Visual Analytics for Understanding Spatial Situations from Episodic Movement Data

Natalia Andrienko, Gennady Andrienko, Hendrik Stange, Thomas Liebig, Dirk Hecker

Fraunhofer Institute IAIS (Intelligent Analysis and Information Systems),
Schloss Birlinghoven, 53754, Sankt Augustin, Germany

Abstract

Continuing advances in modern data acquisition techniques result in rapidly growing amounts of geo-referenced data about moving objects and in emergence of new data types. We define episodic movement data as a new complex data type to be considered in the research fields relevant to data analysis. In episodic movement data, position measurements may be separated by large time gaps, in which the positions of the moving objects are unknown and cannot be reliably reconstructed. Many of the existing methods for movement analysis are designed for data with fine temporal resolution and cannot be applied to discontinuous trajectories. We present an approach utilizing Visual Analytics methods to explore and understand the temporal variation of spatial situations derived from episodic movement data by means of spatio-temporal aggregation. The situations are defined in terms of the presence of moving objects in different places and in terms of flows (collective movements) between the places. The approach, which combines interactive visual displays with clustering of the spatial situations, is presented by example of a real dataset collected by Bluetooth sensors.

Introduction

The popularity of cellular phones and advances in information and sensor technologies lead the way towards new location recording techniques and thus new types of movement data. ‘Episodic movement data’ refers to data about spatial positions of moving objects where the time intervals between the measurements may be quite large and therefore the intermediate positions cannot be reliably reconstructed by means of interpolation, map matching, or other methods. Such data can also be called ‘temporally sparse’; however, this term is not very accurate since the temporal resolution of the data may greatly vary and occasionally be quite fine. There are multiple ways of data collection producing episodic movement data:

- **Location based:** Positions of objects are recorded only when they come into the range of static sensors. The temporal resolution of the collected data depends on the coverage and density of the spatial distribution of the sensors.
- **Activity based:** Positions of objects are recorded only at the times when they perform certain activities, for example, call by mobile phones, pay by credit cards or send posts to a community website.
- **Device based:** Positions are measured and recorded by mobile devices attached to the objects but this cannot be done sufficiently frequently, for example, due to the limited battery lives of the devices i.e. when tracking movements of wild animals.

Irrespective of the collection method we can identify three types of uncertainty. First, the common type of uncertainty in any episodic movement data is the lack of information about the spatial positions of the objects between the recorded positions (continuity), which is caused by large time intervals between the recordings and by missed recordings. Second, a frequently occurring type of uncertainty is imprecision of the recorded positions (accuracy). Thus, a sensor may detect an object within its range but may not be able to determine the exact coordinates of the object. For a mobile phone call, the localization precision may be the range of a certain antenna but not an exact point in space. Due to these two types of uncertainty, episodic movement data cannot be treated as continuous trajectories, i.e., unbroken lines in the spatio-temporal continuum such that some point on the line exists for each time moment. Third, the number of recorded objects (coverage) may also be uncertain due to the usage of a service or due to the utilized sensor technology. For example, one individual may carry two or more devices with Bluetooth transceivers, which will be registered by Bluetooth sensors as independent objects. On the other hand, the sensors only capture devices with activated Bluetooth services. The activation status may change while a device carrier moves from one sensor to another.

Many of the existing visual and data mining methods designed for dealing with movement data are explicitly or implicitly based on the assumption of continuous objects' movement between the measured positions and are therefore not suitable for episodic data. Interpolation is obviously involved in visual representation of trajectories by continuous lines but it is also implicitly involved in computation of movement speeds, directions, and other attributes characterising the movement (these computations also assume that the positions are precise). The same holds for summarisation of movement data in the form of density or vector fields. Mining methods for finding patterns of relative or collective movement of two or more objects (e.g. meeting or flocking) also require fine-resolution data. Since many of the existing methods are not applicable to episodic movement data, there is a need in finding suitable approaches to analysing this kind of data.

Due to the uncertainties, episodic data are usually not suitable for studying the movement behaviours of individual objects. We suggest aggregation of many individual tracks as a way to compensate for missing data and uncertainties in the spatial and temporal coverage.

By example of episodic movement data, this paper motivates the use of visual analytics approaches in analysing complex data. Visual analytics strives at multiplying the analytical power of both humans and computers by finding effective ways to combine interactive visual techniques with algorithms for computational data analysis (Keim et al. 2008). The main role of the visual techniques in this combination is to enable and promote human understanding of the data, which is necessary for choosing appropriate computational methods and steering their work. Thus, visual analytics can help in understanding the data for data mining tasks, such as distributions, features, clusters, patterns.

Visual analytics approaches are applied to data and problems for which there are (yet) no purely automatic methods. By enabling human understanding, reasoning, and use of prior knowledge and experiences, visual analytics can help the analyst to find suitable ways for data analysis and problem solving, which, possibly, can later be fully or partly automated. In this way, visual analytics can drive the development and adaption of learning and mining algorithms.

In the next section, we describe aggregation of episodic movement data and define the analysis tasks in which the aggregated data can be used. Then, after discussing the relevant literature, we present our visual analytics methods and tools using an example of episodic movement data collected by Bluetooth sensors.

Spatio-temporal aggregation

Episodic movement data consist of records including the following components: object identifier o_k, spatial position p_i, time t_i, and, possibly, other attributes. The spatial position may be specified directly by spatial (geographic) coordinates p=(x,y) or p=(x,y,z) or by referring to a sensor or location having a fixed position in space. A chronologically ordered sequence of positions of one moving object can be regarded as a trajectory, which is spatially and temporally discontinuous.

For temporal aggregation of the data time is divided into intervals. Depending on the application and analysis goals, the analyst may consider time as a line (i.e. linearly ordered set of moments) or as a cycle, e.g., daily, weekly, or yearly. Accordingly, time intervals are defined on the line or within the chosen cycle.

