Capturing the beaten paths: A novel method for analysing tourists' spatial behaviour at an urban destination[1]

Marko Modsching[a], Ronny Kramer[a] and
Ulrike Gretzel[b] and Klaus ten Hagen[a]

[a] Department of Computer Sciences
University of Applied Sciences Zittau/Görlitz, Germany
{K.tenHagen, RKramer, MModsching}@hs-zigr.de

[b] Laboratory for Intelligent Systems in Tourism
Texas A&M University
ugretzel@tamu.edu

Abstract

The spatial behaviour of tourists has a large impact on the economic situation of a tourist destination. Stakeholders such as city councils have a strong interest in evaluating spatial behaviour to avoid misallocation of their scarce resources. In addition, individual service providers need to know where the tourists are in order to offer appropriate services and market them more effectively. The following paper suggests a methodology to analyse the spatial behaviours of tourists based on real tracking data. This method was applied during a field study in the city of Görlitz.

Keywords: tourism; spatial behaviour; tour tracking; restaurant selection; destination management; mobile information system;

1 Introduction

For many sectors in the tourism industry it is essential to know the places and times tourists visit. Restaurants for example use this information to find the best location in order to have a maximum amount of visitors. Real estate brokers also need this information to be able to establish prices. Companies are very often, and for good reasons, willing to pay more to receive a good location for their establishment. Destination management tries to redirect tourist flows to take advantage of the entire

---

[1] Published in "Information and Communication Technologies in Tourism 2006", Andrew J. Frew, Martin Hitz, Peter O'Connor (Editors), Springer Computer Science

destination area and to avoid overcrowding at single places. Unfortunately data about the spatial behaviour is rarely collected and evaluated together with contextual information because of the lack of a suitable method. Credit card companies or telecommunication enterprises are collecting customer specific data to understand their behaviour in order to optimize their product marketing. The only tourist-related information most public organisations like city councils have is the total number of overnight stays, but little is known about tourists' actual behaviours in terms of activities and specific locations visited. The simple capturing of pictures and counting of tourists has it's limitations, since it is hard to distinguish between a tourist and an ordinary inhabitant. But with today's available technologies like wearable GPS receivers it is possible to track the paths tourists are taking. This data can be evaluated in different ways. The most obvious one of course is the visualisation of the track of the tour in a map and the indication of the places the tourists have been at. However, there is a lot more that can be gained from analysing the logged data. This paper presents an approach for analysing the complex spatial behaviour of tourists and for pointing out the actual utilisation of different Points of Interest (PoI). Thereby the success of different marketing actions can be measured and evaluated quantitatively.

2 Related work

A study by Freytag (2003) showed that the spatial behaviour of tourists in Heidelberg is very concentrated, with most tourists visiting the Old Town and some attractions receiving very little attention. Kempermann et al. (2004) tracked visitor behaviour at a theme park and found that significant differences exist between first-time and repeat visitors: First-timers visit a large number of attractions while repeat visitors are much more selective. Hwang, Gretzel and Fesenmaier (2005) analysed the spatial behaviour of international tourists to U.S. cities and found that trip patterns have important implications for destination bundling and cooperative destination marketing efforts.

Traditional ways of measuring tourists' spatial behaviour rely on diary data completed during a visit or survey questionnaires completed after the actual tour or trip. The problem with the first method is that it is highly intrusive, whereas the latter relies on the honesty and memory of people when providing their information. It is quite unsure how credible these data are. The ideal would be to log the movement of a tourist and to measure the time without letting the tourist take note of it. Several such methods have been implemented in the context of transportation and retail studies:

- Dijkstra, Jessurun and Timmermans (2001) implemented a model that simulates the movement of pedestrians by agents. Because of defined rules, the agents either move or wait within different cells. They visualise possible interactions of pedestrians in crowded areas.

- Shoval and Isaacson (2005) compared geographic information systems like GPS to land-based tracking systems; these are units sending signals to antenna stations that calculate the position, by measuring the movement of pedestrians. The main advantages of GPS are the worldwide ability, little costs and exacter positions, whereas land-based tracking systems have the advantages of being unaffected by the weather and work also well in urban regions and indoors.
- Larson, Bradlow and Fader (2005) analysed the paths of shoppers in a supermarket with RFID tags located on their shopping carts. The tracked pathways were clustered to find out typical routes through a grocery store.

