Analytical Workflow of Monitoring Human Mobility in Big Event Settings using Bluetooth

Hendrik Stange
Fraunhofer IAIS
Schloss Birlinghoven
D-53754, Sankt Augustin
hendrik.stange@iais.fraunhofer.de

Thomas Liebig
University of Bonn
Dept of Computer Sciences III
D-53117, Bonn
liebig@informatik.uni-bonn.de

Dirk Hecker
Fraunhofer IAIS
Schloss Birlinghoven
D-53754, Sankt Augustin
dirk.hecker@iais.fraunhofer.de

Gennady Andrienko
Fraunhofer IAIS
Schloss Birlinghoven
D-53754, Sankt Augustin
gennady.andrienko@iais.fraunhofer.de

Natalia Andrienko
Fraunhofer IAIS
Schloss Birlinghoven
D-53754, Sankt Augustin
natalia.andrienko@iais.fraunhofer.de

ABSTRACT

In recent times, consumer research at major social events received significant interest by organizing companies. Understanding the movements and motivations of the customers enables new business strategies and is needed to minimize the risk of investment. The spatiotemporal complexity of major events poses high demands on survey and analytical methods. New technological advances in both event monitoring systems and evaluation methods of movement data provide new insights into the behavioral patterns of customers by preserving their privacy. In this paper we present a work that seeks to systematize the research process of design, collection, and analysis of visitor behavior in a mixed indoor-outdoor event setting using Bluetooth sensor technology. The defined workflow is comprised of 5 steps and designed to answer heterogeneous business questions with respect to customer movement behavior in a single event context. Our approach is applied in a real-world business application for a Formula 1 event.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]

General Terms

Measurement, Economics, Standardization

Keywords

Event Monitoring, Mobility Mining, Visual Analytics, Bluetooth

ISA '11, 1 November 2011, Chicago, IL, USA

1. INTRODUCTION

Many major events like Formula 1 races or grand festivals cost millions of Euro and only take place once every year. Organizing companies of these social events carry a high risk of investment. Their sustained success much depends on the attraction of visitors and how they manage to motivate potential customers to use their services during the stay. One way towards optimized services is by understanding the travel behavior and motivations of customers. Until recently, companies relied on decisions based on past experiences or trial and error. Investments were therefore left to chance. Nowadays, they look for a more systematic approach of controlling their investments and seek for new optimization potential. Intelligent, seamless monitoring systems give insights into local human mobility, dependencies, and potential influence factors.

Monitoring flows of pedestrians across boundaries of indoor and outdoor space allows reviewing the performance of an investment, e.g. building investments, new attractions. Furthermore, the temporal and spatial setup of an event can be evaluated for future optimizations: Could the attractions be easily reached, was the line-up of acts well planned, why did one attraction underperform?

Major events likely exceed 10,000 visitors. Hence, systems monitoring these type of events must meet strict security, business and performance conditions. We are proposing to deploy a Bluetooth based Monitoring System (BMS). Due to the complex nature of big social events the monitoring technology must be flexible and failsafe as it is generally not installed permanently. It should be simple to install without imposing particularly high positioning restrictions and it should not require any expert calibrations. Additionally the technology used should equally operate in a mixed indoor and outdoor setting, also be unsusceptible to weather conditions, changing light conditions and independent of electricity supply. Last but not least the monitoring system must scale well and allow to anonymously track mobile devices on a large scale. In business settings it is also important that the systems must be cost-effective. All requirements are met by the used BMS.

In this paper we present an approach to monitoring spatiotemporal divergent major events for understanding the local and time depend mobility of visitors. Monitoring places and movements deliver answers to key business questions. The proposed workflow is comprised of two phases: survey design and collection of the data utilizing Bluetooth sensors and the subsequent process of knowledge discovery using Mobility Mining methods and Visual Analytics. The first phase ensures an optimal distribution of the sensors choosing an appropriate placement strategy. The second phase evaluates the data under the given research questions and allows for secondary data to be integrated as background knowledge. We will demonstrate the functionality of the approach in a real-world business application.

The paper is organized as follows. A short review of current scientific work will be given in the next section. Section 3 describes our workflow and its steps. In Section 4 we demonstrate our approach at this year's Formula 1 Grand Prix event which took place at the Nürburgring in Germany, and Section 5 concludes the paper and points to the direction of future work.

2. RELATED WORK

Understanding consumer behavior has always been the goal of marketing research. The task is to deliver key information about the needs and wants of a company's customers. Current research mainly focuses on understanding the purchase decision of people. Quantitative and qualitative research plans have been developed and adapted for different economical situations tro:konsument, [6]. Recent work has been done in analyzing new data sources for gaining insights into the travelling behavior of people e.g. for the tourism industry [12], gir:footprint, [3]. With rapid development and access of mobile devices new methods have been proposed for analyzing massive distributed movement data [12], [1]. In general analyzing human movement data has received a lot of attention especially in the field of Visual Analytics [4], [2].

