Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Analyzing Temporal Usage Patterns of Street Segments Based on GPS Data – A Case Study in Switzerland

Jutta Schreinemacher
Fachverband
Außenwerbung e.V.
Franklinstraße 62
Frankfurt, Germany
schreinemacher@
faw-ev.de

Christine Körner
Fraunhofer IAIS
Schloss Birlinghoven
St. Augustin, Germany
christine.koerner@
iais.fraunhofer.de

Dirk Hecker
Fraunhofer IAIS
Schloss Birlinghoven
St. Augustin, Germany
dirk.hecker@
iais.fraunhofer.de

Georg Bareth
University of Cologne
Albertus-Magnus-Platz
Cologne, Germany
g.bareth@uni-koeln.de

Abstract

Mobility has become a key component of our social and economic activities. Depending on our activities, the usage of the road network varies, showing, for example, an increased traffic load in early morning when people go to work. Such usage patterns are of interest not only for traffic management but also for private companies offering location-based services. In this paper we analyze the temporal usage of street segments based on a large GPS data set in Switzerland. We first conduct a clustering analysis which detects groups of segments with similar temporal usage patterns. Afterwards we analyze the patterns with respect to their temporal and spatial characteristics. Our analysis shows that GPS data is qualified for an analysis of temporal usage patterns, identifying shopping and leisure activities.

Keywords: temporal traffic distribution, clustering, street segments, GPS, traffic counts

1 Introduction

The technological progress during the past decade leads to a more and more digitized world. Navigation systems direct our way using digital maps, we listen to podcasts and read electronic books. One further industry sector which is about to experience a revolution by digital media is out-of-home advertisement [13]. Already today digital posters have found their way into e.g. train stations or airports. They allow not only to tailor advertising content to characteristics of the spatial environment but also to the temporal environment of a poster site. For example, commuters passing a poster site during morning rush-hour traffic are likely to be interested in a different content than passers-by in the late morning who are on a shopping tour. In order to place advertisements at the right moment in time and space, statistical knowledge about passers-by and their intentions are required. Over the past years the outdoor advertising industry in Switzerland has conducted comprehensive GPS surveys in order to enable spatially differentiated poster performance measurements of classic (i.e. static) poster sites [9]. We explore in this paper whether the data allows also a differentiation according to the temporal usage of road segments. More precisely, our goal is to determine groups of street segments which possess a similar traffic load over time. In a second step we interpret these patterns considering their temporal as well as spatial characteristics.

The temporal analysis of traffic load is one important task of federal authorities in order to monitor traffic and to steer the development of the road network. For example, the traffic authorities of Switzerland (ASTRA[1]) and Germany (BASt[2]) commission regular traffic count surveys. This data has a very high temporal resolution. However, the coverage of street segments is low. In addition, the surveys include only major roads and allow no inference in, for example, residential areas. Furthermore, the data collection considers only vehicular traffic. Pedestrians, who contribute the majority of passages in the city center, are not available.

In this paper we therefore explore the usability of GPS data for a temporal analysis of road usage. We apply a clustering algorithm to identify groups of street segments which possess a similar traffic distribution over the week and weekday. Subsequently, we interpret the resulting clusters according to their temporal pattern as well as their geographic location.

Our paper is organized as follows. Section 2 discusses related work. Section 3 describes the GPS data sample. Section 4 describes the analysis process containing data aggregation, clustering as well as the temporal and spatial interpretation of the clusters. We discuss our results in Section 5 and conclude the paper with a summary and outlook on future work.

2 Related Work

The analysis of trajectory data is a very active research area and has developed a number of algorithms for the clustering of (parts of) trajectories [6,11,12] as well as various definitions of distance functions between trajectories [10]. Our approach differs from this work as we do not concentrate on trajectories as principal objects of interest. Instead, we evaluate the frequency and temporal distribution of people passing a specified set of locations, in our case street segments. Typically, such information is collected by traffic authorities directly on the level of street segments.