The methodology used in studies like those of shopping activities could provide important insights regarding tourist behaviours if expanded to an urban destination. However, RFID tags are not suitable for tracking tourists' movements in a city as their range is limited to less than 1 meter. Instead the worldwide availability of GPS signals makes GPS receivers the perfect method in order to track changes in tourists' positions together with a timestamp. The stored position data can be mapped and used to analyse the aggregated spatial behaviour of tourists. This analysis indicates which attractions or services are visible to the tourists and where information systems might help in diversifying spatial behaviour and redirecting tourists to certain attractions that might be of greater interest to them than the most obvious options. In addition, spatial data enriched with timestamps provides important information about tourists' time allocations over the course of a visit.

3 Analysis methodologies

Several types of analyses are possible for data that results from a GPS tracking study and can be summarized into two broad categories: 1) Analyses using spatial maps; and, 2) Analyses of activities.

3.1 Spatial maps

Most tourists tend to stay on the beaten tracks when visiting urban destinations (Freytag, 2003). Mobile or web-based information systems are built to enable tourists to discover sights or activities that suit their personal tastes. In order to assess the effectiveness of these and other more traditional marketing measures, spatial distribution metrics can be used. A density map provides insight into the spatial distribution of tourists. In order to compute a density map the map is divided into an equal grid. If a track crosses a cell the number of visitors for that cell is incremented. After all tracks have been analysed the number of visitors to a single cell is divided by the total number of visits to all cells in order to compute the relative number of visitors to each cell.

For a certain time period the relative number of visitors RNoV to grid cell(i,j) can be defined as:

$$RNoV(i, j) = \frac{NoV(i, j)}{TNoV} \quad (1)$$

with NoV(i,j) the number of individual visitors to cell(i,j) and TNoV the total number of visits to cells being tracked in this time period. Obviously $0 \leq RNoV(i, j) \leq 1$ is valid.

The authors of this paper have introduced the spatial distribution metric SDM for a given time period, a destination and a set of tracked tourists defined as:

$$SDM = -\sum_{\forall i, j} RNoV(i, j) \log_2 (RNoV(i, j)) \quad (2)$$

SDM reaches its maximum if all RNoV values are equal. The maximum is $ld(I * J)$ with I and J the number of rows and columns of the grid. The Relative Spatial Distribution Metric is defined as: $RSDM = SDM / \log_2 (I \cdot J)$ (3)

The following table shows that RSDM is a suitable metric to measure the equality of the spatial distribution:

Table 1: RSDM calculation examples

NoV(i,j)		**RSDM**	**NoV(i,j)**		**RSDM**	**NoV(i,j)**		**RSDM**
10	10	1	10	0	0	10	1	0.22
10	10		0	0		0	0	
10	1	0.41	10	5	0.75	10	5	0.9
1	0		5	0		5	2	
10	7	0.96	10	7	0.98	10	8	0.99
5	4		6	5		8	7	

3.2 Activities

Tracked positions and timestamps can be used to find out the activities performed of tourists. Activities can be extracted from the tracking logs using two approaches:

- **Hot Area analysis**: Different regions with a geographic reference and an associated category (restaurant, shop, museum etc.) are modelled in advance. The time a tourist is located in the Hot Area is allocated to the associated activity.

- **Walking speed analysis**: The trajectory speed of the tourists has two modes: Either he/she is walking between sights or walking very slowly or even standing to explore a sight, to shop or to have a meal. Since most tourists slow-down in the same area the sequence of analysis starts with a filter to determine the mode of a tourist. All points for one block of slow-down are associated with one area. If many tourists spend time in slow-down mode in a particular area local expertise is used to determine the most likely kind of activity. Then the time spent in this area is allocated to this kind of activity for all tourists.

4 Field Study Design

All together 15 mobile devices were used for the experiment. The experiment took place in the inner city of Görlitz for about a month in the summer of 2005. Students and scientific assistants handed out the mobile devices together with a GPS receiver and an external antenna in the morning. The GPS antenna had to be put onto the shoulder; the rest could be put into a pocket. The tourists were promised a gift when returning in the afternoon. The starting point was situated at a central point in Görlitz. After receiving the device, tourists were asked to go about their tour as they pleased.

On the mobile device an application was already running, while all buttons were disabled, so that the tourists could not stop the application or start any other. The mobile digital assistant (MDA), a handheld device with an integrated mobile phone, was also connected to the GPS receiver via Bluetooth. This application constantly logged the positions delivered by the GPS receiver. When the tourists returned the mobile device, the application was stopped by an assistant, the mobile device was connected to a PC workstation and the data were stored into a database. A PC application then read the data from the database and generated a map where the routes, presenting the current allocation of tourists, were visible and diagrams with tour characteristics were computed.