Visual Analytics aims at combining the strengths of human and computational data processing. It is used out of two reasons: first, for visualizing the results of the mining process and second, for making analytical tools accessible and usable to business experts. The most common techniques for the visualization of movements of humans are static maps with directed linear symbols [19], animated maps [5], and space-time cubes, where two dimensions represent space and the third dimension represents time [11]. Links between places can be studied when movement data is aggregated into origin-destination matrices and flow maps [16].

In recent time the rapid spread of mobile phones with multiple integrated sensors inspired many researches on tracking people or more generally on collecting movement data. Parallel to recording human outdoor movements using GPS technology new solutions for tracking indoor movements have been required. The reason is that GPS is sensible to physical shields like those of roofs. In recent years Bluetooth based sensor technology has evolved to track people indoors. The focus of scientific works concentrated on accurately locating and following objects. Sensor and analysis methods have been proven to allow for fine-grained positioning and tracking of indoor moving objects [9], [10]. Some work has been conducted recording flows of outdoor movements using Bluetooth [15]. Weghe et. al. have investigated the applicability of Bluetooth for collecting movement data during big events to find answers to security and other questions [20].

Figure 1: Research workflow for monitoring big events

3. EMPIRICAL RESEARCH WORKFLOW

Events being subject to high investment risks and uncertainties are likely heterogeneous in space and time. They take place over 2 or more days, indoors and outdoors with a differing number of attractions. Under these varying settings we are in need of a flexible research concept. Many of such concepts have been proposed for different domains. None of them deals with monitoring major events using Bluetooth sensors to answer high level customer related business questions. However, there are special constraints with recording customer behavior using Bluetooth technologies.

One of the constraints is related to the base technology. The sensors - mobile and fixed - only have a certain range to cover; they are limited in the number of measurements per second and are influenced by surrounding objects and other radio networks like Wireless LAN (sources of interferences). Another issue with Bluetooth is that it needs to be activated and set public on the mobile devices in order to be scannable. This may pose problems with regard to the representativeness of a study. Electricity, size, and portability may present some issues, too. Security restrictions also apply (i.e. fire, accident, vandalism etc.). One other issue with Bluetooth based monitoring technologies may be privacy concerns as they are able to record a unique identifier of a Bluetooth device within their reach making people recognizable.

Our descriptive research design addresses these constraints. It is intended to systematize the proceeding of monitoring and evaluating major events based on Bluetooth technology. The goal of descriptive research is to observe and describe the behavior of visitors of a monitored major event without influencing them in any way. In order to achieve this goal we propose a five-stepped concept which we group into two phases (see Figure 1).

3.1 Empirical data collection

In the initial phase of the process the field survey is conceptualized. In the Survey Design step the main tasks are to understand the business, to define business related research questions and to plan the data collection. The latter task aims at defining a suitable sensor placement strategy for the area being monitored. It can be applicable to include an on-site inspection.

The sensor placement strategy is essential for the success of the monitoring. Most areas of major events may easily exceed $10,000 km^2$ making it too costly to cover all areas. Therefore, placing the sensors must not only follow rules of efficiency and economy but be strategically planned to cover the desired area and measure all defining points of an individual track. We can only understand the movement behavior if we "see" it. In practice we need to set the range of each sensor (there are different types of antennas with different reaches), find neuralgic places which are safe and allow to cover the desired area without any interferences. One example of a neuralgic place would be an entrance area. As we are neither interested in accurately locating people nor devices we do not need to overlap the sensor coverage areas too much as this also produces a lot of noise in the data. Nonetheless, in some situations it might be appropriate to install two sensors for error measurements and to assure that each device has been captured; especially, in major events with many 1,000 devices. Placement strategies should avoid blind spots where people leave the covered area and suddenly return. Or they should intentionally account for this circumstance beforehand.

Major advantages of Bluetooth sensors are the small size of the device, low battery consumption [10], and that there are no sensor adjustments necessary during operation. However, this sensor technology requires to set the intervals of measurement (Bluetooth echo or number of scans per minute). With respect to our research task we are not interested in a low level Bluetooth tracking rather than in detecting main flows of movements between different attractions over space and time. The data collected is stored on a SD-card or send directly over a radio network. For the latter the sensors need to be set up properly during Step 2 of our workflow.

In contrast to other technologies (e.g. light barrier) Bluetooth sensors collect privacy sensitive data. Each Bluetooth chip is uniquely identifiable by a Media-Access-Control address (MAC). A person might become trackable beyond the boundaries of an event. Hence, our BlueTechSensors (BtS) do only save an anonymized identifier valid for the time of the monitored event. We are using an embedded, irreversible, and event specific SHA-256bit encryption algorithm to scramble the MAC-address. That is, we ensure platform independent encryption of the address.

To support the process of hypothesis generation and to increase the validity of the study a combination of different sensor technologies can be included for error measurement and to add additional information. Laser scanners for example are applicable to count the absolute number of people visiting a certain attraction.