The German traffic authority BASt publishes in regular intervals statistics about traffic in Germany [3]. Similar to our analysis they analyze and cluster time variation curves using a daily, weekly or yearly resolution of time. However, in

[1] http://www.astra.admin.ch
[2] http://www.bast.de

difference to our analysis they use traffic count data. As already mentioned in the introduction, such data is available only for selected major roads and only for vehicles. In addition, the analysis considers absolute traffic counts, while we are interested to cluster street segments with a similar *distribution* of the traffic load. Moreover, we perform a visual analysis of the spatial distribution in order to interpret the usage patterns.

A very detailed analysis of traffic distributions has been conducted in [7]. In addition to the analysis of weekly patterns also seasonal variations and variations in weather factors were analyzed. However, the basis for this analysis was data from induction loops, i.e. continuous observations of vehicular traffic for a selection of major road segments.

Extensive analysis of GPS data using methods from visual analytics have been conducted in [1]. The authors formed spatial aggregation units based on a grid and provided visualization techniques for the analysis of temporal variation in speed and direction. The variations were visualized using geographically aligned mosaic diagrams, which indicate for each day of the week and hour of the day the intensity of the analyzed variables. However, the exploration of grid cells with similar usage patterns was left to the visual capabilities of the user. In contrast, our data set contains usage patterns for several thousand street segments, which cannot be analyzed by visual inspection only. We therefore perform a cluster analysis, leaving only the interpretation of the results to the user.

3 Description of Data Source

3.1 Swiss Mobility Survey and Selection of Study Areas

GPS technology has established itself as a new standard in Switzerland for the differentiated evaluation of poster performance according to spatial criteria [9]. Beginning in 2003 the neutral research organization Swiss Poster Research Plus (SPR+)[3] conducted systematic GPS studies in the largest metropolitan areas in Switzerland as well as in a number of smaller conurbations (see Figure 1 left). The participants were equipped with GPS devices for a period between 7-10 days. To date the survey includes more than 10,000 participants which form a representative sample for about two thirds of the Swiss population.

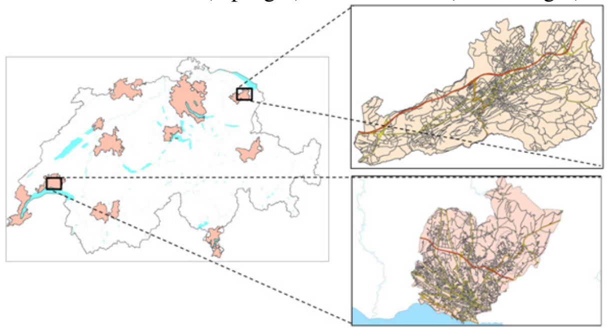
For our analysis we selected two study areas, namely the city of Sankt Gallen and the city of Lausanne, as shown in Figure 1. Both areas possess a high number of GPS test persons compared to the number of inhabitants and street segments. Table 1 provides an overview of the size of both study areas with respect to the number of inhabitants, test persons and size of the street network.

Table 1: Statistics of study areas

City	Inhabitants	Test Persons	Street Segments
Sankt Gallen	72,462	713	4,459
Lausanne	125,885	1,110	7,040

[3] http://www.spr-plus.ch/

Figure 1: Regions of Swiss mobility survey and selected study areas Sankt Gallen (top right) and Lausanne (bottom right)



3.2 Data Subsample

Our temporal analysis shall characterize typical movements on street segments over a time period of one week and a resolution of one hour. In order to obtain representative measurements over one week it is important that each test person provides a high number of measurement days. For example, if a person records movements only on Monday and Wednesday her trajectories do not support the analysis of weekends. However, as movements have a repetitive character [4,14] it can be expected that the person passes similar street segments on the weekend as well (e.g. in her neighborhood). Thus a small number of measurement days increases the risk that we underestimate the number of passages of street segments and thus introduce a bias into our analysis. Similarly, in order to provide a high temporal resolution it is important that each street segment possesses many passages. However, as the introduction of a lower bound for the number of valid measurement days per test person and the number of passages per street segment reduces the size of the data set, we have to find a trade-off between data quality and remaining data size.