For spatial aggregation it is necessary to define a finite set of *places* visited by the moving objects. We can hereby distinguish two different cases:

1. The recorded object positions are limited to a finite set of predefined positions, such as positions of sensors or cells of a mobile phone network.
2. The recorded object positions are arbitrary. This is the case when the positions are received from mobile devices worn by the objects and capable of measuring absolute spatial positions, such as GPS devices.

In the first case the different positions can be directly used as the places for spatial aggregation. In order to analyze the data at a larger spatial scale, the analyst may group neighbouring positions and define places using e.g. convex hulls or spatial buffers or Voronoi polygons around the groups.

In the second case spatial tessellation may give the required set of places (spatial compartments) for aggregation. Often spatial data are aggregated using arbitrary divisions, such as regular grids or administrative districts, which do not respect the spatial distribution of the data. It is more appropriate to define spatial compartments so that they enclose existing clusters of points. However, these clusters may have very different sizes and shapes, which has two disadvantages. First, it is computationally hard to automatically divide a territory into arbitrarily shaped areas enclosing clusters. Second, such areas are likely to differ in their size, and the respective aggregates would be hard to compare to each other. Therefore, we suggest a method that divides a territory into convex polygons of approximately equal size based on the point distribution (Andrienko and Andrienko 2011). The method finds spatial clusters of points that can be enclosed by circles with a user-chosen radius. A concentration of points having a larger size and/or complex shape will be divided into several clusters. The centroids of the clusters are then used as generating points for Voronoi tessellation. The centroids are the points with the minimal average distance to the cluster members. They are usually located inside concentrations of points.

On the basis of the previously defined set of places P, each trajectory is represented by a sequence of visits $v_1, v_2, \ldots, v_n$ of places from P. A visit v_i is a tuple $<o_k, p_i, t_{start}, t_{end}>$, where o_k is the moving object, $p_i \in P$ is a place, t_{start} is the starting time of the visit, and t_{end} is the ending time. Complementarily to this, each trajectory is also represented by a sequence of *moves* $m_1, m_2, \ldots, m_{n-1}$, where a *move* m_i is a tuple $<o_k, p_i, p_{i+1}, t_0, t_{fin}>$ describing the transition from place p_i to place p_{i+1}. Here t_0 is the time moment when the move begins (it equals t_{end} of visit v_i of place p_i) and t_{fin} is the time moment when the move finishes (it equals t_{start} of visit v_{i+1} of place p_{i+1}). Notice that consecutively visited places p_i and p_{i+1} in a discontinuous trajectory are not necessarily neighbours in space.

Having a dual representation of each trajectory, as a sequence of visits and as a sequence of moves, the data can be aggregated in two complementary ways. First, for each place p_i and time interval Δt the set of visits $V(p_i,\Delta t)$ is extracted and the counts of the visits $NV(p_i,\Delta t)$ and different visitors $NVO(p_i,\Delta t)$ are computed:

$$V(p_i,\Delta t) = \{<o_k, p_i, t_{start}, t_{end}> \mid \exists t: t_{start} \leq t \leq t_{end} \text{ and } t \in \Delta t\}$$

$$NV(p_i, \Delta t) = |V(p_i,\Delta t)|$$

$$NVO(p_i, \Delta t) = |\{o_k \mid \exists <o_k, p_i, t_{start}, t_{end}> \in V(p_i,\Delta t)\}|$$

Notice that an object o_k may visit more than one place during the interval Δt. It will be counted in each of the visited places.

If the original data records include additional attributes, various statistics of these attributes can also be computed, such as minimum, maximum, average, median, etc. Hence, each place is characterized by two or more time series of aggregate values: counts of visits NV, counts of visitors NVO, and, possibly, additional statistics by the time intervals.

The second way of aggregation is applied to *links*, i.e., pairs of places $<p_i, p_j>$ such that there is at least one move from p_i to p_j. For each link $<p_i, p_j>$ and time interval Δt the set of moves from p_i to p_j is extracted:

$$M(p_i,p_j,\Delta t)=\{<o_k, p_i, p_j, t_0, t_{fin}> \mid t_{fin} \in \Delta t\}$$

Notice that only the moves that finish within the interval Δt are included. The count of the moves $NM(p_i,p_j,\Delta t)$ and the count of different objects that moved $NMO(p_i,p_j,\Delta t)$ are computed:

$$NM(p_i,p_j,\Delta t) = |M(p_i,p_j,\Delta t)|$$

$$NMO(p_i,p_j,\Delta t) = |\{o_k \mid \exists <o_k, p_i, p_j, t_0, t_{fin}> \in M(p_i,p_j,\Delta t)\}|$$

An object o_k may move through more than one link during the interval Δt. It will be counted for each of the links it passed.

If the original data include additional attributes, it is also possible to compute changes of the attribute values from t_0 to t_{fin}, e.g. as differences or ratios between the values at t_{fin} and t_0, and then aggregate the changes by computing various statistics. Hence, each link is characterized by two or more time series of aggregate values: counts of moves NM, counts of moving objects NMO, and, possibly, additional statistics of attribute changes by the time intervals.

These two ways of aggregation support two classes of analysis tasks:

- Investigation of the presence of moving objects in different places and the temporal variation of the presence. The presence is expressed by the counts of visits and visitors in the places, i.e., NV and NVO, which will be jointly referred to as *presence counts*.
- Investigation of the flows (aggregate movements) of objects between different places and the temporal variation of the flows. The flows are represented by the counts of moves and moving objects for the links, i.e., NM and NMO. These aggregate attributes are often referred to as *flow magnitudes*.