The following step "Data Collection" (Step 2) includes the installation of the sensors in the field. Hardware and software checks are done as well. One important task in this step is to double check the placement plan as major events are likely to change the schedule or modify the locations. During the event there is no further setup necessary. Depending on the type of data transmission (off-line or connected) it may also be necessary to properly set up the network.

3.2 Spatiotemporal analysis of Bluetooth data

The second phase of the research process aims at answering the previously defined research questions. We apply preprocessing methods for Bluetooth data and use Mobility Mining and Visual Analytics to analyze the sensor data and present the results to the business decision makers.

The "Knowledge Discovery Phase" is subdivided into three steps. The first step is "Data Preparation" (see Figure 2). Main tasks of this step are data cleaning and generating movement data from Bluetooth data records. Sensor data is prone to erroneous data, artifacts like spot readings of devices for only seconds, missed readings, parallel measurements due to overlapping sensor areas, device failures (while saving or sending the data), etc. For cleaning the data we use spatial and temporal characteristics of the sensor data adapted from the Sensor Cleaning Pipeline presented in [13] and the Visual Analytics Mantra for massive spatial referenced datasets [14], [12].

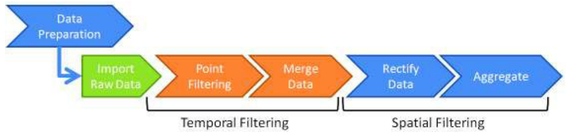

Figure 2: Data Cleaning Process for Bluetooth Data

First, the raw data of each BtS is imported into a database. Then, this data is cleaned beginning with temporal filtering: Point Filtering removes all duplicate entries over time using a time window approach. Depending on the measurement intervals (see section 3.1) this time window needs to be individually shaped also taking into account the sensor placement (overlapping). Erroneous readings or data fragments are deleted. Next, consecutive arbitrary measurements like spot readings (short term measurements) are combined to one meta-point. These measurements are not yet deleted because they might be part of a route between two attractions or indicate a short leaving of an attraction (e.g. smoking break, phone call). Afterwards consecutive data points of the same device and per attraction are merged. For each representation point the duration of stay and the number of measurement points are calculated. This information is later used to decide whether a device has visited an attraction for the entire duration of stay or for some reason left the surveyed area (test of duration plausibility).

Afterwards also a spatial filtering is applied: Rectify Data cleans spatial inhomogeneous measurements like ambiguous jumps between two areas or attractions due to overlapping sensor coverage or measurement errors. Therefore the spatial distance between sensors and sensor coverage areas respectively (e.g. walking distance in meter) are combined with time and duration to calculate speed and position changes per time. The goal is twofold: eliminate bogus position changes and gain first insights in people traveling behavior.

In the final step of the data cleaning process we aggregate the purged data points by space. When monitoring buildings usually all entrances and exits are covered. All data collected on these locations belong to the same attraction. Therefore we first partition the monitored area by means of buffers, tesselation, (dense based) clustering of sensors, similarity in movements or other methods. A plausibility check should be included; we even suggest involving local experts. After the data has been purified the database contains sequences of sensor or attraction visited by multiple devices (persons). This data can be interpreted as information about movements and behavior of visitors of major events. On the one hand, each visit of a person to an attraction can be viewed as an independent spatiotemporal event with a duration, motivation (need/want) and timely order. On the other hand the sequence of visits can be considered as a trajectory of this person. That is, a trajectory contains a sequence of positions (longitude, latitude), time, and transitional information. We call this class of data attraction-

Figure 3: Mining Process on Bluetooth Data

Focus Object	Attractions (1)	Visitors/Device (2)
Events	Tasks: examine the interestingness of places, explore temporal patterns of visits to these places Methods: discovering and clustering of time events, time-sequence analysis, spatio-temporal aggregation of events	Tasks: identify gatherings of people as concentrations of devices in the same area and time Methods: grouping of devices dependent on location and time, density based clustering
Trajectories	Tasks: investigate flows between attractions over time Methods: aggregation of moves, intensity flow maps, sequence analysis, pattern mining	Tasks: discover joint travels, frequently taken routes, similar moves Methods: interaction analysis, clustering of trajectories, subgroup discovery, self organizing maps

Figure 4: Tasks and methods in analysis of Bluetooth-datasets

based movement data.

Analysis of attraction-based movement data in major event context is a relatively new research topic. Andrienko et. al. have been pursuing research on event-based movement data which has similar data structures than our data [3]. We take a systematic approach to the analysis of this special type of movement data. We define a recursive Data Mining workflow comprised of two elements: Visual Analytics and Mobility Mining. Both elements are in a bidirectional circular relationship with each other (see Figure 3); meaning that both analytical field are integrated in an analytical refinement process. In simple terms, Mobility Mining is Data Mining with special focus on mobility data adapting algorithms from machine learning, information retrieval, statistics, and database theory. For the purpose of analyzing attraction-based movement data we developed an interactive visualization and mining process. We used Visual Analytics to understand the recorded event and dynamic data. Based on this understanding we extended the dataset, constructed parametric mobility models, and derived hypothesis to answer the research questions. The results of the modeling and learning were then again analyzed visually and our hypotheses were refined interactively.