To guarantee the privacy of the attendees a juridical assessment about the logged data was made by the research group provet residing in the university of Kassel. All users who took part of the experiment had to read and sign their agreement to give the permission of logging the data. They were promised that all personal data will be deleted after they had reached the stand and they had giving back the device. The only data which is left is some demographical data and information about the computer literacy.

5 Results

From the 65 real tourists of the destination Görlitz who took part in the field study 65% were male. The age of the tourists ranged from 17 to 81 while the average age was 46. Since the age of the majority of the tourists of Görlitz is higher than 56 the tourists willing to participate in the study were younger.

5.1 Spatial distribution of tourists

The number of tourists who visited a certain grid cell was calculated and the specific density level was assigned a colour to map different levels of concentration in the spatial behaviour of visitors to Görlitz. The result is displayed in Fig. 1.

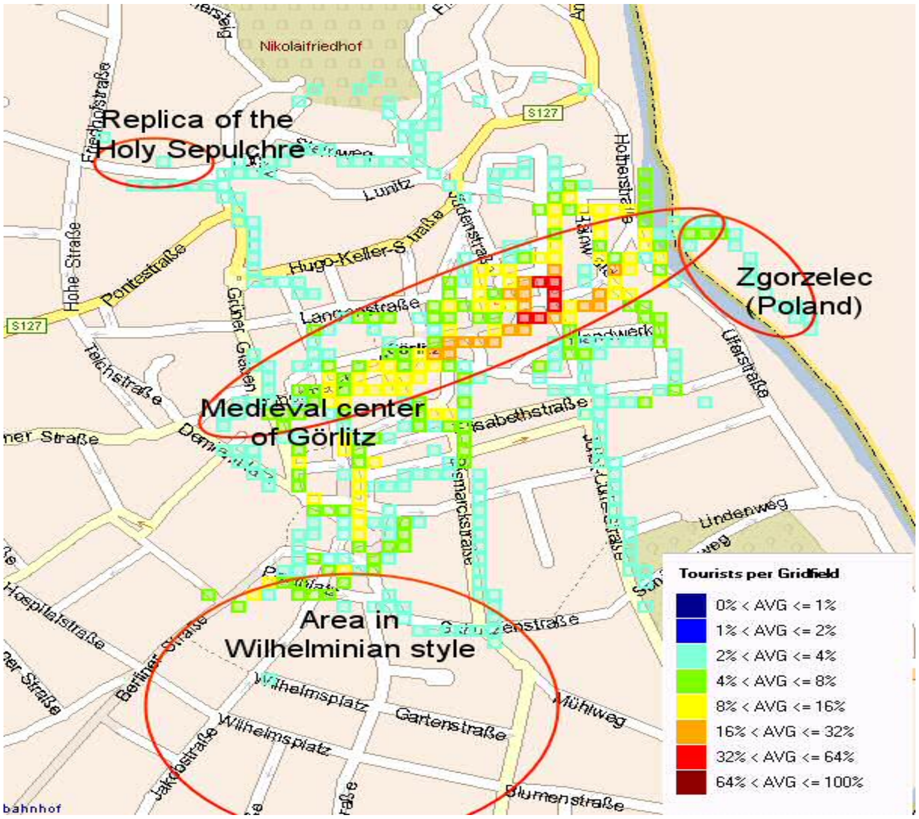

Fig. 1: Spatial Distribution of Tourists

The map indicates that tourists' spatial behaviour is highly concentrated within the medieval city centre and that attractions such as the Replica of the Holy Sepulchre and the Wilhelminian Style Area are hardly frequented at all. The map indicates a clear need for marketing (e.g. creating awareness) and management (e.g. better signage) to diversify tourist flows at the destination because there are a lot of unique sights existing in Görlitz which would surely better visited by the tourists if they had more information about that.

After introducing such strategies, the distribution of tourists could be tracked again to make impacts on behaviour visible. However, to compare two distributions such density maps lack concrete numbers. As explained in the analysis methodologies section, the RSDM metric is a value suitable for measuring the equality of a distribution. The maximum SDM value for the given grid is 14.4 and the measured SDM is 8.4, thus the RDSM value is 0.6. This value gives an impression that the actual distribution is far from being optimal and could surely be improved by certain marketing methods whose effect can be easily measured and compared with the situation before using the RSDM metric as a single figure of merit.