With respect to the number of valid measurement days we decided to introduce a lower bound of 5 days per person. This leads to a reduced data set of 668 test persons in Sankt Gallen and 978 test persons in Lausanne. With respect to the number of passages we decided on a threshold of 20 passages per street segment. The threshold results in a remaining number of 2,672 segments (59.9%) in Sankt Gallen and 4,138 segments (58.8%) in Lausanne for our analysis. Table 2 summarizes the statistics of the data subsample.

Table 2: Statistics of subsample

City	Test Persons	Street Segments
Sankt Gallen	668 (93.7%)	2,672 (59.9%)
Lausanne	978 (88.1%)	4,138 (58.8%)

4 Analysis

4.1 Data Preprocessing

In order to find street segments with a similar development of traffic load over time, the data has to be aggregated and normalized.

The level of aggregation determines which details are still visible in the data. However, a too fine-grained level will result in sparse data. In addition, it will decrease the power of our clustering due to a high number of attributes, also known as curse of dimensionality [8]. The goal of our analysis is to find street segments with a similar course of traffic load over the day as well as over the week. As traffic conditions can change quite fast over the day, we decided to use the hour as daily aggregation unit. Considering traffic behavior over the days of week, however, allows for aggregation as humans are known for their repetitive behavior over time. Typical mobility studies about the traffic load as, e.g., [3] show that traffic during working days differs substantially from traffic during the weekend. Also Saturday and Sunday differ from each other, which is plausible due to the closure of shops on Sundays (and holidays) in most European countries. Between working days differences exist mostly due to commuting behavior before and after the weekend. The authors of [3], for example, therefore aggregate only the working days Tuesday – Thursday and keep separate records for Monday and Friday. However, as the differences are small and concern mostly the major daily traffic peak, we decided not to divide working days further. In summary, we formed the following groups of weekdays for our aggregation:

- Monday – Friday,
- Saturday,
- Sunday and holidays.

Having selected the units of aggregation, we can now formalize the aggregation of passages. Let $h \in \{0, 1, \ldots, 23\}$ index the hours of a day, let $d \in \{1, 2, \ldots, 7\}$ index the days of week, let $g \in \{1, 2, 3\}$ index the groups of weekdays and let $n_{i,d,h}$ denote the number of passages on street segment i on day d at hour h. The aggregation $n_{i,g,h}$ for street segment i over our chosen units of time is then defined as

$$n_{i,g,h} = \begin{cases} \sum_{d=1}^{5} n_{i,d,h}/5 & \text{for } g = 1 \\ n_{i,d=g+4,h} & \text{for } g \in \{2, 3\} \end{cases}. \quad (1)$$

We further normalize the aggregated passages because our analysis focuses on a similar development of the traffic load independent of the actual height of traffic. The normalization takes place per street segment and is defined as

$$n^*_{i,g,h} = \frac{n_{i,g,h}}{\sum_g \sum_h n_{i,g,h}}. \quad (2)$$

4.2 Clustering

We applied clustering, an unsupervised learning method, in order to find groups of street segments with a similar course of traffic load over time. Each street segment is described by a set of 72 attributes which each contain the normalized number of passages in the respective aggregation unit (3 groups of weekdays, 24 hours per weekday group).

We tested three different types of clustering algorithms, namely hierarchical clustering, density-based clustering (DBSCAN [2]) and partitioning clustering (k-means) using the Weka toolkit [5]. We obtained the best results with k-means clustering. The hierarchical clustering posed the problem that the criterion for the final selection of the number of clusters was ambiguous. With DBSCAN the parameter selection turned out to be difficult. We either obtained one single cluster or a very high number of small clusters without clear interpretation. For k-means we varied the number of clusters between 2 and 8 and selected a size of k = 3 clusters for Sankt Gallen and of k = 5 clusters for Lausanne. This selection was based on the development of variance within and between clusters for different k, a comparison of the cluster means and a visual inspection of the resulting clusters in their spatial context. Table 3 shows the resulting cluster sizes.