Andrienko et. al. propose a typification of research tasks and methods [3]. We also defined a focus-object-matrix focusing on the spatial and temporal dimension of the data. Figure 4 shows a non-exhaustive synopsis of tasks and methods. Not all analytical methods are suitable for analyzing high-granular Bluetooth records of movement between only a few stationary and staggered-in-time attractions.

After mining the movement data for answers to the research questions at hand all findings are finally visualized in step 3 of the Knowledge Discovery Phase. In the following section we will apply our workflow for monitoring major events in a case study conducted at the Nürburgring, Germany.

4. MONITORING A RACE EVENT

Every two years the Nürburgring hosts the Formula 1 Grand Prix event in Germany with over 60,000 visitors. During one weekend in July visitors of all ages are offered a variety of attractions reaching from leisure activities (e.g. cart racing, cinema) and music events up to motorsports. All attractions are spatially and chronological distributed. With millions being spent on infrastructure, car racing and shows the investment risk is fairly high. Event monitoring enables organizers to understand their customers for better business decisions.

4.1 Survey Design and Data Collection

In expert interviews supplemented by an on-site inspection ahead of the event we learned about the context in which all movements and spatiotemporal events occur. Generally, it can be noticed that the business questions have two main focuses. (1) The first type of questions is attraction related; (2) the second relates on human movements. Questions of type 1 are:

- What is the duration of stay at an attraction?
- How often are attractions being visited over time?

Second type questions are:

- How do people move around the event area?
- How do visitors move around the area with respect to the point where they entered the premises?

The empirical research plan we developed concentrated on the main spatial and temporal attractions. When we speak of attractions we do so with two different meanings. First, attraction refers to local points-of-interest, e.g. a tribune or the shopping center. Second, temporal attractions are happenings in time, e.g. live music or the race. In order to monitor all attractions we calculated that we need 27 sensors. Three more sensors (green) we installed for error measurement and to account for expected high activities. Figure 5 shows the location of each BtS (red dot) and its estimated coverage area (blue circle). It can be noticed that there are areas with high density of sensors in the center. That is because of the location of two main attractions (Boulevard, RingřWerk) as well as the main tribunes (T4, T4a, Main) in the center area. Outdoor located sensors were battery operated with a runtime of more than 40 hours. All BtS operated offline logging the measurements to a SD-Card because the area was too large (more than 40.000km^2) to setup a wireless LAN and GPRS/LTE is too costly. The size of the area made it economical absurd to capture every move in detail. Furthermore, the research was not designed for real-time monitoring in this particular case study.

During the weekend of the 23rd and 24th July the BtS recorded all Bluetooth activity within their reach with an interval of 1 measurement every 3 seconds. To achieve this rate each sensor was equipped with 3 Bluetooth antennas.

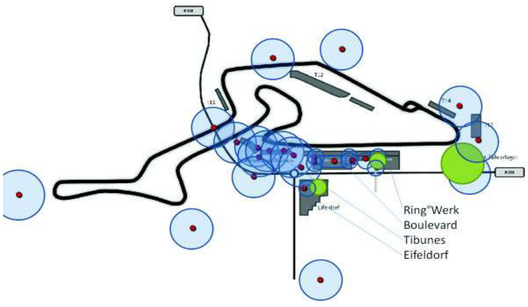

Figure 5: Sensor placement in target area

Each sensor independently measured unique records for each Bluetooth device. The privacy of the device owners was guaranteed by scrambling the MAC address but keeping the MAC-to-ID translation permanent for the weekend (see privacy issues in Section 3.1) so that the visits of a person can be linked into a chronological ordered sequence. Entries in a BtS logfile follow the logic:

```
[time];[sensor_id];[srambled device ID];[weakness]
```

Weakness measures the signal strength of the Bluetooth device within the reach of a BtS. This allows for definite assignment to a sensor. As our primary focus is on the analysis of global movements we are not interested in accurately position an object in space (e.g. triangulation) rather than in a high recall rate and a precise sensor assignment.

4.2 Preparing Bluetooth Data

The measured dataset contains more than 870,000 records produced by over 12,700 unique devices. Figure 6 shows the distribution of the aggregated absolute number of records over the two-day event period. It can already be seen that there are two local minima midday of both days. Furthermore, on Sunday the number of records tends to zero past 7pm whereas on Saturday the recorded numbers only slowly decrease till it reaches its minima at 4 am on Sunday, speaking for more night activity. In the data preparation step of our workflow this raw data is first being cleaned up. As described in detail in Section 3.2 initially we filter out duplicate entries. Two files contained erroneous log entries which have been deleted. We subdivided the measurement sequence of records by time spend at a location. For each device we defined time windows of arbitrary size. Short time readings at a sensor have been yet preserved. This way we do not lose any information; for example about people who leave a sensor area only for a very short time and return. In the second step of the temporal filtering process we merged all consecutive records of one device with identical sensor ID. For each group of records we created one representing entry with the number of data points it represents and the time difference between the max and min time inside a group (*group episode*).