5.2 Activities

Activities discovered by hot-areas. Ten Hagen et. al. (2005) illustrate that virtual areas which are modelled upon a geographic map can be used to trigger certain actions. These areas can also be used to analyse the activities of tourists.

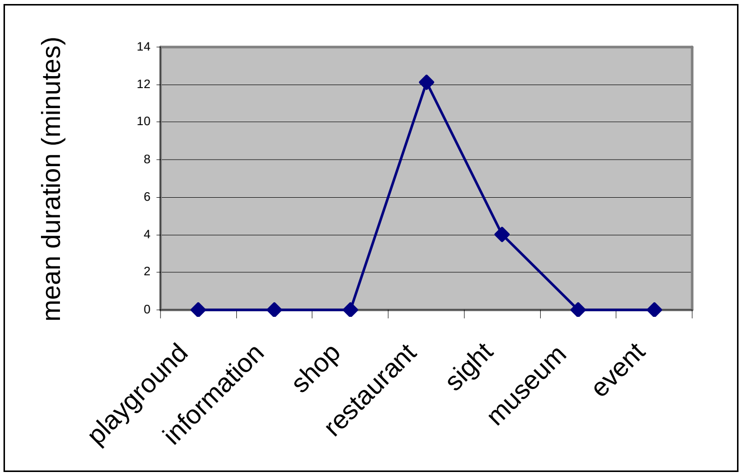

Fig. 2: Time distribution for different activities

Each time a tourist enters and stops inside a hot area the time spent for this activity is increased appropriately for this activity. This data can be used to analyse how tourist allocate time to different activities. Since each activity belongs to a category, by aggregating the time spent for each activity the total time spent in various activity categories can be compared. The data collected during the field trial suggests that tourists spend more time at restaurants than at attractions, which confirms the results of Schmidt-Belz and her colleagues (2003a; 2003b) regarding the importance of restaurants in tourists' needs and time budgets. Figure 2 displays the mean times spent for certain activities.

With the presented hot area method special evaluations for different sights can be used to find out how effectively these Points of Interest are used by tourists. As an example the evaluation of the “Vierradenmühle” which is located in Görlitz shows that tourists stay there on average for only 2.3 min. Tourists who only take a look will need less time and tourists who eat there will need much longer. This indicates that most tourists did not frequent the restaurant; as a matter of fact, only two tourists decided to become patrons of the “Vierradenmühle”.

Automatic discovery of activity areas by walking speed. Since the analysis using hot areas depends on the quality of the localisation technology e.g. GPS, another more robust method had to be found for analysing activities. The locations of the activities might also be discovered automatically considering different walking speeds. Figure 3 shows a typical walking speed distribution for a single tourist. The distribution is clearly bi-modal. Every speed below 3 km/h could be interpreted as being in “slow-down”. All other points of the track with a speed above 3 km/h are interpreted as “walking between activities”. Since the walking speeds depend on the tourist the threshold to separate the two modes is computed for each tourist individually.

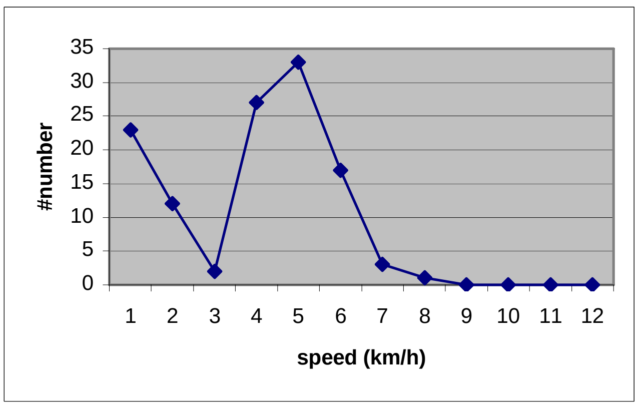

Fig. 3: A typical walking speed histogram of a tourist

The positions the tourists stopped at were drawn into a map afterwards. Since a lot of tourists stop at the same locations many point-clouds are visible on the map. To get a clear conclusion about the centroids of these crowds, the points were clustered with a k-means algorithm shown in Pawas (2004) and works as follows:

```
foreach( Point in Map )
{
  foreach( Cluster in Map )
```

```
    if(GetDistance(Point,Cluster.GetCentroid())==Minimal)
      AddPointToCluster();
}
foreach( Cluster in Map )
  SetCentroid(Mean(Cluster.Positions));
if(Clusters.HasChanged)
  RepeatAll();
```

The map shown in Figure 4 was produced using this clustering algorithm. The initial positions of the centroids could be determined by the user. The circles represent the position of the activities. The darker the circle the more tourists attended the attraction and the greater the circle the longer the tourists stayed there. This map illustrates that by analysing the walking speed, the system identifies the Points of Interest (PoI). After the discovery process human expert knowledge is used to associate the areas of interest with an activity, like a sight or a restaurant. Thus, the automatic discovery process proceeds in the following steps:

1. Filtering of the walking speed of each tourist to identify the individual PoIs and the distribution of durations.
2. Cluster the PoIs of all tourists
3. Compute duration distributions for the “clustered PoI”.
4. Visualization:
    i. Compute a center for each PoI
    ii. Draw a circle with the number of tourists indicated by the colour/shade and the average duration by the size.

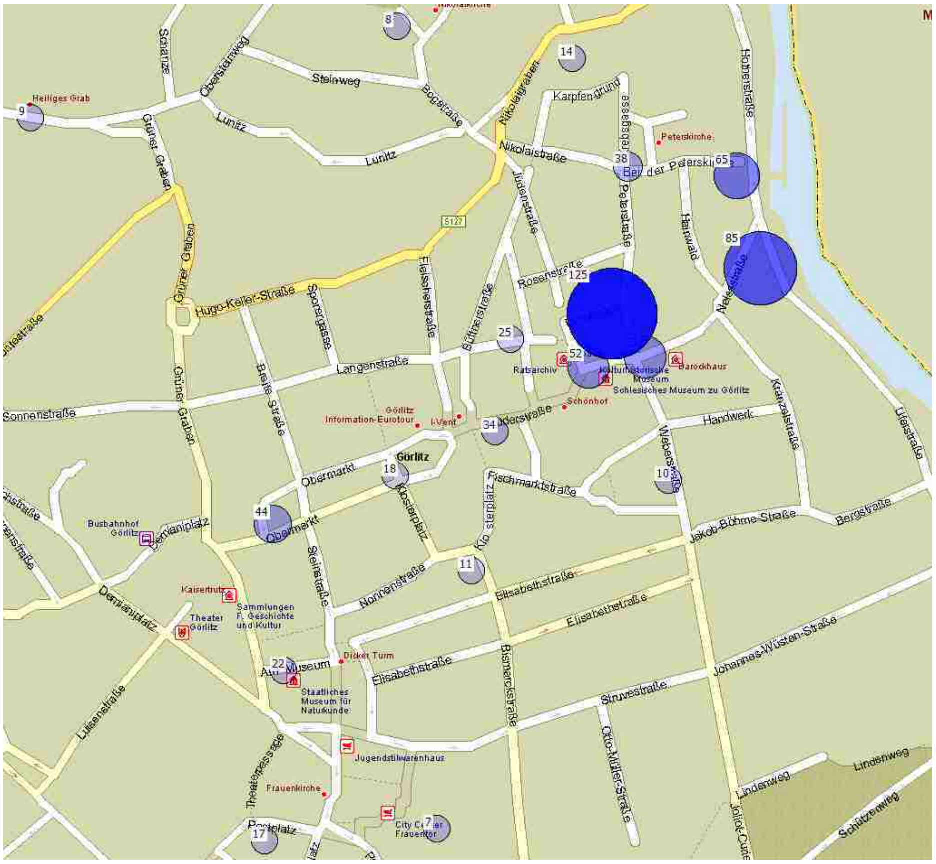

Fig. 4: Automatically discovered activity areas

The walking speed analysis method is much more robust to location errors because of the following reasons:

1. In case of a GPS position between two walkways the navigator might correct the position to one of them. This leads to high-speed jumps, which are filtered out.

2. Systematic -even location dependent- errors of the GPS position are easily be dealt with by a human, who assigns activities to clusters, since these errors are too small to cause an association with an adjacent activity.

3. There is averaging in associating a sequence of slow-speed positions with an Area of Interest (AoI). There’s another averaging when AoIs are clustered together to create an Area of Activity (AoA). These two steps for averaging remove a significant amount of spatial noise in the location measurement.

6 Limitations and future research

There are several things which can happen during a tour that can cause a misinterpretation of the gathered data with the hot area method. For instance, stops for meeting friends or orientating might be considered as sightseeing. Human expert knowledge is crucial in determining whether a stop occurred in conjunction with an attraction or not. However, there will always be a certain extent of noise in the data which will vary depending on the density and types of attractions available at the destination.

