeLines (Plaisant et al. 1996) is dedicated to personal medical histories visualization providing a hierarchical timeline visualization. Events are placed on a time-line accordingly to their occurrences; color encoding and lines with different thickness illustrate the relationship between the events. CloudLines (Krśtajic et al. 2011) use a logarithmic time scale to allow the visualization of recent events together with long term patterns at any time scale. ChronoLense (Zhao et al. 2011) emphasizes on the usage of different lenses (possibly with the user interaction) to navigate through the time series. These techniques are well adapted to explore with precision time series and individual traces but do not aggregate or factorize the event streams and thus do not scale with large event collections or many time series.

Other approaches try to summarize multivariate time series in a high dimensionality setting in order to visualize the data. TimeSeer (Dang et al. 2013) proposes the notion of *scagnostics* to capture the characteristics of the data based on a set of measure (density, skewness, …). Scatterplot matrix and charts are used next to interactively explore the data. ThemeRiver (Havre et al. 2000) is another well know technique to analyze variation and exchange between event streams. RankExplorer (Shi et al. 2012) extends ThemeRiver in a framework for big stream data analysis. In the context of spatio-temporal data analysis, (Tominski et al. 2012) proposes to stack 2D visualization in a 3D representation to analyze trajectories; (Scheepens et al. 2011) uses density fields to capture the multivariate aspect of the time series.

Another way to summarize multivariate time series is to factorize the event streams thanks to a tree structure: LifeFlow (Wongsuphasawat et al. 2011) and EventFlow (Maguire et al. 2013) aggregate time series with respect to common subsequences to provide a tree-like visualization. However, the tree structure grows exponentially with the number of events which hinders the visualization. Graph structures can also be used to analyze time series: Sankey diagram (Riehmann et al. 2005) and derivate techniques (Wongsuphasawat, 2012) are techniques intensively used to analyze the event flow between states of a system. To improve the readability of the Sankey diagram when the graph of states is dense, MatrixFlow (Perer et al. 2012) represents spatio-temporal data with sequences of matrices; MatrixWave (Zhao et al. 2015) improves the use of matrices by apprehending the event flow between the states.

An orthogonal family of approaches is the dimension reduction methods. The objective of this kind of approach is to compress the time series expressed in a high dimensional space into a very few number of dimensions – ideally two – losing the less possible important information. The low dimension projected space can be used to visualize meaningfully the originally high dimensional data. The most used technique, the Principal Component Analysis (PCA) (Candès et al. 2011), which use linear projections to reduce the number of dimensions, has been adapted to temporal data in various works (Yang & Shahabi, 2004) (Liao, 2005). Derived from PCA, Dictionary Learning techniques (Mairal et al. 2009) use combination of weighted atoms to approximate the time series, each atom representing an elementary behavior shared among number of time series. The Non negative Matrix Factorization method (Cichocki et al. 2009) introduces a constraint on the recombination of the atoms, allowing only additive reconstruction. This leads to a better interpretation of the atoms.

Another very active domain in dimension reduction topic is the Representation Learning (Bengio et al. 2013), leaded by the research on deep learning (Lecun et al. 2015). The objective is to learn non linear projection – *embeddings* - of the data in an unsupervised context generally thanks to a deep neural network. A special case is the auto-encoder model, applied successfully to time series (Längkvist et al. 2014). The architecture involves an encoding module and a decoding one. The encoding module projects an input data into a latent space through different neuronal layers with a decreasing number of neurones. The decoder has a mirror architecture, with first layers containing the less neurones and the last layer the same amount as the first layer of the encoding module. The embedding representation in the latent space is learned by a measure of divergence between the input data and the output one: the network learns to reconstruct the original signal with the less distortion possible compressed in a low dimension space. As the layers are connected by non linear functions, the embedding are more expressive than with usual dictionary learning techniques.

Other techniques for dimension reduction use similarity measures to embed the data in a low dimensional space. A widely used approach is the multidimensional scaling (Cox et al. 2001), organizing the data in a 2D space such as the distances in the projected space are close to the distance in the original space. The T-distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten & Hinton, 2008) is privileged for data in high dimension. The algorithm defines two probabilistic models over the distance between each couple of points, the first one in the original space, the second one in the low dimensional representation space. It learns the representation in order to optimize the minimal divergence between the two probability distributions.

3 Cross-scale analysis and dashboards for populations' mobility

3.1 Aggregate visualizations and cross-scale analysis

Aggregate visualizations of trajectory data in mobility analytics can be seen to pertain to two main areas: cross-scale analysis, used to adjust the analysis scale to the scale of the mobility pattern(s) studied; and aggregation of mobility-related parameters and trajectory geometries.

Due to the effect of internal and external factors influencing the movement at different spatial and temporal scales, behaviours may manifest different movement patterns at different scales (Nathan et al. 2008); see Figure 5. Therefore, recently, the importance of scale, and thus of cross-scale analysis has been acknowledged in the literature (Keim et al. 2008; Laube & Purves, 2011; Soleymani et al. 2014), and a number of methods and algorithms capable of investigating the relationships between patterns and processes occurring at multiple spatial and/or temporal scales of movement have been developed. Laube & Purves (2011) systematically explored the effects of computing movement parameters across different temporal scales, and thus demonstrated that adjusting the analysis scale is crucial to obtain meaningful results. Recently, more progress has been made, such as developing new multi-scale measures (e.g., multi-scale straightness index (Postlethwaite et al. 2013)), patterns detection using Brownian bridges in low sampling rate movement data (Buchin et al. 2012), measurement of dynamic interactions in movement (Long & Nelson, 2013), or the use of the discrete wavelet transform for movement classification and trajectory segmentation at different spatio-temporal scales (Soleymani at al. 2017). While relevant methods have been developed for cross-scale analysis in the spatial and temporal domain, it should be noted that the issue of scale also applies to semantic aspects of movement, where research is still very limited. To gain a comprehensive understanding of mobility, cross-scale analysis methods should be developed considering spatial, temporal as well as semantic aspects.