Spatial filtering is then applied on the temporal filtered dataset. Based on the map of the Nürburgring we calculated the expected travel distance between all sensor locations. We decided to use a higher tolerance level in this case study due to the uncertainty introduced by the unknown extends of the BtS coverage areas. The route network was known. Similar to the arbitrary jumps of GPS signals when

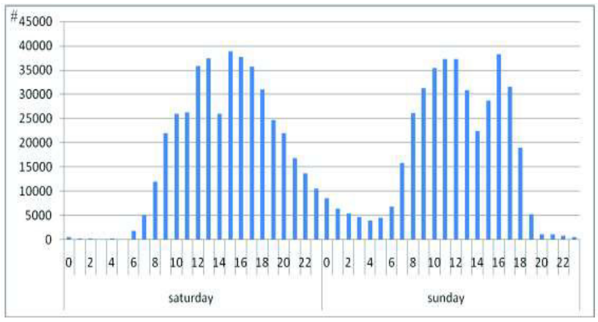

Figure 6: Distribution of aggregated counts of raw records (y) per hour (x) separated by day

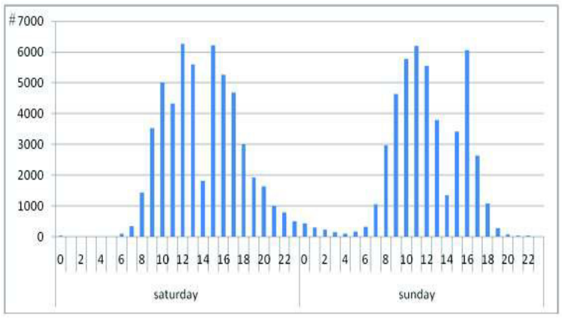

Figure 7: Distribution of aggregated counts (y) of records per hour (x) after spatiotemporal filtering

a device is not moving, there are arbitrary movements in the ordered list of records for each Bluetooth device. The probability of such false movements increases with the number of overlapping sensor areas. A number of scientific work deals with using this information for the localization of a device. We also use signal strength, travel distance, speed, and the group episode information to remove all spurious location changes. At last we aggregated all sensor data of the same attraction. Therefore, in a pre-step we assigned each of the BtS to one spatial attraction. For instance the scanners inside the Boulevard area have been assigned to the attraction Boulevard. For this case study we manually clustered the BtS but doublechecked with a clustering of sensors according to the type and occurrence of events monitored at the locations using Visual Analytics. Results show strong similarity. Again we calculated the min-max-time difference as the duration of stay at an attraction as well as the number of measurement points in this period. The latter is used to assure that a device did not leave the coverage area as this easily leads to false findings if a person continous to travel unnoticed and is thought to be staying at a certain attraction. Thus we require a minimal measurements-per-time ratio (1 record every 30sec).

Next, we generated trajectories based on the filtered and aggregated sequences of visits. We used the center point of each attraction for assigning geo-referenced positions (longitude, latitude) to trajectory elements. After cleaning and aggregating the raw data we have 100,000 records left in the database. Figure 7 shows the distribution of the purified data. The typical course of the original and the new hourly distribution is preserved. The new distribution shows clearer minima and maxima. Especially around 1pm and 2pm the number of records has significantly been reduced. This new class of data we call *attraction-based movement data*.

To gain a better understanding of the prepared dataset we use descriptive statistics and clustering. It turns out that the Bluetooth data sample is comprised of two subsets. The first subset of devices show quick attraction changes and long durations of stay while only causing low numbers of records. The probability of spot measurements is also increased in this group. The other group of devices shows a higher number of attraction visits and the number of records per stay is significantly higher than in the first group. Further analysis revealed that the dataset contains not only Bluetooth chips build into cellular phones or headsets but also navigation and car communication systems. Hence, we not only recorded movements and events of people but also of cars. For training a classifier we manually assigned 5 vendors of Bluetooth devices (like Nokia for cellular phones, TomTom for car navigation) to each class. The feature vector contained speed changes, number of attractions, measurement points per stay, length of the trajectories, and repeated visits. We used a simple classification (decision tree) to assign each vendor to one of the classes pedestrian or car with an accuracy of 90% in the training set.

In the following analysis we will ignore car movements for the moment. For reasons of business privacy we will only show some findings of the knowledge discovery process. The focus will be on the methods used to fulfill the tasks identified in Figure 4.