Table 3: Size of resulting clusters

Cluster-Id	Sankt Gallen	Lausanne
1	2,005	31
2	528	358
3	139	376
4		105
5		3,268

Both clusterings contain one cluster with a very large number of street segments. The size of these clusters decreased only marginally when increasing the number of clusters k. In addition, the clustering of Lausanne contains one very small cluster with only 31 street segments, which remained when decreasing the number of clusters k. A closer inspection of the cluster did not yield interpretable results. However, only a small number of persons contributes to the passages within the cluster. It is therefore likely that the cluster reflects unreliable measurements, and we did not consider it further in our analyses.

4.3 Temporal Cluster Interpretation

Figure 2 and Figure 3 show the cluster means of the clusterings as they develop over time. Remember that the first third of the x-axis corresponds to the average traffic distribution on working days, the second third corresponds to the distribution on Saturdays and the last third to the distribution on Sundays and holidays. Both cities show typical movement patterns on working days, which are also very similar across the different clusters. The working days are characterized by 3 peaks which occur at the index hours 7, 13/14 and 17. These peaks stand clearly for the travel to and from work in the early morning and late afternoon as well as for movement during lunch break or the return trip of part-time workers. One further similarity between both figures is the comparably even curve of Cluster 1 in Sankt Gallen and Cluster 5 in Lausanne. Both clusters contain the majority of segments and possess a high inner-cluster variation. The smoother shape thus results in part from averaging over large numbers. The clusters clearly differ from each other when considering the course of traffic load on the weekend. For Sankt Gallen Cluster 2 shows a predominant activity on Saturday between the index hours 10 and 16, which resembles closely the opening hours of shops in the city center (9.00-17.00 o'clock). Cluster 3 has its major activity on Sundays

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

and holidays with a peak at index hour 15. This suggests places for excursions and recreational activities.

Figure 2: Cluster means Sankt Gallen showing the course of traffic load over the temporal aggregation units

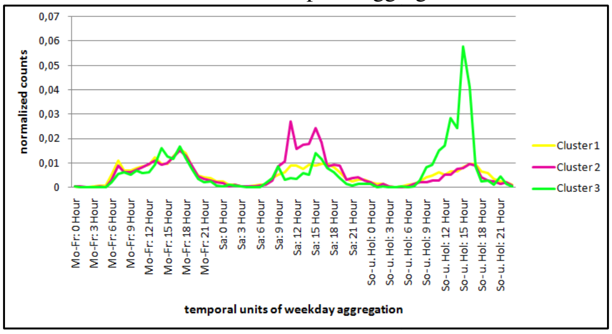
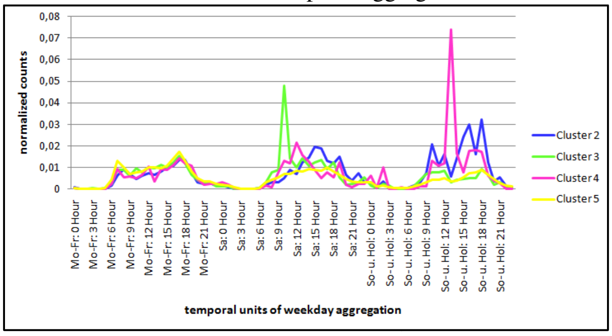


Figure 3: Cluster means Lausanne showing the course of traffic load over the temporal aggregation units



For Lausanne activity on Saturday is shown by Clusters 2, 3 and 4, however, with a shift in their peaks. Cluster 3 shows the highest peak at index hour 10 and remains at a raised level of activity until early evening. Again, this activity corresponds to the opening hours of shops in Lausanne (8.00-18.00 o'clock). Cluster 4 has its peak on Saturday at index hour 12. Afterwards the activity decreases and has two smaller peaks at early evening and Sunday early morning hours. The characteristic peak of Cluster 4 lies on Sunday at index hour 13, however, is accompanied by a raised level of activity between index hour 10 and 18. The time periods on Saturday as well as Sunday correspond to hours of leisure activities, marking especially the times for lunch and dinner. Finally, Cluster 2 shows high activity during Saturday afternoon as well as Sunday morning and afternoon, which also indicates leisure activities.