In both classes of tasks, the aggregated data can be viewed in two ways. On the one hand, the data can be viewed as time series associated with the places or links. The analyst can investigate the individual time series or groups of time series (e.g., clusters of similar time series) using existing methods for time series analysis. On the other hand, the data can be viewed as a sequence of *spatial situations* associated with the time intervals. A *spatial situation* is the distribution of the object presence or flows over the whole territory during a time interval:

$$SSP(\Delta t) = \{NV(p_i,\Delta t) \mid p_i \in P\} \text{ or } SSP(\Delta t) = \{NVO(p_i,\Delta t) \mid p_i \in P\};$$

$$SSF(\Delta t) = \{NM(p_i,p_j,\Delta t) \mid p_i \in P, p_j \in P\} \text{ or } SSF(\Delta t) = \{NMO(p_i,p_j,\Delta t) \mid p_i \in P, p_j \in P\}$$

Here $SSP(\Delta t)$ denotes a spatial situation in terms of presence, which will be further called *presence situation*, and $SSF(\Delta t)$ stands for a spatial situation in terms of flows, further referred to as *flow situation*.

The different views on aggregated movement data are illustrated by maps in Figure 1.

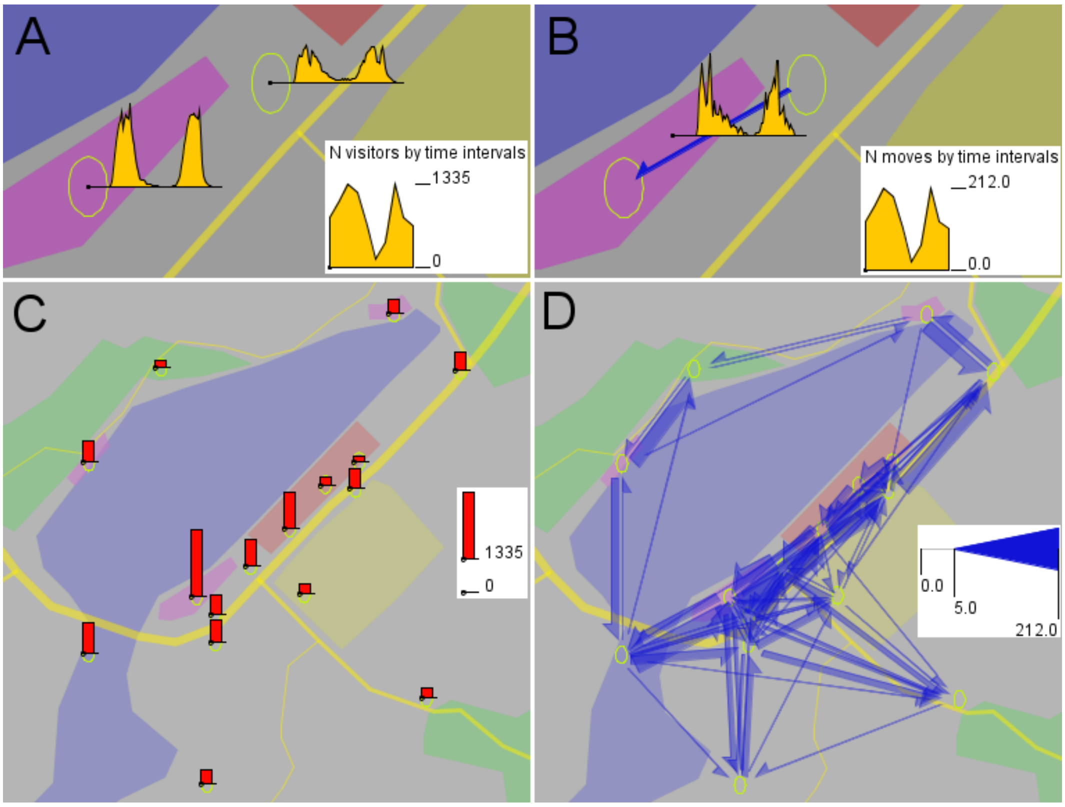


Figure 1. Different views on aggregated movement data. A, B: Time series associated with places (A) and links (B). C, D: spatial situations in terms of presence (C) and flows (D).

In Figure 1A and 1B time series of aggregate values associated with two selected places (A) and with a selected link between two places (B) are represented by diagrams where the horizontal dimension represents time and the height is proportional to the values in different time intervals. The places themselves are represented by ellipses and the link by a special symbol (further referred to as *flow symbol*) in form of a half of an arrow and pointing in the direction of the movement. Such half-arrow symbols are used in order to represent flows between two places in two opposite directions. In Figure 1C and 1D, spatial situations in a selected time interval in terms of presence (C) and flows (D) are shown. The presence counts are shown by proportional heights of the bars drawn in the places and the flow magnitudes by proportional widths of the flow symbols. A map where aggregated movement is shown by flow symbols is called *flow map* (Kraak and Ormeling 2003). Remember that by convention flow symbols (e.g. arrows) represent only counts of items or amounts of goods moving between some places but not the routes of the movement.

In Figure 1D there are many intersections among the flow symbols, which clutter the display. This is a consequence of the discontinuity of the original trajectories, where consecutive recorded positions may be distant in space.

Our research presented in this paper focuses on spatial situations SSP(Δt) and SSF(Δt) and their evolution over time.

Related works

Analysis of movement data is a hot topic in the research areas of machine learning, databases, and visual analytics. However, there are only a few works that address episodic movement data. Due to the complexity of this data type, it is hardly possible to invent a single method enabling full understanding of the data. The existing methods, which are briefly described below, focus on different aspects of the movement and the space where it takes place.

Most of the methods are based on spatial or spatio-temporal aggregation of the data. Aggregates referring to places, such as presence counts, are mainly visualised and analysed in two complementary ways (Jankowski et al. 2010, Wood et al. 2011): animated maps showing the values attained in the places in different time moments and line charts, a.k.a. time graphs, showing the time series in the places by lines. Bak et al. (2009) do not compute presence counts but instead represent each visit of a place by a pixel (tiny rectangle) coloured according to the time of the visit and positioned on a map close to the place location. The arrangement of the pixels produces a kind of place-based aggregation, which is visual rather than computational. Vrotsou et al. (2011) consider aggregated movement data as a weighted directed graph where vertices are the places and arcs are the flows. Various centrality measures are computed for the graph vertices; this can be done by the time intervals. The centrality measures characterize the places in terms of their accessibility, connectivity, links to neighbouring places, etc.