4.3 Analysis of Events

The focus of our analysis of attractions lies on answering questions that are event related. The previously prepared data now contains information about durations of stay and visits of attractions over time. For the case study we identified the most interesting spatial attractions according to the number of visits at these places. To answer the question of what is the duration of stay at an attraction we look at one sample attraction on the Nürburgring area. For every hour of day we calculated the number of devices recorded at one or more sensors covering the attraction. Recall that in our database we only have aggregations of records with a total duration of stay in minutes and a number of measurement points. Therefore, we divided the day into 24 timely ordered periods each of a length of 60 minutes. For these periods we count the number of unique devices which were within the 60-minute-frame at the attraction. For example a person who stayed only 25 minutes of an hour at the attraction will only count second group ([5, 10), [10, 30), [30, ∼)) for that particular hour.

Figure 8 shows an excerpt of our analysis for one area of the Nürburgring. For both racing days we see the number of persons spending less than 5 minutes at the attraction (orange), 5 to 30 minutes (blue), or more than 30 minutes (green). What can be seen is that long durations of stay increase towards 12 o'clock and in the afternoon. Particularly noticeable is the Sunday morning with continuously long periods of stay at the attraction. Both days show a minimum at 2pm. That is the time of the race when people are watching the race from the tribunes. On Sunday after the race we see only one peak at 4pm and afterwards the duration of stay decreases significantly so is also the remaining number of visitors on the area. The peaks of the second group (10-30) shortly before and after the qualifying (2-3pm Saturday) or the race (2-6pm Sunday) can be explained by the structural conditions. Visitors of the main tribune must enter and leave through the attraction we look at.

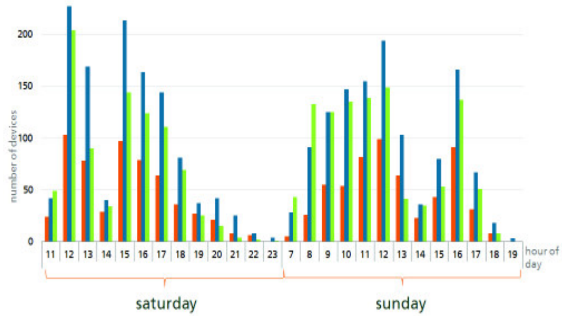

Figure 8: Number of visitors at an example attraction with a duration of stay clustered into 3 groups ([5,10),[10,30),[30,∼)

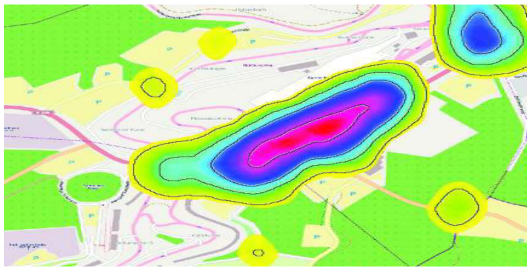

Figure 9: Heat map of the counts of devices at each attraction over the monitoring period; range from red - indicating high counts of devices - to yellow - low counts of devices

If we look at the counts of devices per attraction (adjusted to the number of overlapping sensor areas) the center shows a significant share of devices (see Figure 9). Spatially limited heat maps prove to be suitable for visualizing the aggregated counts of devices over time. We therefore buffered the centroids of the attractions and assigned the number of devices to each region and calculated a heat map. As many visitors enter the area from the north-east we also see a higher number of devices there.

4.4 Analysis of Movements

The objective of movement analysis in event monitoring was twofold: The first research question deals with the movement behavior of visitors depending on the entry point used to the area. The second question was targeted on discovering spatio-temporal patterns of visiting attractions or events in time in the main area of the Nürburgring Racing Track. Each recording of a Bluetooth device labeled pedestrian (see Section 4.2) was considered to be part of a daily trajectory of a person. If the person left the monitored area or turned the device on and off periodically we split the daily sequence into subsequences. The result of the trajectory generation was over 16,000 geo-referenced sequences of attractions created by 12,185 devices. Note that we used the centroids of the attraction geometries as position for the visit (step of

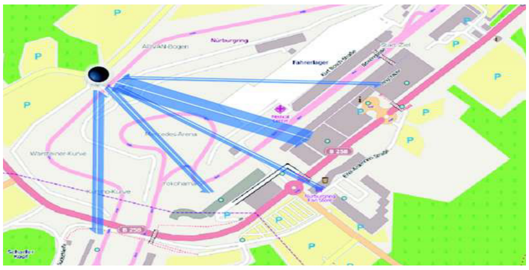

Figure 10: Directed aggregated moves (blue arrows) starting or ending in the northern entry point (blue point)

the trajectory).

One task was to investigate the flows between attractions over time. The local experts already have a feeling of what the major paths are in and out of the central areas. To this day they were lacking a weighting of the paths and general dependencies between entry points and attractions.

Mapping of movement flows is a standard cartographic technique to visualize flows of movements [17]. For massive movement data it is recommended to aggregate the moves first. Therefore, we count the number of moves () starting or ending for every pair of attractions (A, B). The resulting records (A, B,) are referred to as *aggregated moves*. The used visualization tool interprets the direction of aggregated moves by displaying arrows at the end of the connecting lines between the centroids of the paired attractions. The counts of moves correspond to the thickness of the arrows. Coinciding starts and ends do not exist in our setting as they are interpreted as duration of stay at one attraction.

To answer the question of how visitors move around the area with respect to the point where they entered the premises it is of most interest to the business decision makers to know the primary targets. Figure 10 visualizes the aggregated moves of one out of 10 entry points using a flow map. For this visualization we ignored all short stops at attractions on the path towards the actual target. The threshold for deciding the first target after entering the area was defined as a 10 minute stay at one attraction. It can be seen that the majority of people entering the premises through the marked point first visit the Boulevard area in the center of the map section.

For analyzing global flows we also utilize flow maps based on unfiltered aggregated moves between all sensor locations. An example is presented in Figure 11 which depicts flows on a low granular spatial scale. The blue lines represent actual movements of devices between sensors. The red colored arrows indicate the flows of devices between areas. The arrows in the center of the map indicate no favorite directions of flows but in general it can be depicted that most movements take place in the center. The biggest is inside the Boulevard (shopping area). This finding must be critically interpreted as our coverage is not intended to track people on a low level. Thus, there might still be overlapping. Nonetheless, we do see that the dependency assumed in our heat map interpretation in Section 4.3 can be substantiated. The flows from the north-east entry point are strong towards the center.

After looking at flows and dependencies of places we are

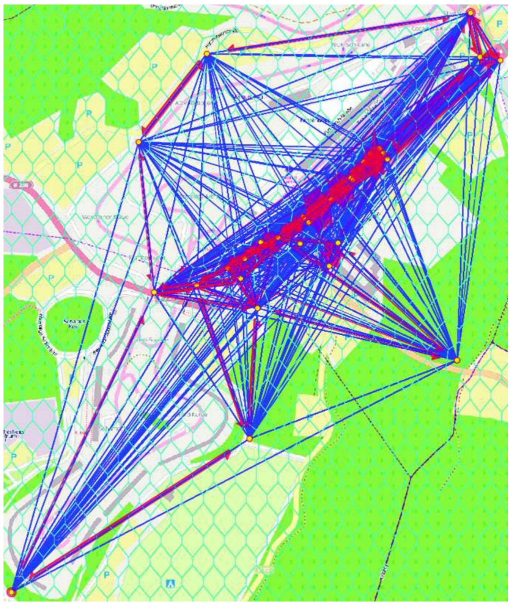

Figure 11: Flows of visitors as directed aggregated moves (red arrows) between all sensor areas

nowinterested in finding patterns in movement behavior. One way of interactively analyze large amounts of movement data is by using self-organizing maps (SOM). By varying the clustering parameters we can do the analysis with a different number of clusters or thresholds. Figure 12 shows an example of the SOM-clustering of movements. The grey lines represent subsequences of movements between sensor areas. The background color indicates the temporal cluster of movements. Saturday morning and night we have a purple like coloring indicating a similarity of movements. Sunday morning does not show a particular different movement behavior as it could be expected when thinking back to the durations of stay-analysis. Apparent is that both times around the races (qualifying and main race) show similar movement characteristics (dark yellow colored). The only difference is the duration as it is twice as long on Sunday. This is substantiated by the official schedule as the qualifying only takes place for 1 hour whereas the race lasts for 2 hours. Before the races takes place or when they end we see another type of movement (red colored) which can be interpreted as people going to the tribunes to watch the race or leaving. This is especially obvious on Sunday after the main race. The movement behavior lets assume that people are leaving the area massively. Apparently people arrive distributed over the morning hours so that there is no similar cluster found. On Saturday we do not see such a "collective-leaving". One reason for that could be that there are still many attractions and events (e.g. live music).

5. CONCLUSION

In this paper we have presented a systematic workflow for monitoring the movement behavior of visitors of major events using Bluetooth capturing technology. The workflow begins with planning the data collection with special focus on the placement strategy of the sensors. Shortly we have described our BlueTechSensor and addressed the pri-

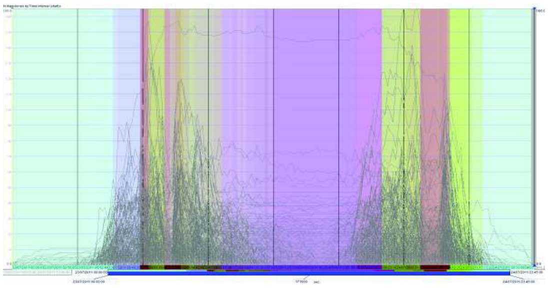

Figure 12: Clustering of situations according to people's moves between sensor areas using self-organizing-maps

vacy issue with using Bluetooth sensors technology. Next, we looked closely at the data preparation of Bluetooth data. This is because false treatment of errors or loss of information may easily lead to spurious results and misleading hypothesis in the Knowledge Discovery Process.

The interactive design of the mining process enabled many new analytical combinations. We have defined possible analysis tasks and presented several analysis approaches appropriate to solve these tasks. The proposed workflow proved suitable for answering the identified business questions. Decision makers gained new insights in the movement behavior of their "customers".

The advantages with Bluetooth capturing technology lie in their flexibility and robustness. They can be used in a mixed indoor and outdoor setting and produce one single and homogeneous database. Hence, the data must not be joined by complex algorithms. The main advantage is that Bluetooth sensor technology allows for uniquely tracking devices in space and time without any interaction.

More than 12,000 unique devices have been recorded over two days of the case study event. One surprise during this research was to find that over 55% of the recorded Bluetooth data was related to car movements. In the future we plan to investigate which Mobility Mining methods developed for GPS-based studies can be adapted and used for constructing mobility models and conducting research about the representativeness of Bluetooth samples and sample bias.

6. ACKNOWLEDGMENTS

We would like to thank the Nürburgring GmbH for all the support and the provided valuable expert knowledge.

7. REFERENCES

[1] G. Andrienko and N. Andrienko. Designing visual analytics methods for massive collections of movement data. *Cartographica*, 42(2):117–138, 2007.

[2] G. Andrienko and N. Andrienko. Extracting patterns of individual movement behaviour from a massive collection of tracked positions. *Technical Report TZI Technologie-Zentrum Informatik*, 42:1–16, 2007.

[3] G. Andrienko, N. Andrienko, P. Bak, S. Kisilevich, and D. Keim. Analysis of community-contributed space- and time-referenced data (example of flickr and panoramio photos). In *Proceedings of the IEEE Visual Analytics Science and Technology*, pages 213–214. VAST, 2009.

[4] G. Andrienko, N. Andrienko, and S. Wrobel. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations*, 9(2):38–46, 2007.

[5] N. Andrienko, G. Andrienko, and P. Gatalsky. Supporting visual exploration of object movement. In *Proceedings of the Working Conference on Advanced Visual Interfaces AVI*, pages 217–220. AVI, May 2000.

[6] P. Atteslander. *Methoden der empirischen Sozialforschung*. de Gruyter, Berlin, New York, 2000.

[7] J. Dykes and D. Mountain. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Computational Statistics and Data Analysis*, 43:581–603, 2003.

[8] F. Girardin, F. C. F. D. Fioro, and C. Ratti. Digital footprinting: Uncovering tourists with user-generated content. In *Pervasive Computing*, pages 36–43. IEEE, 2008.

[9] J. Hallberg, M. Nilsson, and K. Synnes. Positioning with bluetooth. In *Proceedings of the 10th International Conference on Telecommunications*, pages 954–958. ICT, 2003.

[10] S. Hay and R. Harle. Bluetooth tracking without discoverability. In *Proceedings of the 4th International Symposium on Location and Context Awareness*. LoCA, 2009.

[11] T. Hägerstrand. What about people in regional science? *Papers of the Regional Science Association*, 24:7–21, 1970.

[12] P. Jankowski, N. Andrienko, G. Andrienko, and S. Kisilevich. Discovering landmark preferences and movement patterns from photo postings. *Transactions in GIS*, 14(6):833–852, 2010.

[13] S. Jeffery, G. Alonso, M. Franklin, W. Hong, and J. Widom. Declarative support for sensor data cleaning. In *Lecture Notes in Computer Science*, pages 83–100, 2006.

[14] D. Keim. Scaling visual analytics to very large data sets. *WWW, http://infovis.uni-konstanz.de/ events/ VisAnalyticsWs05/pdf/ 03DanielKeim.pdf*, 2005.

[15] S. Leitinger, S. Pavelka, and M. Wimmer. Erfassung von personenströmen mit der bluetooth- tracking-technologie. In *Angewandte Geoinformatik 2010*. 22. AGIT-Symposium Salzburg, 2010.

[16] D. Phan, L. Xiao, R. Yeh, P. Hanrahan, and T. Winograd. Flow map layout. In *Proceedings IEEE Symposium on Information Visualization*, pages 219–224. InfoVis, 2005.

[17] T. Slocum, R. McMaster, F. Kessler, and H. Howard. *Thematic Cartography and Geovisualization*. Pearson Prentice Hall, New York, 2009.

[18] V. Trommsdorff. *Konsumentenverhalten*. Kohlhammer GmbH, Stuttgart, 2003.

[19] I. Vasiliev. Mapping time. *Cartographica*, 34(2):1–51, 1997.

[20] M. Versichele, T. Neutens, M. Delafontaine, and N. V. de Weghe. The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the ghent festivities. *Applied Geography*, 32(2):208–220, 2009.