4.4 Spatial Cluster Interpretation

In this section we consider the spatial distribution of the street segment clusters in order to gain a better understanding of the road usage clusters and to verify assumed activities connected to the patterns.

Figure 4 shows a map of Sankt Gallen where we colored the street segments according to their cluster id. Segments of Cluster 1 are widely distributed over the city. They contain highways, access roads as well as roads in typical residential areas. For all these types of roads travel activities to and from work are characteristic, while during the weekend an unspecific usage over time is plausible. The segments of Cluster 2 reside predominately in the city center. Figure 5 shows this part of the map in more detail. More precisely the segments are close to the abbey church near the historic city center and the pedestrian area, the train station as well as in shopping quarters west of the city wall. Thus, our interpretation that Cluster 2 represents streets with a characteristic usage for shopping is correct.

Figure 4: Clustering Sankt Gallen

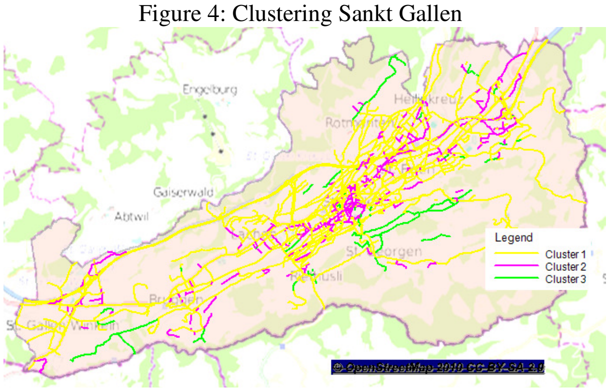


Figure 5: Sankt Gallen city center with shopping facilities

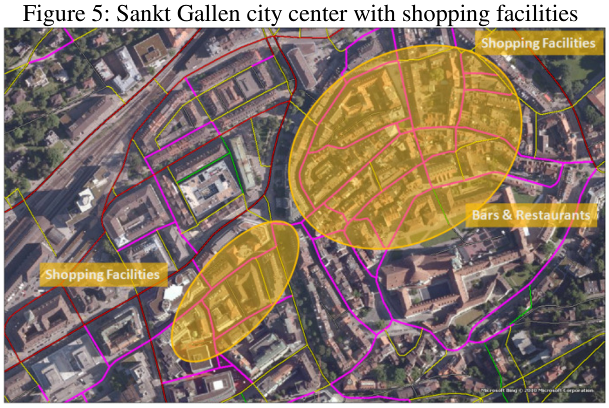


Figure 6: Sankt Gallen recreational areas

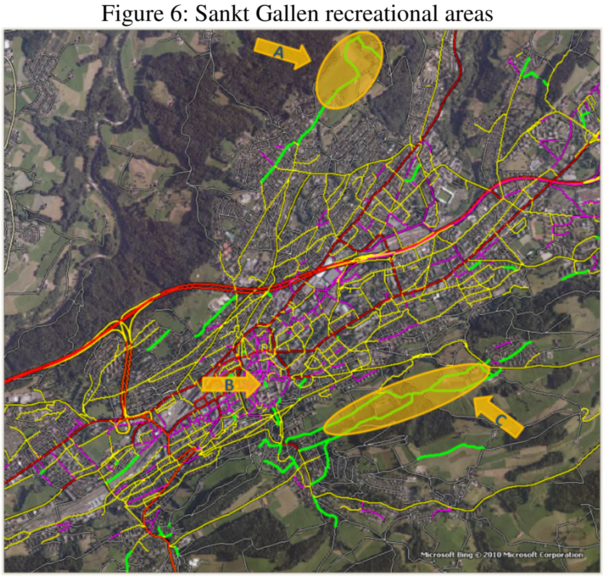


Figure 6 shows details of areas containing Cluster 3 street segments. In general, these segments reside in the northwest and southeast of Sankt Gallen where parks and recreational areas lie. For example, in the area marked with A resides a wildlife park with restaurant and in the area marked with C an open-air swimming pool. Furthermore, the street segment marked with B, which resides in the city center, belongs to a little park inside the abbey church. These examples confirm our assumptions from the temporal analysis of cluster 3 that the contained street segments are predominately used for recreational and leisure activities.