Flows are traditionally visualized on flow maps. Changes of flows over time can be shown on animated flow maps which show one time interval at each moment (Jankowski et al. 2010, Wood et al. 2011). Many works in visualization focus on reducing clutter on flow maps (e.g. Boyandin et al. 2010, Phan et al. 2005) but a good general solution does not exist. Flows are also visualised in the form of origin-destination (OD) matrices (e.g. Guo et al. 2006), which are free from occlusions but lack spatial context. To alleviate the problem of missing spatial context in OD matrices, Wood et al. (2010, 2011) have devised an algorithm to generate so called OD maps, in which multiple OD matrices are arranged according to the geographic positions of the places.

The techniques mentioned above focus on places or links (flows) between them. Andrienko et al. (2011) describe an analysis focussing on discontinuous trajectories constructed from episodic movement data. The trajectories are generalized and then clustered according to the similarity of the routes, i.e., sequences of the visited places.

Investigation of temporal variation of spatial situations is not well supported by existing tools. It is difficult and time consuming to investigate the situations in all time moments one by one or to compare each situation with all others. We suggest an approach based on clustering of spatial situations. Earlier Andrienko et al. (2010) applied clustering to presence situations derived from quasi-continuous trajectories but did not deal with flow situations. Examination and comparison of the clusters was done using static images of individual situations. We advance this technology by summarizing clusters and representing them on interactive maps and enabling computation and visualization of differences between the clusters.

Example dataset

We present our approach by example of a dataset collected using Bluetooth sensors (Bruno and Delmastro 2003). The sensors were installed in 17 places of interest (POIs) in an area where car races take place. The POIs included parking places, entrances to the area and to spectators' tribunes, the information centre, places with shops and restaurants, and other attractions. The data were collected during two consecutive days, Saturday and Sunday. The devices having Bluetooth transceivers, such as mobile phones and digital cameras, which were carried by the visitors coming close to the sensors, were anonymously registered by the sensors. In total, the sensors registered 12,185 different devices and made 792,694 time-stamped records. After cleaning, we have built from these records discontinuous trajectories reflecting the movements of 9,226 device carriers, which is about 15% of the total number of the visitors in these two days. It is reasonable to assume that the people from the sample behaved similarly to the other visitors. The procedure of data collection, cleaning, and preparation for the analysis is described by Stange et al. (2011).

The data have been aggregated spatially by the POIs and temporally by 30-minutes intervals. For the 17 places, there are 276 links.

Analysis of presence

Clustering of spatial situations in different time intervals (i.e., SSP(Δt) or SSF(Δt)) by similarity reduces the workload of the analyst: instead of exploring each situation separately, it is possible to investigate groups of similar situations. Besides, an appropriate visual representation of the clustering results can disclose the patterns of the temporal variation: whether similar spatial situations occur adjacently or closely in time or may be separated by large time gaps, whether the changes between successive intervals are smooth or abrupt, whether the variation is periodic, etc.

For the clustering, the presence situation in each time interval Δt is represented by a feature vector consisting of the presence counts in all places: $\{NV(p_i,\Delta t) \mid p_i \in P\}$ or $\{NVO(p_i,\Delta t) \mid p_i \in P\}$. Any partition-based clustering algorithm can be applied to these feature vectors. In our example, we apply the k-means method from the Weka library (www.cs.waikato.ac.nz/ml/weka/). The method uses the Euclidean distance between feature vectors as the measure of dissimilarity. The results of the clustering are immediately visualized. The centres of the clusters are projected onto a two-dimensional colour space as shown in Figure 2 left; the cluster centres are represented by dots. This is done by means of Sammon's mapping (Sammon 1969). The projection display is used for three purposes: first, for assigning colours to clusters so that close clusters receive similar colours, second, for testing the sensitivity of the clustering results to the parameters of the algorithm (k in our example), and, third, for detecting very close clusters that can be united. Thus, in our example we have tried different values of k from 5 to 20 and found that starting from k=11 increasing the value of k results only in new dots appearing very closely to one or more other dots while the number and relative positions of the dots in the remaining space do not change. Closeness of dots means that the respective clusters do not substantially differ. Hence, we take the result for k=11; however, it contains a concentration of five dots close to each other, i.e., the respective clusters are very similar. Decreasing k does not unite these clusters but decreases the number of the other clusters whose centres are not so close. This means that the clustering algorithm tends to produce more clusters where the data density is higher. To decrease the number of close clusters while preserving the clusters that are less similar, we apply the clustering tool only to the members of the five close clusters and set k to 2. In the result, the five chosen clusters are replaced by two clusters. In total, we have eight sufficiently dissimilar clusters of the presence situations.

The clusters are represented in a summarized way on a multi-map display as shown in Figure 3. For the reason of data confidentiality, we do not use a real cartographic background but show the data on top of a schematic drawing. Each of the small maps represents a cluster; the map caption has the colour of the cluster. To obtain a summary of a cluster, the descriptive statistics of the presence counts for the places (minimum, maximum, sum, mean, median) are computed from all situations included in the cluster. One or more of these statistics can be visualized on the multiple maps. In Figure 3, the mean numbers of place visitors are represented on the maps by proportional heights of the bars.