The study methodology might have encouraged a self-selection process since the sample contains more males and more young people than the general population of visitors to Görlitz. It could be that the prospect of having to use a mobile technology or privacy concerns discouraged some individuals from participating in the study. Future studies need to make sure that a representative sample is included. This is probably dependent on being able to provide incentives that could help overcome participation barriers.

The position data in the field trial was collected by a combination of an MDA and Bluetooth GPS receiver. This combination is prone to errors because the Bluetooth connection or the operating system may break down. New studies of spatial behaviour should be conducted with standalone GPS loggers. These loggers should be handed out at various points within the destination to reduce potential bias in the data and make sure that tourists are being tracked from the very beginning of their tour. Also, there needs to be a sufficient amount of drop-off points (e.g. hotel receptions) to ensure that the data collection methodology does not interfere with the actual tourist behaviour.

Recommender systems have been proposed as tools to influence tourist behaviour by providing specific suggestions tailored to the preferences of the particular user (Nguyen, Calvada & Ricci, 2004). The methodology presented in this paper can not only be used as input for the generation of tours but also as a tool to evaluate whether the provision of attraction-related information has an impact on tourists’ actual behaviour when touring a city destination. Future research in this area will provide important input regarding drivers of tourist spatial behaviour and will also offer valuable insights regarding the design of tourist information systems.

7 Conclusion

The results of this field study clearly show that it is possible to rigorously analyse the spatial behaviour of tourists in a city. Especially with a rising number of tourists it becomes more important for city councils as well as business owners to know where tourists really go and what activities they engage in. With the presented analysis methodology the spatial behaviour of tourists can be analyzed quantitatively. This provides important input to planning processes but can also be used to evaluate the implementation of certain measures such as specific marketing campaigns by comparing pre-implementation data with spatial behaviour after the measure was introduced. The steadily decreasing cost of mobile devices and GPS tracking devices will make this methodology ever more affordable and will hopefully also increase its actual use for research purposes.

References

Djikstra, Jan; Jessurun, Joran; Timmermans, Harry (2001): "A Multi-Agent Cellular Automata Model of Pedestrian Movement" *Pedestrian and Evacuation Dynamics, Schreckenberg and Sharma (eds.)*, Springer-Verlag, Berlin.

Freytag, Tim (2003): „Städtetourismus in Heidleberg – Ergebnisbericht zur Gästebefragung 2003.", Geographisches Institut der Universität Heidelberg.

Hwang, Y.-H., Gretzel, U., and D. R. Fesenmaier (2005). Multi-city Trip Patterns: International Tourists to the U.S. *Annals of Tourism Research*, forthcoming.

Larson, Bradlow, Fader (2005): "An Exploratory Look at Supermarket Shopping Paths"

Kempermann, Astrid; Chang-Hyeon, Joh; Timmermans, Harry (2004): "Comparing First-time and Repeat Visitors' Activity Patterns in a Tourism Environment" *Consumer Psychology of Tourism, Hospitality and Leisure Vol.3*, CAB International.

Nguyen Q.N., Cavada D., Ricci F. On-Tour Interactive Travel Recommendations. In Information and Communication Technologies in Tourism 2004, Proceedings of the International Conference, Cairo, Egypt, January 26 - 28, 2004. Pages 259-270. Springer.

Pawan Lingras, Chad West (2004): "Interval Set Clustering of Web Users with Rough K-Means", Journal of Intelligent Information Systems

Schmidt-Belz, Barbara; Posland, Stefan (2003a): „User Validation of a mobile Tourism Service"; Workshop *HCI mobile Guides*, Udine (Italy).

Schmidt-Belz, Barbara, Laamanen, Heimo; Posland, Stefan; Zipf, Alexander (2003b): „Location-based Mobile Tourist Services – First User Interaction" *Information and Communication Technologies in Tourism 2003, Andrew Frew et al. (eds.)*, Springer Computer Science.

Shoval, Noam; Isaacson, Michal (2005): "The Application of Tracking Technologies to the Study of Pedestrian Spatial Behaviour" *The Professional Geographer*

Ten Hagen, K. et al.(2005b): „Context driven adaptive tour computation and information presentation", First International Workshop on Managing Context Information in Mobile and Pervasive Environments; Ayia Napa, Cyprus.

Acknowledgements

This work is part of the IKAROS project funded by the federal ministry for education and research (BMFT) and part of the VESUV project in cooperation with Siemens AG, Microsoft's European Innovation Center (EMIC), the city of Görlitz and Fraunhofer Institute IGD. VESUV is supported by the Federal Ministry for Economics and Labour (BMWA).