Figure 7: Clustering Lausanne

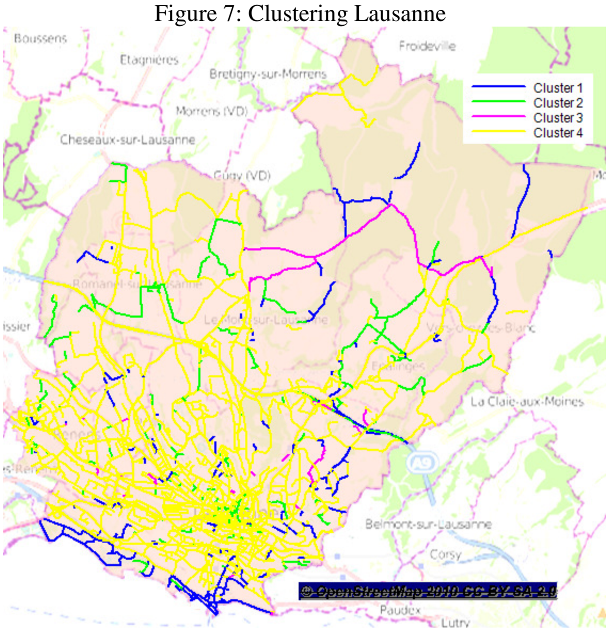


Figure 8: Lausanne, recreational areas along Lake Geneva

Figure 7 shows a map of Lausanne with street segments colored according to their cluster id. Similar to Sankt Gallen does Cluster 5 with the majority of street segments not allow to draw specific, location-dependent conclusions about the usage behavior. However, Clusters 2, 3 and 4 allowed for further interpretation by visual analysis. Figure 8 shows details for Cluster 2. Its segments are located along the Lake Geneva. The area marked A contains sports facilities which are most likely used for sport events during the weekend. The segment allowing entrance to the area marked B gives access to an open-air swimming pool, which also contains access to the lake. Finally, area C contains the harbor and promenade, including a landing place for ferries to France and a metro/subway station. All of these areas are intended for leisure and recreational activities, and confirm the temporal patterns described in the previous section.

The visual analysis of Cluster 3, which showed a characteristic peak on Saturday morning as well as a high activity during the opening hours of shops, showed that the cluster contains in part collections of street segments in the inner city of Lausanne as well as in its surrounding suburbs. However, segments are also located within residential areas and on access roads. Thus, the segments are not only characterized by shopping activities but also by access functionality and homely activities. Clearly, the two latter show in their temporal behavior the departure and arrival times of people. Finally, the spatial distribution of segments belonging to Cluster 4 are distributed over Lausanne and have primarily an access function to residential areas or major roads. Thus, the cluster symbolizes especially movement to and from residential areas, which is plausible considering main peaks around lunch and dinner time as well as peaks on late Saturday evening and early Sunday morning. We suppose that these characteristics show mainly on access roads instead of directly within the residential areas because these streets are likely to contain more passages and thus yield more stable results.

5 Discussion

The goal of our analysis was to determine whether the GPS data set collected by SPR+ is sufficient to draw inference about temporal usage patterns of street segments. Our results show that such an analysis is possible, however, within limits. First, we were able to include about 60% of the street segments in each city in our analysis when applying a threshold of 20 passages per segment. Clearly, a threshold of 20 is a lower bound and allows for noise in the data. This is clearly one reason that contributed to the two large non-divisive clusters in Sankt Gallen and Lausanne. However, compared to about 360 permanent traffic counting points available by the Swiss Federal Roads Authority [15] for whole Switzerland, the GPS data set possesses a good ground for temporal analyses and allows explicitly to analyze inner-city traffic and pedestrian movement. Second, we obtained clusters with temporally distinguished usage patterns. The visual inspection of these clusters showed that especially shopping and recreational activities have a unique temporal usage pattern. However, the clusters in Lausanne showed also that temporal patterns cannot distinguish clearly between consecutive activities. For example, the usage of access roads is closely connected to the usage in residential areas where individual movements typically start or end. Third, the clustering showed that the most characteristic time span to distinguish usage patterns is the weekend. This means with respect to data collection that observation periods up to 10 days should be placed to cover two weekends. Even better would be an extension of the survey, for example, to 3 weeks.