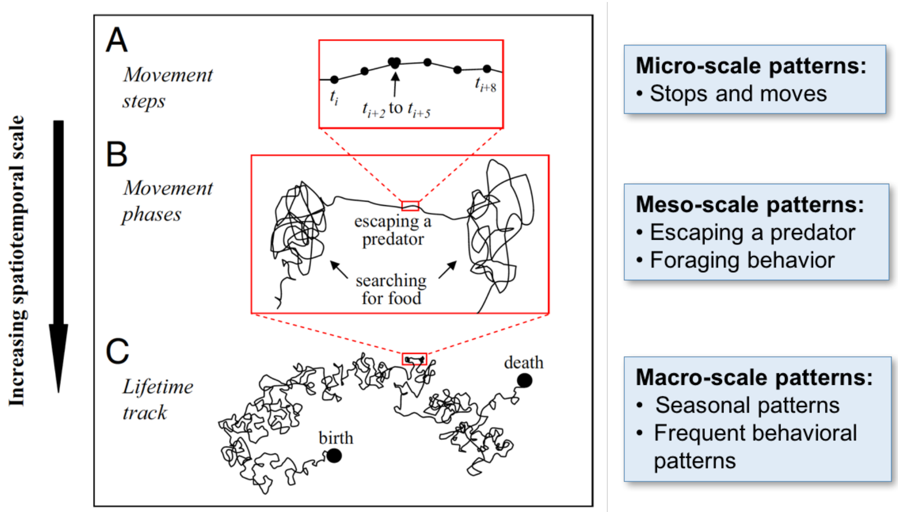

Figure 5: Mobility behaviours may manifest different movement patterns at different scales (i.e., spatial, temporal, or thematic scale). The above shows an example in animal ecology (adapted from Nathan et al. 2008).

Aggregation in support of visualization may affect either the computation of mobility-related parameters or the summarization of the trajectory geometries. Aggregation of mobility-related parameters may take place over different windows of time, different (possibly hierarchical) spatial units, or combination thereof, and may take the form of simple aggregation to advanced data modelling algorithms (Zhang et al. 2012). Andrienko & Andrienko (2008) introduce methods that allow aggregating parameters related to mobility and trajectories in the temporal, spatial and spatio-temporal domains.

Summarizing trajectory geometries typically follows two strands. In the first strand, the geometries (and associated movement parameters) are aggregated to cells of a tessellation. Examples of this approach include Lee et al. (2009), who use a tree data structure for aggregation, as well as Andrienko & Andrienko (2011), where the spatial aggregation units consist of Voronoi cells that capture the essential characteristics of the original trajectories around a subset of 'significant points'. In the second strand of approaches, the geometry of bundles of trajectories is represented by a 'placeholder' that approximates the shape and position of the trajectory bundle. Examples of this group include the TRACLUS clustering algorithm (Lee et al. 2007) and various decedents thereof, algorithms for generating centroid trajectories under positional uncertainty (Pelekis et al. 2011), and algorithms for creating median trajectories (Figure 6; Buchin et al. 2013).

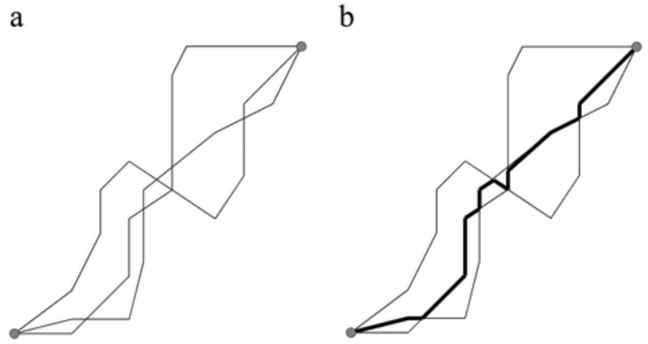

Figure 6: Trajectory aggregation: median trajectory. a) Three trajectories with a common start and end point; b) a median trajectory (bold) representing these three trajectories (Buchin et al. 2013).

Trajectories can also be aggregated from the semantic perspective (Wan *et al.* 2017), besides the geometry-centered methods. Geographical context is one of the essential semantics for trajectories. Semantic enrichment associates special semantics of geographic entities to trajectories by spatial co-location for more complex trajectory modeling (Alvares *et al.* 2007). The geographical entities can be a variety of types, including roads (Yuan *et al.* 2010, Mattheis *et al.* 2014, Quddus and Washington 2015), points-of-interest (POIs, Wu *et al.* 2015a, Arslan *et al.* 2018), land use parcels, and meaningful locations derived from the raw trajectory (Siła-Nowicka *et al.* 2016). As the geographical contexts may also have a semantical hierarchy, e.g., the taxonomy of POIs and land use, semantics-based trajectory aggregation may also happen at different scales, which thus leads to possible comparison analytics across different semantical scales, similarly to the cross-scale analysis from the geometrical perspective.

3.2 Dashboards

Few (2006, p. 26) proposed a commonly accepted definition of a dashboard as "a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so that the information can be monitored at a glance". It often combines text and graphs, with an emphasis on graphics (e.g., maps) to visually present the overview information (Figure 7). Dashboards allow users to explore their data, "not only in terms of spatio-temporal aspects, but also in terms of attribute aspects" (Rahman 2017, p. 1). By nature, dashboards are especially used for providing an overview as they visualize the most essential information at a glance. Very often a dashboard consists of a set of visualizations and controls, allowing interactions such as selection, filtering, and drilling down (Zhang et al., 2012). The use of dashboards has become popular in many fields, such as marketing (Krush et al. 2013), public health (Lechner & Fruhling 2014), construction (Guerriero et al. 2012), urban development (Scipioni et al. 2009), education (Maldonado et al. 2012), and transport monitoring (Rahman 2017).

There are different ways to categorize dashboards. Few (2006) classified them into three groups according to their roles: strategic, analytical and operational. Pappas & Whitman (2011) and Rahman

(2017) provided a detailed comparison of these three groups, in terms of supporting scenarios, timeframe, graph presentation, interactivity, and update frequency.

- Strategic dashboards: They are the most popular type. They provide a quick overview of the data with the notion to enable decision makers to monitor the health and opportunities of the business. Very simple graphs that show what is going on without much interactivity work well for this type of dashboards (Rahman 2017). They don't need real-time data, but still need regular update, e.g., daily, weekly or monthly.
- Analytical dashboards: This group of dashboards often offer greater context than just simple overview in strategic dashboards. They show trends or patterns reflected in the data, and enable further explorations, such as drilling down into the underlying details. Like the previous group, simple graphs work well for analytical dashboards, but with extensive interactivity to allow users to explore the details. They often use historical data.
- Operational dashboards: They represent the most dynamic type compared to the other two groups. They help to monitor situations and act as soon as possible according to particular conditions (Rahman 2017). Highly dynamic graphs such as animated displays work best. They can warn users of outliers or when something goes wrong. Usually, real-time data or near real-time data are required for operational dashboards.

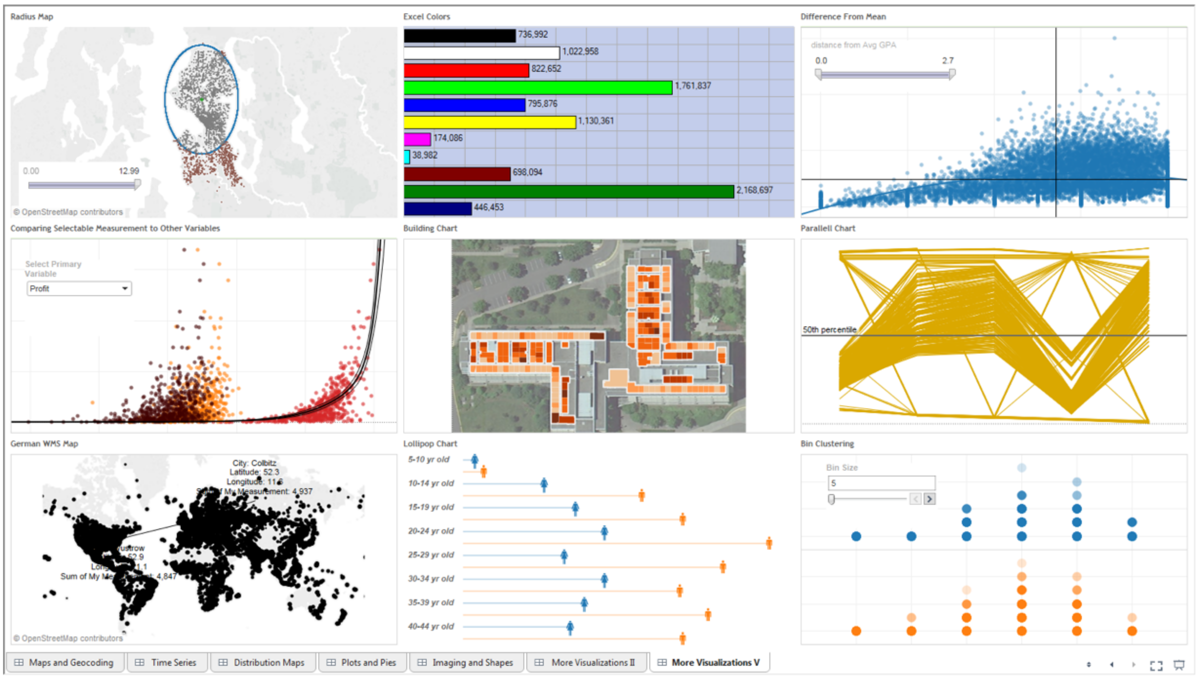

Figure 7: Dashboard examples created in the Tableau software[3]. Different visualization methods can be integrated, such as charts, maps, and even simple texts.

Few (2006) reviewed many existing dashboards, and identified 13 common pitfalls of dashboard design, including: exceeding the boundaries of a single screen, supplying inadequate context for the data, displaying excessive detail, expressing measures indirectly, choosing inappropriate media of display, introducing meaningless variety, using poorly designed display media, encoding quantitative data inaccurately, arranging the data poorly, ineffectively highlighting what's important, cluttering the screen with useless decoration, misusing or overusing color, and designing unattractive visual displays.

To effectively communicate information, a dashboard should be properly designed, particularly using the right visualization (Rahman 2017). Ware (2012) and Few (2006) argued that research on visual

[3] URL: https://qph.ec.quoracdn.net/main-qimg-000650654f2338d4828bb89ab106ffa8.

perception provides empirical evidences on this aspect, such as those on short-term visual memory, visual encoding of rapid perception, and gestalt theory. Few (2006) further proposed two fundamental principles for appropriate visualizations: 1) it must be the best means that is commonly found, 2) it can be still functional even in a small space. Six types of visualizations were proposed: graphs, images, icons, drawing objects, text, and organizers. Pappas & Whitman (2011) further proposed guidance for choosing the right visualizations for dashboards, such as providing interactivity for strategic and analytic dashboards, and allowing comparisons. Zhang et al. (2012), ActiveWizards (2018) and Machlis (2017) all provide detailed reviews of commercial visualization systems that might be used for the implementation of dashboards.

Mobility is one area that stands to benefit from dashboard development. The broader interest or task in analyzing mobility data is understanding or predicting movement and mobility patterns and behaviors (Andrienko *et al.* 2007). Examining how people move within built and natural environments is relevant to a variety of disciplines concerned with the day to day activities that people carry out (e.g., work, shopping, recreation) and many typical use-cases of mobility data visual analytics relate to transportation, including movements of vehicles, bicycles, and pedestrians within routes or transport systems and other interactions with built and natural environments (Menin *et al.* 2019, Nazemi and Burkhardt 2019). It is possible to visualize smaller sets of trajectories as collective individual lines, though large and complex movement data become visually overwhelming and visualization alone is insufficient for making inferences about movement behaviors or patterns (Andrienko *et al.* 2007). While the problem of large data volumes and overwhelming visual displays has been addressed by work in visual analytics, dashboards for mobility analytics require additional consideration in developing at-a-glance views that highlight relevant mobility results. This challenge requires a clear framework for guiding the design and implementation of the mobility-analytics-oriented dashboards. It also needs novel aggregation visualization methods, such as the methods in Section 6.3.1. to support technical solutions.

4 Evaluating visual analytics procedures through eye-tracking

Eye tracking is the process of measuring and recording individual's gaze positions and eye movements. This technology is being increasingly used not only in psychology and product design but also in visualization and human-computer interaction sciences for evaluation of visual displays and user interfaces. Researchers employ eye tracking to understand how their designs are actually used and, possibly, even get insights into users' ways of reasoning and problem solving. In evaluating one design, researchers want to check if users' behaviours correspond to the supposed ways of use, see where the users may have difficulties, and understand how the design can be improved. In evaluating two or more alternative designs, researchers want to know not only which design is better in terms of task completion times and error rates but also why it is better: How does the use of this design differ from the use of the others? What is more difficult, confusing, or inconvenient in the other designs?

Eye tracking produces large amounts of data that are quite hard to analyse. The standard tools and methods for analysis of these data have rather limited capabilities. They can show where the users look first and where they look most and can compute basic measures (fixation count, time to the first fixation, statistics of the fixation durations and saccade lengths, etc.). However, these methods are hardly suitable for studying the spatio-temporal structure of eye scan paths, in particular, how the movements change over time while the user carries out a given task. For these purposes, we adopt movement analysis methods (see Andrienko et al. 2012).

4.1 Eye tracking vs. geographic movement data

Eye tracking data consist of records about the positions and times of gaze fixations. Each record includes the following components: user identifier, time, position in the display space (x- and y-coordinates), and fixation duration. The records may also include other attributes, e.g., stimulus identifier when different stimuli are used in the data collection. The temporally ordered sequence of records of one user referring to one stimulus is further called eye trajectory or scanpath, as in the literature on eye tracking.

Geographical movement data have the same structure: moving object identifier, time, and position (in geographical space) defined by coordinates; additional attributes may also be present. The structural similarity suggests that both classes of data may be analyzed using the same methods. However, there is a significant difference between eye movements and movements of physical objects governed by inertia: eye movements include instantaneous jumps (saccades) over relatively long distances (Dodge et al. 2009). The intermediate points between the start and end positions of a jump are not meaningful; it cannot be assumed that there exists a straight or curved line between two fixation positions such that the eye focus travels along it attending all intermediate points. This prohibits the use of methods involving interpolation between positions, as in creating movement density surfaces (Willems et al. 2009). Hence, not all movement analysis methods are valid for eye trajectories.

Another concern is whether the tasks for which a method was developed are relevant to eye movement analysis. For example, the methods intended to analyze collective simultaneous movements of multiple objects can hardly be useful in analyzing eye trajectories since simultaneous eye movements of two or more users viewing the same image are usually not tracked. Even if such data were collected, the eye foci of different users are unlikely to interact in the screen space similarly to interactions of material moving objects. Hence, not all movement analysis methods are meaningful for eye trajectories.

4.2 Analytical tasks in eye tracking studies

We use the term ‘analytical task’ to denote possible interests of eye movement analysts, i.e., the questions they may seek to answer. The possible types of tasks have been in part extracted and generalized from the eye tracking-related literature and in part generated during the study, when the evaluation group posed their questions and the technology group, from their side, applied the methods to the data and looked what could be learned.

The possible tasks can be divided into two major categories: tasks focusing on areas of interest (AOIs) and tasks focusing on movements. The first category deals with the distribution of the user’s attention over a display. It can be subdivided into several task types according to the following aspects:

- whether the AOIs are predefined (e.g., certain targets the users are supposed to search for) or need to be extracted from the data (e.g., elements/parts of an image attracting more attention);
- whether the evolution of the attention over time is of interest;
- whether the analyst needs general results for the entire set of users or looks for essential differences between individuals or groups (e.g., experts versus novices);
- whether the study is focused on a single display or compares two or more displays.

Common for these tasks is that only the fixations are analyzed and not the saccades or transitions between the AOIs. For example, Çöltekin et al. (2009) compare two interfaces by analyzing fixation durations and fixation counts for predefined AOIs.

In the second task category, the movements are of primary interest. AOIs are important, but the focus is on transitions between them and their temporal order. Analysts want to discover the users’ strategies in visual exploration, search, and performing given tasks. They also want to understand whether and where the users have difficulties. Movement-focused tasks are indispensable in evaluation of information displays and user interfaces. This task category can be subdivided as follows:

- Examine the general characteristics of the movements, e.g., prevalence of long or short movements, presence of sharp turns, path complexity, etc.

- Examine the spatial patterns of the movements, e.g., jumps across large areas or gradual scanning, spatial clustering or dispersion, radial or circular moves, etc.
- Study the relation of the movements to the display content and/or structure, e.g., correspondence to the arrangement of the display elements, movements along available lines or figure boundaries, connections and transitions between the AOIs, etc.
- Understand individual viewing or searching strategies, compare to expected or theoretically optimal strategies.
- Understand general viewing or searching strategies of multiple users, find and interpret different types of activities.
- Find typical paths, e.g., as frequent sequences of attended AOIs.
- Detect and investigate indications of possible users' difficulties: returns to previous points, repeated movements, and cyclic scanning behaviours.

Like the AOI-focused tasks, the movement-focused tasks can be additionally classified according to the following aspects:

- whether the evolution of the eye movements over time is of interest;
- whether different users or groups are compared;
- whether different displays are compared.

Analysis of eye tracking data usually involves many tasks, which may require several analysis methods. The next section briefly reviews the methods that have been previously applied to eye tracking data. It shows that AOI-focused tasks are better supported by the standard methods than movement-focused tasks, which are therefore will be given more attention in our project.

4.3 Methods

There are many statistical metrics that can be derived from eye tracking data. Poole and Ball (2006) systemize these metrics and their possible interpretations. For example, high saccade/fixation ratio indicates more processing, large saccade amplitudes indicate more meaningful cues (as attention is drawn from a distance), etc. However, eye movements cannot be fully understood just from those numbers. Visual analysis is essential for further insight.

The most popular tool to visually analyze eye tracking data is the attention heatmap (Bojko, 2009) showing the distribution of users' attention over the display space. Heatmaps can be easily generated using standard eye tracking software. They can visualize counts of fixations, counts of different users who fixated on different areas, absolute gaze duration, and relative gaze duration (percentage to the total time spent). Attention heatmaps may be useful for AOI-focused tasks. In comparative studies (different time intervals, different users, or different images) several heatmaps are compared. Eye tracking analysts also try to determine users' search strategies by analyzing series of heatmaps generated for consecutive time intervals (Poole & Ball, 2006), which show how the users' attention foci change over time. However, the characteristics of the eye movements, the links between the attention foci, and the paths followed during the search remain unclear.

Another visualization technique provided by standard software is the gaze plot, which represents fixations by circles with sizes proportional to the fixation durations and connects consecutive fixations by lines. Eye movement analysts usually admit that this method is not suitable for large data due to enormous overplotting (Çöltekin et al. 2010).

A common method suitable for movement-focused tasks is scanpath comparison (Duchowski et al. 2010) based on computing the degree of dissimilarity between two scanpaths. The latter are represented as strings where the symbols designate the AOIs and are arranged in the order of attending the AOIs; then a distance function based on string editing is used (Duchowski et al. 2010). The function computes the cost of transforming one string into another by means of deletions, insertions, and substitutions. This can be extended to account for the fixation durations and distances between the AOIs (von der Malsburg & Vasishth, 2011). In analyzing multiple scanpaths, pairwise distances may be averaged (Duchowski et al. 2010) or used to cluster the paths by similarity (Çöltekin et al. 2010). The matrix of pairwise distances can be fed to a projection algorithm, e.g., multidimensional scaling, and the projection can be visualized (Çöltekin et al. 2010) for finding groups of similar scanpaths.

Opach and Nossum (2011) admit that scanpath comparison may be ineffective in case of large variance among eye trajectories. The authors even conclude that the method requires the visual stimuli to be specially designed to minimize the possibilities of different viewing strategies. Thus, this method works well enough in text reading studies (von der Malsburg & Vasishth, 2011) and psychological tests (Duchowski et al. 2010) where the AOIs (words, numbers, letters, etc.) are predefined and supposed to be viewed in a particular order. Çöltekin et al. (2010) represent scanpaths in a generalized way: the possible AOIs are assigned to classes according to their semantics or function; the scanpaths are transformed to sequences of class labels and thereby become more comparable; the analysis is based on these sequences.

The scanpath comparison methodology does not provide a way to see the original scanpaths. The analyst has to deal with the strings, which may be not easy to understand, especially when the symbols represent automatically extracted AOIs and therefore lack semantics. Çöltekin and Kraak (2010) suggest that the space-time cube (STC) (Kraak, 2003) can be used to visualize eye trajectories. It is good for detailed exploration of a single trajectory and even for multiple trajectories when there is not much diversity among them (Çöltekin and Kraak, 2010). Eye trajectories have also been analyzed using the movement summarization method originally developed for geographic data (Fabrikant et al. 2008; Ooms et al. 2012). The successful uses of this method and STC show that geographic movement analysis methods can also be useful in eye movement analysis. Within the project, we are going to perform a systematic investigation of the potential of these and other techniques for eye movement studies.

5 References

ActiveWizards, 2018. A Comparative Analysis of Top 6 BI and Data Visualization Tools in 2018. Available at https://activewizards.com/blog/a-comparative-analysis-of-top-6-bi-and-data-visualization-tools-in-2018/

Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A.F., Moelans, B., Vaisman, A., Fernandes, J.A., Moelans, B., and Vaisman, A. (2007) A model for enriching trajectories with semantic geographical information. In: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems - GIS '07.

Andrienko G, Andrienko N, Dykes J, et al. (2008b) Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. Inf Vis 7(3/4): 173–180.

Andrienko N., G. Andrienko, N. Pelekis, and S. Spaccapietra (2008c). Basic Concepts of Movement Data. In Mobility, Data Mining and Privacy – Geographic Knowledge Discovery, edited by F. Giannotti, D. Pedreschi, 15–38. Heidelberg: Springer Verlag. ch. 1.

Andrienko, G., Andrienko, N. (2008) Spatio-temporal aggregation for visual analysis of movements. In IEEE Symposium on Visual Analytics Science and Technology, pages 51–58.

Andrienko, G., Andrienko, N. and Bartling, U. (2008a) Visual analytics approach to user-controlled evacuation scheduling. Inf. Vis., vol. 7, no. 1, pp. 89–103, 2008.

Andrienko, G., Andrienko, N. Burch, M., and Weiskopf, D. (2012) Visual Analytics Methodology for Eye Movement Studies. IEEE Transactions on Visualization and Computer Graphics (Proceedings IEEE VAST 2012, best paper award), 18(12): 2889-2898, Dec. 2012.

Andrienko, G., Andrienko, N., and Fuchs, G. (2016c) Understanding movement data quality. Journal of Location Based Services, 10(1):31–46, 2016. doi: 10. 1080/17489725.2016.1169322. URL https://doi.org/10.1080/17489725.2016.

Andrienko, G., Andrienko, N., and Wrobel, S. (2007) Visual analytics tools for analysis of movement data. ACM SIGKDD Explorations Newsletter, 9 (2), 38.

Andrienko, G., Andrienko, N., Bak, P., Keim, D. and Wrobel, S. (2013) Visual Analytics of Movement. Berlin, Germany: Springer

Andrienko, G., Andrienko, N., Chen, W., Maciejewski, R., and Zhao, Y. (2017) Visual Analytics of Mobility and Transportation: State of the Art and Further Research Directions. IEEE Transactions on Visualization and Computer Graphics, 24(1):34–44, August 2017. ISSN 1077-2626. doi: 10.1109/TVCG.2017.2744322.

Andrienko, N. and Andrienko, G. (2012) Visual analytics of movement: An overview of methods, tools and procedures. Information Visualization, 12(1):3– 24, 2012. doi: 10.1177/1473871612457601.

Andrienko, N. and Andrienko, G. (2013) A visual analytics framework for spatio-temporal analysis and modelling. Data Mining Knowl. Discovery, vol. 27, no. 1, pp. 55–83.

Andrienko, N. and Andrienko, G. (2017) State Transition Graphs for Semantic Analysis of Movement Behaviours. Information Visualization, 17(1):41–65, 2017. doi: 10.1177/1473871617692841

Andrienko, N., Andrienko, G. (2011) Spatial generalization and aggregation of massive movement data. IEEE Transactions on Visualization and Computer Graphics, 17(2): 205–219.

Andrienko, N., Andrienko, G., and Rinzivillo, S. (2015) Exploiting spatial abstraction in predictive analytics of vehicle traffic. ISPRS Int. J. Geo-Inf., vol. 4, no. 2, pp. 591–606

Andrienko, N., Andrienko, G., and Rinzivillo, S. (2016b) Leveraging spatial abstraction in traffic analysis and forecasting with visual analytics, Inf. Syst., vol. 57, no. 1, pp. 172–194, Apr. 2016.

Andrienko, N., Andrienko, G., Fuchs, G., and Jankowski, P. (2016a) Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. Inf. Vis., vol. 15, no. 2, pp. 117–153, 2016.

Arslan, M., Cruz, C., and Ginhac, D. (2018) Semantic Enrichment of Spatio-temporal Trajectories for Worker Safety on Construction Sites. Procedia Computer Science, 130, 271–278.

Beecham, B. and Wood, J. (2014) Exploring gendered cycling behaviours within a large-scale behavioural data-set. Transp. Planning Technol., vol. 37, no. 1, pp. 83–97, 2014.

Beecham, R., Wood, J., and Bowerman, A. (2014) Studying commuting behaviours using collaborative visual analytics. Comput., Environ. Urban Syst., vol. 47, pp. 5–15, Sep. 2014.

Bengio, Y., Courville, A., & Vincent, P. (2013) Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35(8), 1798-1828.

Bojko, A. (2009) Informative or misleading? Heatmaps deconstructed. In J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2009, LNSC 5610, pp. 30-39.

Buchin, K., Buchin, M., van Kreveld, M., Löffler, M., Silveira, R.I., Wenk, C., Wiratma, L. (2013) Median Trajectories. Algorithmica, 66: 595-614. https://doi.org/10.1007/s00453-012-9654-2

Buchin, K., Sijben, S., Arseneau, T.J.M., Willems, E.P. (2012) Detecting Movement Patterns Using Brownian Bridges, in: Proceedings of the 20th International Conference on Advances in Geographic Information Systems, SIGSPATIAL ’12. ACM, New York, NY, USA, pp. 119–128. https://doi.org/10.1145/2424321.2424338

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011) Robust principal component analysis. Journal of the ACM 58(3), p. 11.

Chae, J., Thom, D., Jang, Y., Kim, S. Y. , Ertl, T. and Ebert, D. S. (2014a) Public behavior response analysis in disaster events utilizing visual analytics of microblog data. Comput. Graph., vol. 38, pp. 51–60, Feb. 2014.

Chu, D. et al. (2014a) Visualizing hidden themes of taxi movement with semantic transformation. in Proc. Pacific Vis. Symp., 2014, pp. 137–144.

Cichocki, A., Zdunek, R., Phan, A. H., & Amari, S. I. (2009) Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation. John Wiley & Sons.

Çöltekin, A., B. Heil, S. Garlandini, S.I. Fabrikant (2009) Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis, Cartography and Geographic Information Science, 36(1): 5-17.

Çöltekin, A., S.I. Fabrikant, M. Lacayo (2010) Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. International Journal on Geographical Information Science, 24(10): 1559-1575.

Cox, T.F. & Cox, M.A.A. (2001). Multidimensional Scaling. Chapman and Hall.

Dang, T N, Anand, A, & Wilkinson, L. (2013) TimeSeer Scagnostics for high-dimensional time series. IEEE Transactions on Visualization and Computer Graphics, 19(3), 470-483.

Dodge, S., R. Weibel, E. Forootan (2009) Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. Computers, Environment and Urban Systems, 33(6): 419-434.

Duchowski, A.T., J. Driver, W. Tan, A. Robbins, B.N. Ramey, S. Jolaoso (2010) Scanpath comparison revisited. In Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA 2010), pp. 219-226.

Fabrikant, S.I., S. Rebich-Hespanha, N. Andrienko, G. Andrienko, D.R. Montello (2008) Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes. The Cartographic Journal, 45 (3), pp. 201-215.

Ferreira, N., Poco, JVo, ., H. T., Freire, J. and Silva, C. T. (2013a) Visual exploration of big spatio-temporal urban data: A study of New York City taxi trips. IEEE Trans. Vis. Comput. Graphics, vol. 19, no. 12, pp. 2149–2158, Dec. 2013.

Few, S., 2006. Information Dashboard Design. Oreilly & Associates Incorporated.

Fredrikson, A., North, C., Plaisant, C., and Shneiderman, B. (1999a) Temporal, geographical and categorical aggregations viewed through coordinated displays: A case study with highway incident data in Proc. Workshop New Paradigms Inf. Vis. Manipulation (NPIVM), New York, NY, USA, 1999, pp. 26–34.

Gazis, D. C. (2002a) Traffic Theory. Boston, MA, USA: Kluwer.

Greenewald, K., Tsiligkaridis, T., & Hero, A. (2013) Kronecker sum decompositions of space-time data. Proceedings of IEEE CAMSAP.

Guerriero, A., Zignale, D., Halin, G., 2012. A Zoomable Location-Based Dashboard for Construction Management, in: Cooperative Design, Visualization, and Engineering, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 207–210. https://doi.org/10.1007/978-3-642-32609-7_29

Havre, S, Hetzler, B, & Nowell, L. (2000) Themeriver: Visualizing theme changes over time. In Proceedings of the IEEE Symposium on Information Visualization.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G. (2008) Visual Analytics: Definition, Process, and Challenges, in: Information Visualization, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 154–175. https://doi.org/10.1007/978-3-540-70956-5_7

Keim, D.A., Kohlhammer, J., Ellis, G., and Mansmann, F. (2010) Eds., Mastering the Information Age—Solving Problems With Visual Analytics. Goslar, Germany: Eurographics Association, 2010.

Kieu, L. M., Bhaskar, A., and Chung, E. (2015) Passenger segmentation using smart card data. IEEE Trans. Intell. Transp. Syst., vol. 16, no. 3, pp. 1537–1548, Jun. 2015.

Kraak, M.-J. (2003) The space-time cube revisited from a geovisualization perspective. In Proceedings of the 21st International Cartographic Conference, Durban, South-Africa, pp. 1988-1995.

Krśtajic, M., Bertini, E., & Keim, D. A. (2011) CloudLines Compact display of event episodes in multiple time-series. IEEE Transactions on Visualization and Computer Graphics, 17(12), 2432-2439.

Kruger, R., Thom, D., and Ertl, T. (2015) Semantic enrichment of movement behavior with foursquare—A visual analytics approach. IEEE Trans. Vis. Comput. Graphics, vol. 21, no. 8, pp. 903–915, Aug. 2015.

Kruger, R., Thom, D., Worner, M., Bosch, H. and Ertl, T. (2013) TrajectoryLenses—A set-based filtering and exploration technique for long-term trajectory data. Comput. Graph. Forum, vol. 32, pp. 451–460, Jun. 2013.

Krush, M.T., Agnihotri, R., Trainor, K.J., Nowlin, E.L. (2013) Enhancing organizational sensemaking: An examination of the interactive effects of sales capabilities and marketing dashboards. Industrial Marketing Management, Business Models – Exploring value drivers and the role of marketing 42, 824–835. https://doi.org/10.1016/j.indmarman.2013.02.017

Laharotte, P. A., Billot, R., Come, E., Oukhellou, L., Nantes, A. and El Faouzi, N. E. (2015) Spatiotemporal analysis of Bluetooth data: Application to a large urban network. IEEE Trans. Intell. Transp. Syst., vol. 16, no. 3, pp. 1439–1448, Jun. 2015.

Längkvist, M., Karlsson, L., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognition Letters, 42, 11-24.

Laube, P., Purves, R. (2011) How fast is a cow? Cross-Scale Analysis of Movement Data. Transactions in GIS 15, 401–418. https://doi.org/10.1111/j.1467-9671.2011.01256.x

Lechner, B., Fruhling, A. (2014) Towards Public Health Dashboard Design Guidelines, in: HCI in Business, Lecture Notes in Computer Science. Springer, Cham, pp. 49–59. https://doi.org/10.1007/978-3-319-07293-7_5

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436.

Lee, D.-W., Baek, S.-H., Bae, H.-Y. (2009) aCN-RB-tree: Update Method for Spatio-Temporal Aggregation of Moving Object Trajectory in Ubiquitous Environment. International Conference on Computational Science and Its Applications (ICCSA '09), 10.1109/ICCSA.2009.30.

Lee, J.-G., Han, J., Whang, K.-Y. (2007) Trajectory Clustering: A Partition-and-Group Framework. ACM SIGMOD International Conference on Management of Data, 593-604. doi:10.1145/1247480.1247546

Li, X., A. Çöltekin, M.-J. Kraak (2010) Visual exploration of eye movement data using the space-time cube. In Proceedings of GIScience 2010, pp. 295-309, Springer.

Liao, T. W. (2005). Clustering of time series data—a survey. Pattern recognition, 38(11), 1857-1874.

Liu, D. et al. (2017) SmartAdP: Visual analytics of large-scale taxi trajectories for selecting billboard locations. IEEE Trans. Vis. Comput. Graphics, vol. 23, no. 1, pp. 1–10, Jan. 2017.

Long, J.A., Nelson, T.A. (2013) Measuring Dynamic Interaction in Movement Data. Transactions in GIS 17, 62–77. https://doi.org/10.1111/j.1467-9671.2012.01353.x

Ma, Y., Lin, T., Cao, Z., Li, C., and Chen, W. (2016) Mobility viewer: An Eulerian approach for studying urban crowd flow. IEEE Trans. Intell. Transp. Syst., vol. 17, no. 9, pp. 2627–2636, Sep. 2016.

Machlis, S. (2017). 22 free tools for data visualization and analysis. Computerworld, https://www.computerworld.com/article/2507728/enterprise-applications/enterprise-applications-22-free-tools-for-data-visualization-and-analysis.html

Maguire, E., Rocca-Serra, P., Sansone, S.-A., Davies, J., & Chen, M. (2013) Visual compression of workflow visualizations with automated detection of macro motifs. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2576–2585.

Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2009). Online dictionary learning for sparse coding. In Proceedings of ICML.

Maldonado, R.M., Kay, J., Yacef, K., Schwendimann, B. (2012) An Interactive Teacher's Dashboard for Monitoring Groups in a Multi-tabletop Learning Environment, in: Intelligent Tutoring Systems, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 482–492. https://doi.org/10.1007/978-3-642-30950-2_62

Mattheis, S., Khaled Al-Zahid, K., Engelmann, B., Hildisch, A., Holder, S., Lazarevych, O., Mohr, D., Sedlmeier, F., and Zinck, R. (2014) Putting the car on the map: A scalable map matching system for the Open Source Community. Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI), P-232, 2109–2119.

Menin, A., Chardonnel, S., Davoine, P.-A., and Nedel, L. (2019) eSTIMe: Towards an All-in-One Geovisualization Environment for Daily Mobility Analysis. In: 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, 39–46.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., Smouse, P.E. (2008) A movement ecology paradigm for unifying organismal movement research. PNAS 105, 19052–19059. https://doi.org/10.1073/pnas.0800375105

Nazemi, K. and Burkhardt, D. (2019) Visual analytical dashboards for comparative analytical tasks - A case study on mobility and transportation. Procedia Computer Science, 149, 138–140.

Ooms, K., G. Andrienko, N. Andrienko, P. De Maeyer, V. Fack (2012) Analysing the spatial dimension of eye movement data using a visual analytic approach. Expert Systems with Applications, v.39 (1): 1324-1332.

Opach, T., A. Nossum. Evaluating the usability of cartographic animations with eye-movement analysis (2011) In Proceedings of the 25th International Cartographic Conference, Paris.

Pappas, L., Whitman, L. (2011) Riding the Technology Wave: Effective Dashboard Data Visualization, in: Human Interface and the Management of Information. Interacting with Information, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 249–258. https://doi.org/10.1007/978-3-642-21793-7_29

Pelekis, N., Kopanakis, I., Kotsifakos, E.E., Frentzos, E., Theodoridis, Y. (2011) Clustering uncertain trajectories. Knowledge and Information Systems, 28: 117-147. https://doi.org/10.1007/s10115-010-0316-x

Perer, A. & Sun, J. (2012) Matrixflow: Temporal network visual analytics to track symptom evolution during disease progression. In Proceedings of AMIA.

Plaisant, C., Milash, B., Rose, A., Widoff, S., & Shneiderman, B. (1996) LifeLines: Visualizing personal histories. Proceedings of CHI.

Poole, A., L.J. Ball (2006) Eye tracking in human-computer interaction and usability research: current status and future prospects. In C. Ghaoui (Ed.), Encyclopedia of Human-Computer Interaction, pp. 211-219, Idea group.

Postlethwaite, C.M., Brown, P., Dennis, T.E. (2013) A new multi-scale measure for analysing animal movement data. Journal of Theoretical Biology 317, 175–185. https://doi.org/10.1016/j.jtbi.2012.10.007

Quddus, M. and Washington, S. (2015) Shortest path and vehicle trajectory aided map-matching for low frequency GPS data. Transportation Research Part C: Emerging Technologies, 55, 328–339.

Rahman, A. (2017) Designing a dashboard as geo-visual exploration tool for origin-destination data (Master thesis). ITC.

Riehmann, P., Hanfler, M., & Froehlich, B. (2005) Interactive sankey diagrams. Proceedings of the IEEE Symposium on Information Visualization.

Scheepens, R., Willems N., & van de Wetering, H. (2011) Composite density maps for multivariate trajectories. IEEE Transactions on Visualization and Computer Graphics, 17(12), 2518-2527.

Scipioni, A., Mazzi, A., Mason, M., Manzardo, A., 2009. The Dashboard of Sustainability to measure the local urban sustainable development: The case study of Padua Municipality. Ecological Indicators 9, 364–380. https://doi.org/10.1016/j.ecolind.2008.05.002

Shi, C, Cui, W, Liu, S, Xu, P, Chen, W, & Qu, H. (2012) RankExplorer: Visualization of ranking changes in large time series data. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2669-2678.

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J.A., Demšar, U., and Fotheringham, A.S. (2016) Analysis of human mobility patterns from GPS trajectories and contextual information. International Journal of Geographical Information Science, 30 (5), 881–906.

Soleymani, A., Pennekamp, F., Dodge, S., Weibel, R. (2017) Characterizing change points and continuous transitions in movement behaviors using wavelet decomposition. Methods in Ecology and Evolution, 8(9): 1113-1123. doi: 10.1111/2041-210X.12755.

Soleymani, A., Pennekamp, F., Petchey, O.L., Weibel, R. (2015) Developing and Integrating Advanced Movement Features Improves Automated Classification of Ciliate Species. PLoS ONE, 10(12): e0145345. DOI: 10.1371/journal.pone.0145345.

Thomas, J. J. and Cook, K. A., Eds. (2005) Illuminating the Path: The Research and Development Agenda for Visual Analytics. Piscataway, NJ, USA: IEEE Press, 2005.

Tominski, C, Schumann, H, Andrienko, G, & Andrienko, N. (2012) Stacking-based visualization of trajectory attribute data. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2565-2574.

van der Hurk, E., Kroon, L., Maro´ti, G., and Vervest, P. (2015) Deduction of passengers’ route choices from smart card data. IEEE Trans. Intell. Transp. Syst., vol. 16, no. 1, pp. 430–440, Feb. 2015.

van der Maaten, L.J.P.; Hinton, G.E. (2008) Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research. 9, 2579–2605.

von der Malsburg, T., S. Vasishth (2011) What is the scanpath signature of syntactic reanalysis? Journal of Memory and Language, 65(2): 109-127.

von Landesberger, T., Brodkorb, F., Roskosch, P., Andrienko, N., Andrienko, G. and Kerren, A. (2016) MobilityGraphs: Visual analysis of mass mobility dynamics via Spatio-temporal graphs and clustering. IEEE Trans. Vis. Comput. Graphics, vol. 22, no. 1, pp. 11–20, Jan. 2016.

Wan, Y., Zhou, C., and Pei, T. (2017) Semantic-Geographic Trajectory Pattern Mining Based on a New

Similarity Measurement. ISPRS International Journal of Geo-Information, 6 (7), 212.

Ware, C. (2012) Information Visualization, Third Edition: Perception for Design, 3 edition. ed. Morgan Kaufmann, Waltham, MA.

Willems, N., H. van de Wetering, J.J. van Wijk (2009) Visualization of vessel movements. Computer Graphics Forum, 28(3): 959-966.

Wongsuphasawat, K. & Gotz, D. (2012) Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2659–2668.

Wongsuphasawat, K., Guerra Gomez, J. A., Plaisant, C., Wang, T. D., Taieb-Maimon, M., & Shneiderman, B. (2011) LifeFlow: Visualizing an overview of event sequences. In Proceedings of SIGCHI.

Wood, Slingsby, J., A., and Dykes, J. (2011) Visualizing the dynamics of Lon- don's bicycle hire scheme. Cartographica, vol. 46, no. 4, pp. 239–251, 2011.

Wu, F., Li, Z., Lee, W.-C., Wang, H., and Huang, Z. (2015) Semantic Annotation of Mobility Data using Social Media. Proceedings of the 24th International Conference on World Wide Web - WWW '15, 1253–1263.

Yang, K. & Shahabi, C. (2004). A PCA-based similarity measure for multivariate time series. In Proceedings of the 2nd ACM international workshop on Multimedia databases.

Yang, X., Zhao, Z., and Lu, S. (2016) Exploring spatial-temporal patterns of urban human mobility hotspots. Sustainability, vol. 8, no. 7, p. 674, 2016.

Yuan, J., Zheng, Y., Zhang, C., Xie, X., and Sun, G.-Z. (2010) An Interactive-Voting Based Map Matching Algorithm. In: 2010 Eleventh International Conference on Mobile Data Management. Kansas City, MO: IEEE, 43–52.

Zhang, L., Stoffel, A., Behrisch, M., Mittelstadt, S., Schreck, T., Pompl, R., Weber, S., Last, H., Keim, D. (2012) Visual analytics for the big data era — A comparative review of state-of-the-art commercial systems, in: 2012 IEEE Conference on Visual Analytics Science and Technology (VAST). pp. 173–182. https://doi.org/10.1109/VAST.2012.6400554

Zhao, J, Chevalier, F, Pietriga, E, & Balakrishnan, R. (2011) Exploratory analysis of time-series with chronolenses. IEEE Transactions on Visualization and Computer Graphics 17(12), 2422-2431.

Zhao, J., Liu, Z., Dontcheva, M., Hertzmann, A., & Wilson, A. (2015) MatrixWave: Visual Comparison of Event Sequence Data. Proceedings of CHI.