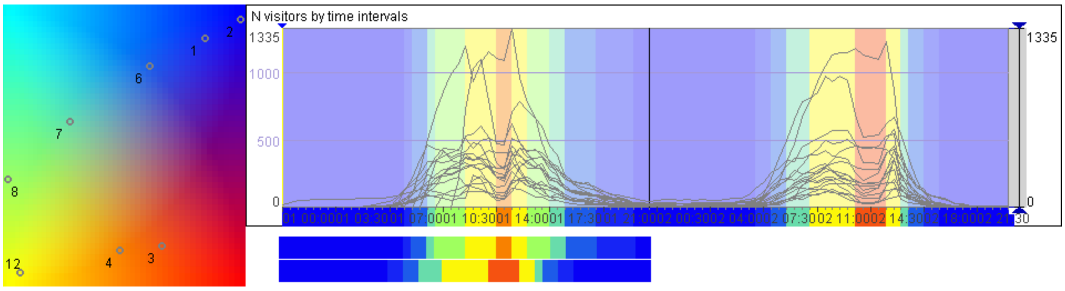


Figure 2. The presence situations in different time intervals have been clustered by similarity. The cluster colours are propagated to the respective time intervals.

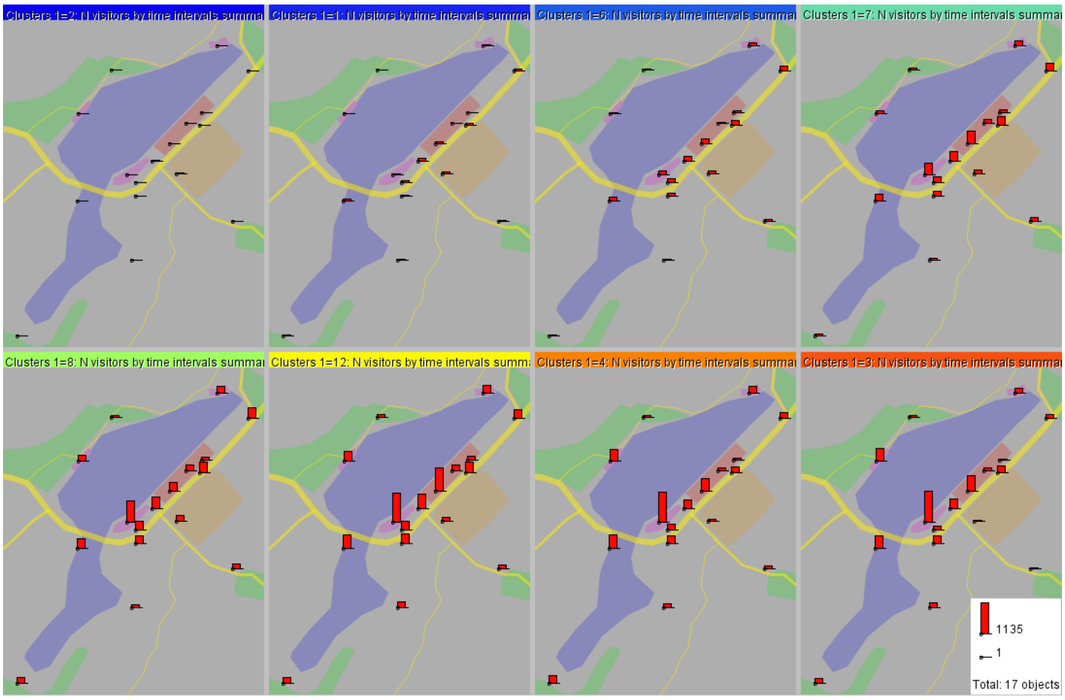

Clusters 1=2: N visitors by time intervals summai
Clusters 1=1: N visitors by time intervals summai
Clusters 1=6: N visitors by time intervals summai
Clusters 1=7: N visitors by time intervals summai
Clusters 1=8: N visitors by time intervals summai
Clusters 1=12: N visitors by time intervals summ:
Clusters 1=4: N visitors by time intervals summai
Clusters 1=3: N visitors by time intervals summai
1135
1
Total: 17 objects

Figure 3. The presence situations are summarized by the clusters. The mean values of the presence are shown by proportional heights of the bars.

The colours of the clusters can be used for colouring time intervals in temporal displays such as a time graph (Figure 2 top right) and a time mosaic (Figure 2 bottom), which can be used for exploring the temporal patterns of the variation of the presence situations. The time graph shows the time series of place visitors. The time axis spans from 0 o'clock of day 1 till the midnight of day 2. The black vertical line separates the two days. The background colouring reveals the patterns of the temporal variation. We see that the time intervals in the nights and early mornings as well as late evening on the second day are in the same cluster. The presence values are close to zero in these times. We also see a similarity between the temporal patterns on the first and second day. The days can be conveniently compared using the time mosaic. Here the time intervals are represented by coloured rectangles arranged in two rows corresponding to the two days. Both displays show us that the time intervals starting from 14:00 and 14:30 on day 1 (i.e., the time from 14:00 till 15:00) and from 13:30, 14:00, 14:30, and 15:00 on day 2 (i.e. the time from 13:30 till 15:30) were quite particular as their colours differ much from the colours of the preceding and following intervals. We checked this finding against the official schedule of the events during the two days and detected that these were the times of the qualifying race on day 1 and the main race on day 2. The presence situations during the two races were quite similar (which is reflected in similar colours) but not enough to be included in the same cluster. The multi-map display shows us that the situations are characterized by very high values of presence in the places from which the races are observable and low values in the other places. The differences between the two groups of places are higher on the second day than on the first day, which is explainable by a higher interest of people in the main race than in the qualifying race.

The situations immediately before and after the races were sufficiently similar to be included in the same cluster. They are characterized by high presence of people at the spectator tribunes but also relatively high attendance of the exhibition and shopping places. The situations after the qualifying race on day 1 changed more gradually than after the main race on day 2.

These examples demonstrate that clustering of presence situations combined with spatial and temporal displays of the clusters can help an analyst to understand the spatio-temporal variation of the presence of people over a territory.

Analysis of flows

The spatio-temporal variation of flows is explored analogously to the variation of the presence except that the clustering and visualization tools are applied to the flow situations instead of the presence situations. The flow situation in each time interval Δt is represented by a feature vector consisting of the flow magnitudes: $\{NM(p_i,p_j,\Delta t) \mid p_i \in P, p_j \in P\}$ or $\{NMO(p_i,p_j,\Delta t) \mid p_i \in P, p_j \in P\}$. We perform the clustering in the same interactive and iterative way as described in the previous section. Finally we obtain 12 clusters adequately representing the differences in the flow situations. Figure 4 shows the projection of the cluster centres onto the colour space and the propagation of the cluster colours to the time graph of the counts of moves and to the time mosaic. Figure 5 shows summaries of the clusters on multiple maps where mean flow magnitudes are represented by proportional widths of the flow symbols. Minor flows (with the magnitude below 10) are hidden for reducing the display clutter.

The flow situations in the night times are characterized by low values of the flow magnitudes. Unlike the presence situations, the flow situations during the times of the qualifying and main races were very similar to the situations in the night, i.e., very few people moved over the area. As one could expect, the situations before and after the races were essentially dissimilar: major flows before the races were directed towards the tribunes and after the races people mostly moved in the opposite directions. However, there are also some unexpected patterns. The collective movement behaviour before the qualifying race (day 1) was more variable than that before the main race (day2): there are three different clusters of flow situations (8, 6, and 5) before the qualifying race and only one cluster (5) before the main race; see the marks above the time mosaic at the bottom of Figure 6. Also the movements after the races are represented by two clusters (10 and 12) on day 1 and one cluster (10) on day 2. To support examining the differences between the clusters, the summary values of flow magnitudes for one selected clusters can be subtracted from the respective values for the other clusters.

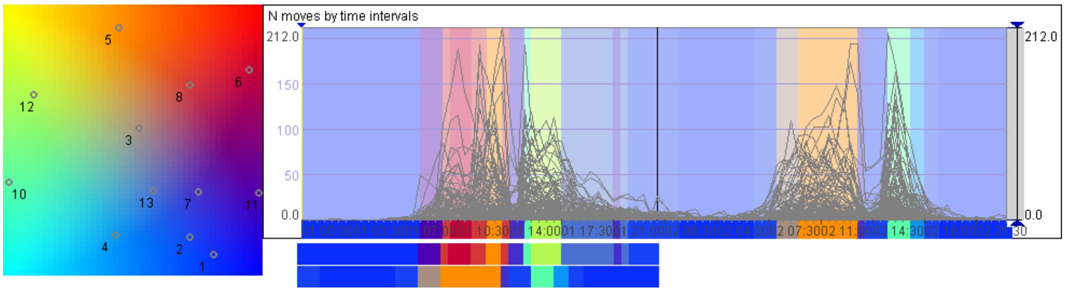


Figure 4. The flow situations in different time intervals have been clustered by similarity. The cluster colours are propagated to the respective time intervals.

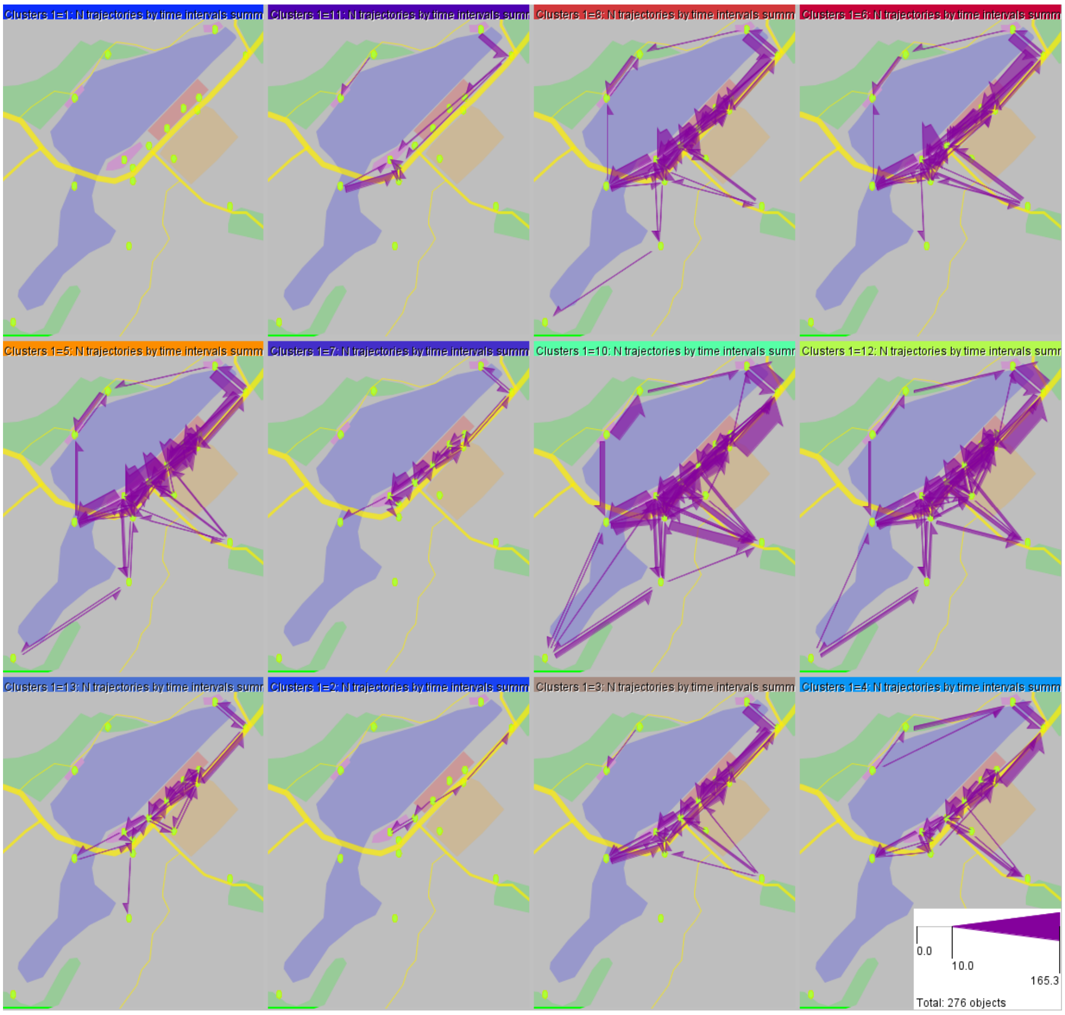

Clusters 1=1: N trajectories by time intervals summ
Clusters 1=11: N trajectories by time intervals summ
Clusters 1=8: N trajectories by time intervals summ
Clusters 1=6: N trajectories by time intervals summ
Clusters 1=5: N trajectories by time intervals summ
Clusters 1=7: N trajectories by time intervals summ
Clusters 1=10: N trajectories by time intervals summ
Clusters 1=12: N trajectories by time intervals summ
Clusters 1=13: N trajectories by time intervals summ
Clusters 1=2: N trajectories by time intervals summ
Clusters 1=3: N trajectories by time intervals summ
Clusters 1=4: N trajectories by time intervals summ
0.0
10.0
165.3
Total: 276 objects

Figure 5. The flow situations have been summarized by the clusters. The mean flow magnitudes are shown.

In Figure 6 A, B clusters 6 and 5 are compared in this way with cluster 8 (the values for cluster 8 are subtracted) and in C cluster 12 is compared to cluster 10 (the values for cluster 10 are subtracted). The differences in the values are shown by flow symbols coloured in two complementary colours: purple is used for positive differences and green for negative. The widths of the symbols are proportional to the absolute values of the differences. Small differences (with the absolute values below 10) are hidden for improving the display legibility.

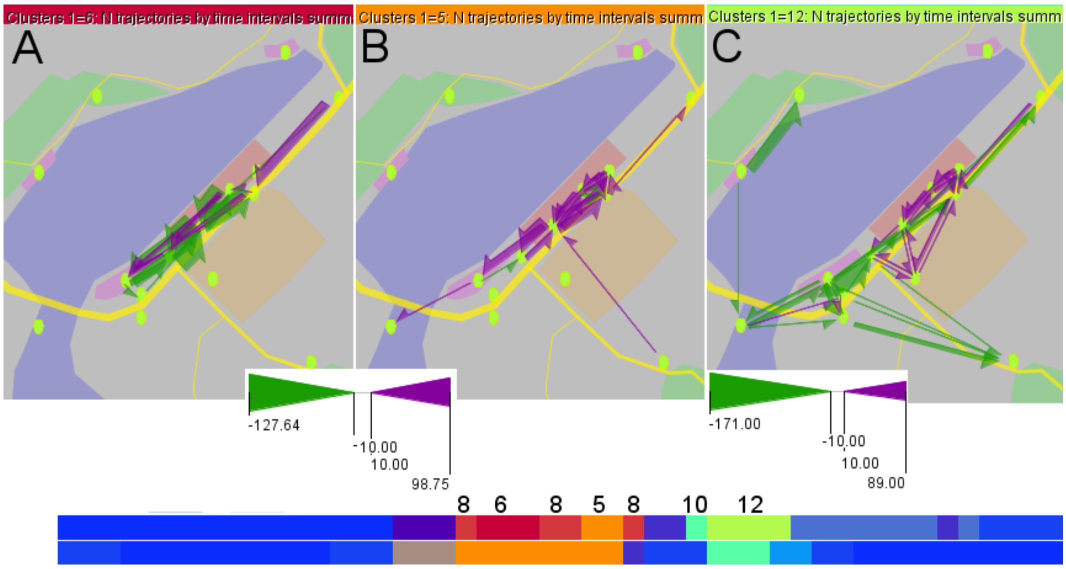

Clusters 1=6: N trajectories by time intervals summ
Clusters 1=5: N trajectories by time intervals summ
Clusters 1=12: N trajectories by time intervals summ
A
B
C
-127.64
-10.00
10.00
98.75
-171.00
-10.00
10.00
89.00
8
6
8
5
8
10
12

Figure 6. Differences between temporal clusters of flow situations. A, B: differences in the movements before the races. Clusters 6 (A) and 5 (B) are compared with cluster 8. C: differences in the movements after the races. Cluster 10 is compared with cluster 12. Bottom: the clusters involved in the comparison are marked on a time mosaic display.

We observe that the flow situations in cluster 6, which took place from 10:00 till 11:30, are distinguished by notably lower magnitudes of short-distance flows between the exhibition/information places and the spectator tribunes than in cluster 8, which took place before 10:00 and after 11:30. At the same time, the magnitudes of the flows from more distant places towards the tribunes were somewhat higher in cluster 6 than in cluster 8. It can be guessed that in the interval 10:00-11:30 people were attracted to the race circuit, probably, by some event. This can explain the decrease of the short-distance flows. The increase of the longer-distance flows can be attributed to people who could have arrived to the area later and immediately moved towards the tribunes to watch the event. Checking against the timetable revealed that it was the practice time. This time is also distinguishable among the clusters of the presence situations (Figure 2). Comparison of clusters 12 and 10 (Figure 6C) tells us that from 15:30 till 17:30 on day 1 (the time of cluster 12) there were more people moving between the places of exhibition, information, and shopping and fewer people moving towards the parking places than in the first half an hour after the qualifying race (day 1) and two hours after the main race (day 2) (these are the times of cluster 10).

Conclusion

The examples show how complex movement data can be analysed by means of visual and computational methods. In fact they demonstrate that collective behaviours of moving objects can be studied and understood by analysing episodic movement data despite the low spatial and temporal coverage and resolution of such data. Aggregation of episodic movement data to some extent compensates for their spatial and temporal sparseness and allows extraction of valuable information about collective movement behaviours of large numbers of moving objects. We argue that this information cannot be extracted by purely computational techniques. Visualisation is essential for enabling human interpretation and involving human knowledge and thinking. However, only visual methods are insufficient due to the large amounts of data (numerous places, combinatorial number of flows, and long time series), their complexity (the spatial, temporal, and attributive components are hard to visualise together in an easily perceivable way), and massive intersections of the flows in space, which makes them extremely hard to visualise in a comprehensible way. Hence, only a combination of visual and computational methods can turn episodic movement data into knowledge.

We combine visual and interactive techniques with computational clustering of spatial situations emerging due to movements of multiple objects. Spatial displays and interactive operations enable comparison of the space-related properties of the clusters of situations. Temporal displays show the arrangement of the clusters in time and enable perception and investigation of the temporal patterns in the variation of the collective movement behaviour. The methodology is applicable to episodic movement data collected in various ways.

References

Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., Wrobel, S. A Conceptual Framework and Taxonomy of Techniques for Analyzing Movement. Journal of Visual Languages and Computing, 2011, v.22 (3), pp.213-232

Andrienko, G., Andrienko, N., Bremm, S., Schreck, T., von Landesberger, T., Bak, P., Keim, D. 2010. Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatiotemporal Patterns. Computer Graphics Forum, 29(3), pp. 913-922.

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

Bak, P., Mansmann, F., Janetzko, H., Keim, D. A. 2009. Spatio-temporal Analysis of Sensor Logs Using Growth-Ring Maps. IEEE Transactions on Visualization and Computer Graphics, 15(6), pp. 913-920.

Boyandin, I., Bertini, E., Lalanne, D. 2010. Visualizing the World's Refugee Data with JFlowMap. In Poster Abstracts at Eurographics/ IEEE-VGTC Symposium on Visualization.

Bruno, R., Delmastro, F. 2003. Design and Analysis of a Bluetooth-based Indoor Localization System. In: Proc. Personal Wireless Communications (PWC), IFIP-TC6 8th International Conference, pp. 711-725.

Guo, D., Chen, J., MacEachren, A., and Liao, K. 2006. A visualization system for space-time and multivariate patterns (VIS-STAMP). IEEE Transactions on Visualization and Computer Graphics, 12(6), pp.1461–1474

Jankowski, P., Andrienko, N., Andrienko, G., Kisilevich, S. 2010. Discovering Landmark Preferences and Movement Patterns from Photo Postings. Transaction in GIS, 2010, v.4 (6), pp.833-852

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G. 2008. Visual Analytics: Definition, Process, and Challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (editors). Information Visualization – Human-Centered Issues and Perspectives. Lecture Notes in Computer Science, Vol. 4950, Springer, Berlin, pp.154-175

Kraak, M.-J., and Ormeling, F. 2003. Cartography: visualization of spatial data. Second edition. Pearson Education Limited, Harlow, UK

Phan, D., Xiao, L., Yeh, R., Hanrahan, P., and Winograd, T. 2005. Flow Map Layout. In Proc. IEEE Symposium on Information Visualization InfoVis 05, Minneapolis, Minnesota, USA, 23-25 October, 2005, pp.219-224

Sammon, J. W. 1969. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers, 18, pp.401–409

Stange, H., Liebig, T., Hecker, D., Andrienko, g., Andrienko, N. 2011. Analytical Workflow of Monitoring Human Mobility in Big Event Settings using Bluetooth. Third International Workshop on Indoor Spatial Awareness ISA 2011, November 1, 2011, Chicago, USA

Vrotsou, K., Andrienko, N., Andrienko, G., Jankowski, P. 2011. Exploring City Structure from Georeferenced Photos Using Graph Centrality Measures. In Proc. Machine Learning and Knowledge Discovery in Databases (PKDD 2011), Lecture Notes in Computer Science, Vol. 6913, pp.654-657

Wood, J., Dykes, J., Slingsby, A. 2010. Visualization of origins, destinations and flows with OD maps. Cartographic Journal, 47(2), pp. 117–129.

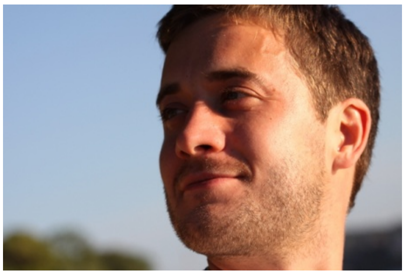
Wood, J., Slingsby, A., Dykes, J. 2011. Visualizing the Dynamics of London's Bicycle Hire Scheme. Cartographica, 46(4), pp.239-251.

Biographies

Natalia Andrienko received her Master degree in Computer Science from Kiev State University in 1985 and PhD equivalent from Moscow State University in 1993. Since 1997, she has been working at GMD, now Fraunhofer IAIS. Since 2007, she is a lead scientist responsible for the visual analytics research.

Gennady Andrienko received his master degree in Computer Science from Kiev State University in 1986 and PhD equivalent from Moscow State University in 1992. Since 1997, he has been working at GMD, now Fraunhofer IAIS, where he is a lead scientist responsible for the visual analytics research.

Thomas Liebig, research scientist at Fraunhofer IAIS and university of Bonn, was born in 1980 and received his diploma in computer sciences from university of technology, Chemnitz, in 2007. Current research focuses on annual average daily traffic (AADT) estimation, Bluetooth tracking and models for correlations within trajectories.



Hendrik Stange is a project manager at the Knowledge Discovery department of the Fraunhofer IAIS since 2007. He studied Business Information Science specializing on data mining and knowledge management at the Otto-von-Guericke University Magdeburg. His current research interests focus on real-time mobility mining, spatial business intelligence and mobile communications.

Dirk Hecker is the leader of the Mobility Mining Group at Fraunhofer IAIS. He studied Geography at the University of Cologne and received his diploma in 2004. Dirk Hecker has been principal investigator in several recent industry funded projects on spatial learning. Current research focusses on real time traffic estimation, mobile analytics and micro-simulation.