6 Summary and Future Work

In this work we analyzed temporal usage patterns of street segments based on large GPS surveys in two Swiss cities. We identified groups with a similar temporal traffic distribution by clustering and interpreted the results based on temporal and spatial background knowledge. We provided a detailed analysis for the two Swiss cities Sankt Gallen and Lausanne for which we could clearly identify traffic patterns related to shopping and leisure activities.

In future work we plan to extend our analysis by including data sources about sociodemographic information of the test persons as well as environmental information of the street segments. Such an extension could either be integrated into the existing clustering process or be carried out subsequently following a multi-stage approach. In addition we would like to increase the granularity of our results by interpreting not only the clusters in total, but also their characteristic peaks. One possibility to do so would be to analyze diary data from other mobility surveys and to merge the annotated patterns with the results of our analysis.

7 Acknowledgement

We would like to thank SPR+ for the provision of the GPS data set in Switzerland.

References

[1] G. Andrienko and N. Andrienko. Spatio-temporal aggregation for visual analysis of movements. *In Proc. of IEEE Visual Analytics Science and Technology (VAST 2008)*, pages 51-58, IEEE Computer Society Press, 2008.

[2] M. Ester, H.-P. Kriegel, J. Sander and X. Xu: A density-based algorithm for discovering clusters in large spatial databases with noise. *In Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*. pages 226–231, AAAI Press, 1996.

[3] A. Fitschen and H. Nordmann. Verkehrsentwicklung auf Bundesfernstraßen 2008 (Traffic development on federal roads 2008). *Berichte der Bundesanstalt für Straßenwesen (V 191)*, NW-Verlag, Bremerhaven, 2010.

[4] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779-782, 2008.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, Volume 11, Issue 1. 2009.

[6] J.-G. Lee, J. Han and K.-Y. Whang. Trajectory clustering: a partition-and-group framework. *In Proc. of the 2007 ACM SIGMOD International Conference on Management of Data (SIGMOD'07)*, 2007.

[7] W. Meijermars. Analysis of urban traffic patterns using clustering. PhD Thesis, *TRAIL Thesis Series T2007/3*. 2007.

[8] T. Mitchell, Machine Learning. McGraw Hill, 1997.

[9] M. Pasquier, U. Hofmann, F. H. Mende, M. May, D. Hecker and C. Körner. Modelling and prospects of the audience measurement for outdoor advertising based on data collection using GPS devices (electronic passive measurement system). *In Proc. of the 8th International Conference on Survey Methods in Transport*, 2008.

[10] N. Pelekis, G. Andrienko, N. Andrienko, I. Kopanakis, G. Marketos, and Y. Theodoridis. Visually exploring movement data via similarity-based analysis. *JIIS Online First*, pages 1-49, 2011.

[11] C. Piciarelli, C. Micheloni, G.L. Foresti. Trajectory-based anomalous event detection, *IEEE Transactions on Circuits and Systems for Video Technology*, 18(11):1544-1554, 2008.

[12] S. Rinzivillo, D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko and G. Andrienko. Visually driven analysis of movement data by progressive clustering. *Information Visualization*, 7(3):225-239, 2008.

[13] F. Rotberg, S. Schömann-Finck and O. Schwede, editors. *Digitale Außenwerbung 2011, Digital Signage verändert den Außenwerbemarkt (Digital out-of-home advertising, Digital signage changes the out-of-home advertising market)*. invidis consulting GmbH, 2011.

[14] R. Schlich and K. W. Axhausen. Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation*, 30:13-36, 2003.

[15] Swiss Federal Roads Autority (ASTRA). Swiss automatic road traffic counts. http://www.astra.admin.ch/verkehrsdaten/00297/index.html, accessed 2012.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke