Abstract

In this paper, proximity-based Bluetooth tracking is postulated as an efficient and effective methodology for analysing the complex spatiotemporal dynamics of visitor movements at mass events. A case study of the Ghent Festivities event (1.5 million visitors over 10 days) is described in detail and preliminary results are shown to give an indication of the added value of the methodology for stakeholders of the event. By covering 22 locations in the study area with Bluetooth scanners, we were able to extract 152,487 trajectories generated by 80,828 detected visitors. Apart from generating clear statistics such as visitor counts, the share of returning visitors, and visitor flow maps, the analyses also reveal the complex nature of this event by hinting at the existence of several mutually different visitor profiles. We conclude by arguing why Bluetooth tracking offers significant advantages for tracking mass event visitors with respect to other and more prominent technologies, and outline some of its remaining deficiencies.

Keywords

Bluetooth, tracking, mass event, geographical information technology, geographical information systems

1. Introduction

In the last few years the representation and analysis of large volumes of trajectory information of objects moving through geographical space has become a major topic of interest in research domains such as geographical information science (Ahlgqvist et al., 2010; Shaw, Yu, & Bomboim, 2008), computer science (Bogorny, Kuijpers, & Alvares, 2009; Orlando et al., 2007), visual analytics (Andrienko & Andrienko, 2007) and urbanism (Van Schaick & Van der Spek, 2008). This burgeoning academic interest has emerged as a result of the increased feasibility and affordability of collecting detailed data about spatiotemporal phenomena triggered by the widespread adoption of location-aware technologies. Past studies have focused on the movements of various kinds of objects including vehicles (Quiroga & Bullock, 1998), animals (Laube et al., 2007), bank notes (Brockmann, Hufnagel, & Geisel, 2006) and typhoons (Terry & Feng, 2010), but the majority of research has been devoted to human movement in different contexts and at various scales. Some examples are the movement of athletes on a pitch (Laube, Imfeld, & Weibel, 2005), tourists on a regional (Ahas et al., 2008) and local scale (Kemperman, Borgers, & Timmermans, 2009; O'Connor, Zerger, & Itami, 2005; Shoval & Isaacs, 2007a), and customers in a supermarket (Hui, Fader, & Bradlow, 2009). In these contexts, advanced tracking technologies complement more traditional qualitative methods, such as shadowing (Quinlan, 2008) and collecting travel diaries (Axhausen et al., 2002).

Within research on human behaviour, particular attention has been devoted to the collective behaviour of crowds at mass events such as street parades, festivals, public assemblies, sporting events, and exhibitions (Batty, Desyllas, & Duxbury, 2003; Helbing, Johansson, & Al-Abideen, 2007; Zeitz et al., 2009). Tragic events such as the recent stampede during the 2010 edition of the Love Parade in Duisburg show that it is vital to have accurate information on the spatiotemporal flow of visitors at
mass events. However, the collection of quantitative movement data by the use of advanced tracking technologies, with GPS (global positioning system) as the main example, in these contexts raises serious issues concerning feasibility. Distributing and recollecting tracking units to a sufficient number of individuals within the crowd is indeed a labour-intensive and expensive process. Additionally, GPS usage in dense urban settings and inside buildings engenders problems due to signal distortion.

To date, video surveillance on the basis of closed circuit television (CCTV) has been the customary approach to capture human motion in crowded environments. Technological advancement in the last decade has led to a large number of distinct research topics related to video surveillance including crowd density estimation, crowd behaviour monitoring, and face detection and recognition in crowds (Saxena et al., 2008). Despite substantial progress made in recent years, the use of video data to track individual movements within crowds remains a challenging task. First, dense packing and constant interactions among individuals make it difficult to unambiguously distinguish between individuals in a crowd. Second, limited viewing angles and changes of illumination and weather conditions highly complicate the visual recognition of individual spatial patterns from imagery. Finally, difficulties also arise with respect to the reconstruction of individual movements across multiple camera views, which is a necessity for larger study areas. Hence, current applications of video surveillance have achieved to record the spatiotemporal paths of only few objects in limited spatial environments (Dee & Velastin, 2007), severely limiting its use as a tracking technology in the context of mass events.

In response to these issues and given the ubiquity of Bluetooth-enabled devices such as mobile phones and personal digital assistants (PDA) carried around by their owners, Bluetooth technology has increasingly been suggested as a simple and low-cost alternative for the reconstruction of spatial behaviour (Bullock et al., 2010; Fatah & Mottram, 2007; Leitinger et al., 2010; Van Londersele, Delafontaine, & Van de Weghe, 2009; Versichele et al., 2010; Wasson, Sturdevant, & Bullock, 2008). Designed as an open and wireless communication technology by Ericsson in 1994, Bluetooth has become a well-known and implemented standard for wireless exchange of data between devices. If devices are set to be discoverable, the movement of the devices – and by extension their users – can be reconstructed by means of a unique Media Access Control (MAC) address that gets broadcasted in the discovery process. Because this fixed MAC address cannot be linked to any personal information such as names or phone numbers (contrary to the friendly name of the device), tracked individuals remain anonymous avoiding potential privacy infringements. Applications of Bluetooth tracking include the estimation of travel times on highway segments (Haghi et al., 2009; Wasson et al., 2008), public transport usage in Graz (Weinzerl & Hagemann, 2007), movement behaviour in a shopping mall (Millonig & Gartner, 2008), and functioning as an extension on social data gathered from Facebook in the Cityware project (Kostakos & O’Neill, 2008).

Because Bluetooth allows for non-participatory, unannounced and simultaneous tracking of a large number of individuals, it is particularly useful to study visitor flows at mass events. Despite its potential, only pilot studies using Bluetooth tracking at mass events have been reported (Leitinger et al., 2010; Van Londersele et al., 2009). This paper aims to significantly augment the current knowledge by reporting on a recent and comprehensive experiment using Bluetooth as a tracking technology. The experiment was carried out at the Ghent Festivities, one of the largest outdoor cultural events in Europe which lasts for 10 days and attracts around 1.5 million visitors. This setting offers a challenging test bed in terms of crowd size, duration of the event and spatial extent of the study area. The aim of this case study was to explore the potential of Bluetooth tracking for studying the spatiotemporal dynamics of visitors at mass events by highlighting a selection of analytical possibilities with the gathered data and showing the corresponding results.

The remainder of the paper is organized as follows. Section 2 gives a brief discussion of Bluetooth as a tracking technology. Section 3 describes the background and experimental design of the case study, and in section 4 we describe the pre-processing of the tracking data. In section 5, we present the results of this study. Finally, we contextualize these results, argue why Bluetooth tracking has the potential to become a valuable methodology for studying the dynamics associated with mass events, and outline some of its remaining deficiencies in section 6.
2. Bluetooth as a tracking technology

2.1. Working principle

Bluetooth is a short-range, low-power and open protocol for implementing Wireless Personal Area Networks (WPAN) between mobile devices. It operates on the Industrial, Scientific and Medical (ISM) frequency band (2.4 GHz). To minimize potential interference with other technologies using this crowded radio band (Wi-Fi, microwave ovens, etc.), Bluetooth employs a frequency hopping technique (Bensky, 2007). For technical details about the Bluetooth protocol and the discovery and communication process, the reader is referred to (Peterson, Baldwin, & Kharoufeh, 2006). In brief, the process can be summarized as follows. To set up a Bluetooth network (piconet) in which all devices follow the same hopping frequency to communicate with each other, a Bluetooth device first needs to be discovered by a master device during an inquiry phase. This master device enters the inquiry substate and starts transmitting inquiry packets, which triggers unknown devices within the detection range of the master device that are in the inquiry scan substate (‘discoverable’) to respond by transmitting an identification message containing their MAC address and class of device (COD) code. In a next phase, called the page phase, the master and slave devices synchronize their internal clocks to align their frequency hopping schemes and start communicating.

All Bluetooth scanners used in this study functioned as master devices, continuously broadcasting inquiry messages and listening for the responses of other discoverable devices that passed within their radio range. This way, mobile devices were detected without the need for an actual connection with the scanner. All further processing was done in the scanner. In comparison with other studies that did need a connection (Hay & Harle, 2009), the methodology used in this experiment is better suited for mass events where the participation of the tracked individuals should be minimal. Every time a device was detected, its MAC address, COD code, and the timestamp of the detection were registered. Additionally, the received signal strength intensity (RSSI) of the inquiry response was registered. This intensity value is inferred from the received power level with which the response packet was detected by the scanner and is theoretically negatively correlated with the distance between the scanner and the detected device (Hossain & Soh, 2007). Although it is technically possible to also register the friendly name of the device (which can be changed by the users), this significantly slows down the scanning process. Additionally, part of the tracked population might be identified by the inclusion of names or phone numbers. As this would raise serious privacy issues, the friendly name was not registered.

In theory, the RSSI values registered by different scanners can be used to calculate the position of a mobile device through multilateration (Bensky, 2007; Kelly, 2010). Although some studies have reported acceptable accuracies in indoor environments, they have also outlined problems with multipath fading due to obstructions in these environments (Feldmann et al., 2003; Zou & Pollard, 2006). Because of the complex environmental setting and the resulting unpredictability of the propagation of Bluetooth signals, positioning in this case study was done through the proximity principle, where the position of a detected mobile device is approximated to the position of the sensor by which it is detected (Bensky, 2007). This way, the path of a mobile device is reconstructed by means of a sequence of time intervals spent within the detection ranges of different Bluetooth scanners. The precision of the resulting trajectories ultimately depends on the detection range of the Bluetooth scanners, and on their number and distribution over the study area. In theory, Bluetooth devices are classified into three power classes. Class 1, 2 and 3 devices support theoretical ranges of 100 m, 10 m and 1 m respectively. The observed detection range, however, depends on how many obstructions the signal encounters between the scanner and the mobile device (buildings, furniture, clothing, people, etc.).

2.2. Equipment

Fig 1 shows the hardware components used in this Bluetooth tracking experiment. A Bluetooth ‘scanner’, as referred to in the remainder of this paper, is a combination of a computational unit running the scanning software and processing and storing the results (1), a power source (2), and a USB cable (3) to connect the computational unit with a Bluetooth sensor (4-5). The heart of the
computational unit was an ALIX motherboard (alix2d2, alix3d2), equipped with a 1Gb CompactFlash card for storing logfiles. The operating system was an adapted version of Voyage Linux (based on Debian 5, 2.6.32.15 kernel), and the scanning software, Gyrid (version 0.3.3, http://github.com/Rulus/Gyrid), was developed at our research group. It is a python implementation built around the BlueZ Bluetooth stack (version 4.63).

In order to control the detection range, different Bluetooth sensors were used. In this study, we used a combination of class 2 (D-Link DBT-122) and class 1 (Sena Parani UD-100) devices. These are respectively shown as numbers 4 and 5 in Fig 1. At most locations, class 2 devices without an external antenna were used. Where a larger detection range was necessary, a class 1 device with a replaceable antenna was used. Two types of omnidirectional antenna were available with gains of 3 (6) and 5 dBi (7). The higher the gain of the antenna, the higher the detection range was. Experiments showed that the detection ranges deviated considerably from the theoretical ranges in case of line-of-sight detections (up to 100 m for the class 2 device, and 300 m for the class 1 device with a 5dBi external antenna).

3. Background and experimental design

3.1. Description of the event and study area

The Ghent Festivities are a 10-day cultural and theatre festival taking place in the historic city centre of Ghent in Belgium. Besides stage events at the larger public squares (starting in the evening), there are also random small street acts during the day resulting in an almost continuous flow of events. The festival has grown from a small event in the sixties attracting only local residents, to a genuine mass event attracting visitors from larger distances. Because of the size and the openness of the event – most activities in the festival are free, and there are no explicit entrance or exit points – collecting quantitative data such as visitor counts is challenging. The resulting lack of quantitative data acts as a bottleneck for research into the spatiotemporal dynamics of visitor movements. Exemplary to this is the issue of calculating the total number of visitors that attend the festival, which has traditionally been estimated by using proxy variables such as the daily amount of waste collected in the centre and the number of tram or bus tickets sold (Verbeke et al., 2003). As such, estimations vary but the general consensus is that 1.5 million (non-unique) visitors attended the festival in 2010. Other than this rough figure and the use of video technology by the police department to give a qualitative indication of crowdedness, little is known about the general movement patterns of these visitors within and around the festival site, how long they stay at the festival, the number of days they visit the festival, their transport mode, etc. This paper aims to highlight the potential of Bluetooth tracking for understanding such crowd dynamics, and thereby significantly complement the (little) quantitative data that is currently available.

In order to manage such a large event in a relatively small area, a body of regulations has been prescribed by local authorities and other stakeholders (police department, fire department, festival organizers and residents). Specific rules apply in the officially documented 4.5 km$^2$ festivities zone, surrounding the historic part of the city centre. This zone comprises 11 public squares which act as major attractors because of on-stage performances, bars, food stands, fairs, etc. Because of the large size of the event, mobility issues regarding the movement of visitors to and from the event are also important. Consequently, the study area was defined larger than the festivities zone sensu stricto, and we also focused on the two main train stations in Ghent, and a park and ride facility in the southwest. In this park and ride facility visitors could park their car and subsequently take a tram to the city centre. A general overview of the study area is depicted in Fig 2. A summary of the different locations that were covered with a Bluetooth scanner is given in the next section.

3.2. Selection of scanner sites

Given the limited range of the Bluetooth scanners and the size of the event, a full coverage of the entire study area was impossible from a practical point of view. Instead, a careful selection of strategic
coverage sites was made after consultation with local policy makers and urban experts with the purpose of collecting as many significant individual movements as possible. The spatial distribution of the selected locations is depicted in Fig 2. In total, 22 locations were covered. In order to capture the main bulk of movements between the 11 public squares, a scanner was placed at each square (scanners 1-11). A selection of points of access around the festivities zone was also covered (scanners 12-19). One of these was located in the busiest shopping street of Ghent (scanner 12). The two main train stations of Ghent were also covered with a Bluetooth scanner at locations where the majority of train passengers needed to pass by. Finally, the tram station next to the park and ride facility in the southwest was covered by two scanners in order to be able to track all passengers. In the rest of the analyses, these two scanners are regarded as one scanner by merging all detections. All scanners were operating for the entire duration of the event, except for scanners 13-19 (points of access) which were only operational during the two first days of the festival. As mentioned above, the type of sensor was chosen in function of the desired detection range at each site. Inside the festivities zone, the scanners in the smaller squares (scanners 5, 6, 8, 9 and 11) were equipped with a class 2 Bluetooth dongle and the scanners in the larger squares (scanners 1, 2, 3, 4, 7 and 10) with a class 1 Bluetooth dongle. The type of antenna for the latter dongle was chosen in function of the size of the square that needed to be covered (the larger the area, the higher the gain). As such, scanners 1, 2, 3, 4 and 7 used a 5 dBi external antenna, and scanner 10 a 3 dBi external antenna. All scanners outside of the festivities zone used a class 2 Bluetooth dongle.

4. Pre-processing

The raw data consisted of log files on the different scanners having the following format: timestamp of detection, MAC address of the detected device, COD code of the detected device, RSSI of detection. After merging the logfiles of the different scanners, the dataset consisted of 263,680,889 loglines. In order to obtain a compressed dataset, the scanners were programmed to create a second set of log files during the scanning process in the following compressed format: timestamp, MAC address of the detected device, COD code of the detected device, in/out/pass. A buffer time of 10 seconds was used to create detection time intervals from the detection time points. ‘In’ was written when a device entered the detection range of the sensor, and ‘out’ was written when the device left the range. ‘Pass’ was used for solitary detections with no prior or later detections within 10 seconds. In this way, the dataset was compressed by 91% to 23,889,850 loglines. Fig 3 shows an extract of both types of logged data. Over the entire duration of the event, 102,467 unique devices were detected over all covered locations. The majority of these were detected at least once inside of the festivities zone (88,763 devices or 87%).

The remainder of the analyses were done using a geographical information system for moving objects (GisMo), implemented in java and developed at our research group.

Because this study uses mobile devices as a proxy for detecting the movements of their mobile users, we have analysed the different types of devices that were detected by the Bluetooth scanners. As mentioned above, detectable devices respond to the inquiry call of the scanners by transmitting their class of device code. This code directly corresponds to certain types of devices. The scheme that is thereby followed was created by the Bluetooth SIG (Special Interest Group) and distinguishes between six major classes (Audio/Video, Computer, Imaging, Network Access Point, Peripheral and Phone). Additionally, some devices do not publish their COD. The minor classes give more information about the specific kind of device within each major class. They distinguish, for example, between cell phones and smartphones, laptops and desktops, etc. Taking both major and minor classes into account, the dataset was divided into three general kinds of devices: phones (91%), handsfree car kits (7%) and other devices (2%). The car kits represent cars, while the phones represent persons. The other devices could not be directly linked to a (moving) person. Because this study focuses on the movements of persons, only phones were withheld in the remainder of the analyses. The resulting dataset consisted of 80,828 phones or visitors that were detected at least once in the festivities zone during the event.

Because the Ghent Festivities span multiple days, some of the visitors visit the event for more than one day. Consequently, there is a need for a distinction between visitors (the person itself modelled as
a moving object with a fixed identifier) and their visits (modelled as trajectories generated by the moving object). Because of the continuous flow of events, a fixed time point at which to split trajectories was not available. Accordingly, an algorithm using a maximum time gap was developed. Trajectories containing gaps between detections with a duration of more than this maximum time gap were split into two subtrajectories. After careful exploration of the dataset, this parameter was set at five hours. As a consequence, the 80,828 detected phones/visitors that were detected at least once inside of the festivities zone were responsible for 152,487 trajectories/visits in total.

To make predictions about the entire visitor population, observed numbers of detected phones/trajectories need to be extrapolated to estimated numbers of visitors/visits. In order to do this, the detection ratio (the percentage of visitors that gets detected by means of a mobile device with a visible Bluetooth interface with respect to the entire population) needs to be known. To this end, we compared visual counts of passing people in smaller passageways with the number of Bluetooth devices detected during the same time period by a mobile scanning setup (laptop + class 2 Bluetooth dongle). In order to preserve privacy, we did not record any images in this process. Ten such experiments were conducted at eight different locations, each of them over a time period of 15 minutes. Averaging the results yielded a general estimation of the detection ratio of $11.0 \pm 1.8\%$. This ratio can be used to roughly estimate total numbers of visitors/visits from numbers of detected devices/trajectories. Care should be taken, however, when generalising insights gathered from tracked visitors to the entire visitor population. This is due to the possibility of sample bias resulting from different detection ratios in different population segments, which is further discussed in section 6.3.

5. Results

In the remainder of this section, we will highlight some of the analytical possibilities of Bluetooth tracking data in the context of mass events by showing a selection of concrete results from the Ghent Festivities.

5.1. Total number of visitors and visits of the event

Given the detection ratio of $11.0 \pm 1.8\%$, we can estimate the total number of unique visitors and visits (or ‘non-unique’ visitors). This leads to an estimation of around 735,000 unique visitors (minimum: 630,000, maximum: 880,000) and around 1.4 million visits (minimum: 1.2 million, maximum: 1.7 million). The estimation of 1.5 million visitors made by the city department apparently lies close to the estimation made with the gathered Bluetooth data.

5.2. Total number of unique visitors per location during one day

The aggregated number of phones that is detected over a certain time interval on a Bluetooth scanner gives an indication of the total number of visitors that passed within the detection range of a sensor during that time interval. Because of the strategic locations of the Bluetooth scanners, the tracking data can be used to estimate the number of visitors that visit each location in and around the festivities zone. This is illustrated for the third event day of the festivities in Fig 4, which shows the number of phones detected at each public square. It is immediately clear that some squares attract considerably more visitors than others. Square 3, for example, attracts more than three times as many visitors than square 11 (7,639 vs. 2,347 detected phones) over the course of a day. In general, the public squares in the southwest of the festivities zone attract most visitors, reflecting the fact that the largest and most popular squares are situated there as well as the main points of access. By multiplying these counts with the previously defined detection ratio, it is possible to make a rough estimation of the total number of visitors that visited each location. The busiest square for example attracted around 70,000 visitors on that specific day.

5.3. Varying number of visitors in the entire festivities zone over time (day, hour)

By aggregating the number of detected phones over regular time periods, it is possible to calculate the crowdedness at a location over time. Additionally, a number of locations can be generalized into one
zone to estimate zonal crowdedness. For this case study, we calculated the crowdedness in the festivities zone at a time resolution of both one day and one hour. This is illustrated in Fig 5. Aggregating over one hour time windows results in a very smooth curve with sharp troughs in the morning (usually around 7 am). The peaks are also usually sharp and situated around 11 pm except for days 2, 5 and 9 where a broader peak in the late afternoon is observed. These correspond to two Sundays and the national day of Belgium (21/07), and these days are known to attract more daytime visitors (such as working couples with children). As a result, the sharp peaks around midnight do not appear because of the relatively larger crowdedness earlier in the afternoon. The three busiest days are immediately visible: the fourth day is the most crowded with almost 10,000 detected phones or 90,000 unique visitors in the festivities zone between 11 and 12 pm. The little peak during the build-up of the first day is due to the opening parade on the first day attracting specific visitors leaving immediately afterwards. To aggregate over daily periods, one should carefully consider how to define a day. Looking at the hourly crowdedness, it is clear that it does not make any sense to define days starting and ending at midnight because that is generally the most crowded period of the day. Doing so would cause the Bluetooth observations to be segmented by unnatural breaks. Consequently, we have considered the starting point of an ‘event day’ to coincide with the on average least crowded moment of a day, i.e. 7 am. The daily aggregates again show the three busiest days with day 4 peaking at almost 20,500 detected phones or 190,000 visitors.

5.4. Changing distribution of the crowd in the festivities zone over time

Although the varying crowdedness in the festivities zone already offers valuable insights for all stakeholders of the festival, it does not take the spatiotemporal dynamics of the crowd within the zone into account. As a first and general approach to shed light onto these dynamics, we have examined the changing distribution of the crowd over the different public squares in the festivities zone over time. This was done by slicing the data into time periods of one hour, and counting the number of phones detected at each public square during each time period. Subsequently, these numbers for each time period were summated over the 10 event days. The resulting chart in Fig 6 shows the changing distribution of the crowd in the festivities zone during a general event day. A clear trend is visible where the crowd is evenly distributed over the entire centre for most of the day, but condenses to a smaller set of squares during the night. Especially locations 8, 9 and 10 seem to attract a specific night audience (together they attract around 50% of the entire audience between 5 and 6 am). This well-known phenomenon is caused by visitors gathering in this area after midnight and staying until the morning. Although it is not shown in this figure, this trend was visible during every day.

5.5. Returning visitors

Because of the fixed MAC address assigned to each device, it is possible to investigate how many detected visitors visit the event for more than one day. This was done by counting the number of trajectories (visits) per phone (visitor). If a trajectory comprised more than one day, only the first day was considered to prevent potential errors from visitors staying longer than 7 am and being registered as two-day visitors. The result is shown in Fig 7. There is a dominance of one-day visitors with respect to returning visitors (65% vs. 35%). Furthermore, the share of returning visitors decreases with an increasing number of visit days (from almost 14,000 phones for 2 days to 131 phones for 10 days).

The share of returning visitors at each public square separately was used to gain insight into the tendency of returning visitors to visit the same squares in the festivities zone. The used procedure was the same as described in the previous paragraph. The result is depicted in Fig 8. Squares 9 and 10 are characterised by a higher than average degree of returning visitors (21.6 ± 0.2% for these two squares vs. 13.5 ± 3.4% for the other squares), whereas locations 5 and 11 exhibit a lower degree. The anomalous observations at locations 9 and 10 can again be attributed to their functionality as night hubs of the event. Apparently, visitors of this area seem to return for more than one day more often than the average visitor at other locations. Square 5 is a smaller square in the centre of the festivities zone, and a significant portion of the more general public (consisting largely of one-day visitors) needs to pass this location to walk around the centre. The result at square 11 is influenced by its proximity to the border of the festivities zone and hence to movements that are not related to the event.
5.6. Transportation mode

Bluetooth tracking data can additionally offer semantic data concerning transportation mode by careful selection of the sites covered with Bluetooth scanners: train users can be distinguished in a train station, tram users in a tram stop, car users in a parking lot, etc. In this case study, train users were detected by two scanners in the two train stations of Ghent. Visitors making use of the park and ride facility outside the centre were detected at the main tram stop upon entering the tram taking them to the centre. Users of other transport modes were harder to track because of a lack (pedestrians, cyclists) or an excess (car and bus users) of fixed departure points. Despite the fact that not all transport modes were detected, these two groups do represent visitors coming to Ghent from larger distances. As an illustration, we calculated their relative shares within the total visitor population for the different days of the event. The result is shown in Fig 9. The relative share of tram users clearly varies more over time (3-7%) than the share of train users (5-6%). Remarkably, the share of tram users follows the exact same trend as the daily visitor counts in Fig 5. Apparently, there is a systematically larger share of visitors making use of the park and ride facility and going to the city centre by tram on the busier days. This is interesting as it could indicate a change in composition of transport modes of the public over the different days.

5.7. Visit duration

Technically, there are two methods of estimating the duration of a visit to the Ghent Festivities. The first method involves measuring the total duration of detection by generalising the detections over the different public squares into one zone. Because of the incomplete coverage of the study area, however, gaps occur in between the detections. Algorithms can merge all co-located detection intervals within a certain time threshold of each other, but the correct choice of this threshold is problematic without accurate and systematic comparisons with the ground truth. Consequently, we chose to estimate the duration of a visit in a second way. Because some access points around the festivities zone were covered during the first two days, the duration between two detections at an access point separated by one or more detections in the festivities zone can also be used as a proxy for the duration of a visit. The resulting distribution in Fig 10 exhibits a wide spread around its median value of 3.5 hours. The majority of this subpopulation of visitors seems to visit the event for only a few hours, while the large positive skew shows that other visitors stay much longer in the festivities zone (1,036 trajectories or nearly 11% of the sample stay for at least 7 hours). Because there seems to be a large variability in visit duration among (certain types of) visitors, the median value should be regarded as a very rough generalisation of this distribution.

5.8. Flow analysis

Although our methodology can only analyse flows of visitors carrying discoverable devices, the discovered patterns and trends can aid stakeholders to make well-informed decisions regarding crowd-management and security in general. By making a time series of these flow diagrams, it is possible to investigate the time-dependency of visitor flows and link them to factors that potentially influence the movements such as the time and order of performing artists at the different locations. In this paper, we limit the flow analysis to the third event day (19/07 – 20/07) and study the flows without an in-depth look at their influencing factors. A flow from location A to location B is defined as the number of mobile devices that is subsequently detected at A and B over a maximum time period of 30 minutes, without being detected at any other locations in between. Fig 11 shows four characteristic snapshots of the visitor flows during this event day. Regular time periods of 30 minutes were used to generate the different snapshots, and flows were attributed to their respective time period according to their departure time. The first snapshot (a) shows that the majority of the flows in the afternoon take place in the southwest of the festivities zone. A clear west-east and north-south axis is visible. Later at night (b), movements seem to condense along the west-east axis. The large flow from location 2 to 1 – which was also the largest flow of the entire event day – is related to the performance of a popular artist at location 1 at 11 pm which temporarily caused square 1 to be closed to prevent overcrowding. Around 3 am (c), flows pointing to the night hub in the northeast (squares 8, 9 and 10) dominate. Later in the morning (d), there are still considerable flows to this area but flows moving back to the
southwest dominate more. Since there are no programmed events at this time in the southwest, these flows largely consist of visitors leaving the event.

6. Discussion and conclusion

The discussion is organised as follows. First, the added value of Bluetooth tracking in the context of mass events is discussed by reviewing the results generated from the Ghent Festivities. Next, we describe why Bluetooth has a high potential as a tracking technology in this specific niche of use-cases. We conclude by highlighting some remaining issues and weaknesses concerning the methodology, and make suggestions for further research.

6.1. The added value of Bluetooth tracking for mass events

We were able to generate a huge dataset of movement trajectories with a limited number of scanners. Around 100,000 devices were detected over 10 days. This is equal to the amount that was detected by 450 nodes over 4 months in the Cityware project, based in the United Kingdom (Kostakos & O’Neill, 2008), and about 5 times higher than the amount detected by 16 scanners over 2 days in the Donauinselfest (Leitinger et al., 2010). Comparing the total number of detections in the raw data with the number of detected devices, we can see that each device was on average detected 240 times by one or more of the scanners. This is a first and strong indication that the employed technology and system are able to manage the huge flow of information generated by tracking visitors of very large mass events such as the Ghent Festivities.

In order to determine the true representativeness of the gathered tracking data, the extracted information needs to be confronted with reality. One such figure is the number of devices / trajectories in the database as representatives of visitors / visits, which can be estimated by means of a detection ratio. Our detection ratio estimate of 11% seems higher than the ones of 7% found in two other studies (O’Neill et al., 2006; Weinzerl & Hagemann, 2007). The increasing penetration of the Bluetooth technology in recent years might be responsible, but ultimately there is a need for more automated and accurate ratio estimations if Bluetooth tracking results are to be extrapolated to an entire population of moving individuals. Nevertheless, the estimated number of visits calculated from our dataset seems to closely correspond to the estimation made by the city department (1.4 vs. 1.5 million respectively). This indicates that the detection ratio is reasonably accurate and that there is no significant problem with devices somehow escaping the detection ranges of our scanners (O’Neill et al., 2006).

Both the daily and hourly variations of crowdedness in the festivities zone showed a smooth profile with recurring peaks and troughs. The crowd density profiles for the separate public squares (not shown) followed similar patterns, highlighting the suitability of Bluetooth tracking for studying spatiotemporal crowd density variations at specific locations or zones. Apart from distinguishing between crowded and less crowded days, we were also able to identify three days characterised by a relatively larger share of afternoon visitors.

Large events such as the Ghent Festivities are known to attract a wide spectrum of visitors with different tendencies and expectations. This is noticeable in the rest of the analyses. While the public squares in the southwest of the festivities zone attract the main share of visitors (Fig 4), the squares in the northeast clearly function as a night-hub of activities (Fig 6). Additionally, visitors of this night-hub seem to return more regularly than visitors that avoid this area (Fig 8). Despite the clear dominance of one-day visitors (Fig 7), the large size of the event makes that the relatively small share of visitors that attend the event for several days nevertheless constitutes a group of considerable size.
As an example, 2,173 phones were detected during 7 or more days, which represent a group of around 20,000 visitors.

It was possible to make an estimation of the average visit duration, but the resulting distribution (Fig 10) again pointed to the existence of two (or more) types of visitors: a majority that only visits the event for a few hours and a relatively smaller group that stays considerably longer. As a result, the median value of 3.5 hours does not reflect all visitors and probably lies closer to the average visit duration of the majority of the public than that of the longer staying visitors.

Because of the coverage of two train stations and a large park and ride facility outside the centre, the transport mode of roughly 10% of the visitors could be determined (either train or tram after parking the car). A preliminary analysis showed that the share of visitors using the park and ride facility was disproportionately higher on days where more visitors in general attended the event. This could reflect a temporal change in composition of transport modes, but more research and a higher coverage of transport modes are necessary to confirm this hypothesis.

A concise flow analysis exposed a typical spatiotemporal pattern over the course of one event day that was in agreement with on-field experiences of the city and police department. Together with qualitative estimations of crowd density by means of video cameras distributed over the festivities zone, they are especially valuable for assisting crowd management specialists.

6.2. The Bluetooth niche

If a tracking technology is to be suitable for studying spatiotemporal dynamics of crowds at mass events, it needs to comply with certain requirements. In this section, we explain why Bluetooth tracking can potentially fill this niche by comparing with two other technological candidates for this use-case: GPS and mobile positioning of cell phones (Ahas et al., 2008; González, Hidalgo, & Barabási, 2008).

First and foremost, Bluetooth tracking offers the ability to track a very large number of individuals in an easy and relatively inexpensive way. A large population sample is often necessary to understand the complex dynamics involving crowd movements at mass events. Other tracking technologies, such as GPS, only reach a small subset of individuals because of the low penetration rate in the general audience (Ratti et al., 2006). Additionally, they involve the tracked individual in a direct way (e.g. by distributing and later recollecting logging units). This renders the tracking process labour-intensive, and makes it prone to possible bias since individuals might behave differently because they are aware of being tracked or because certain population segments might be more inclined to cooperate with such experiments and hence be overrepresented in the resulting dataset. Mobile positioning datasets are usually even larger than Bluetooth tracking datasets, but cooperation with mobile operators has proven to be difficult (Ahas et al., 2008). The ultimate cost of the technology will depend on the number of scanners (around 200 euro’s per device) used, but will generally be fairly low if expressed per tracked individual.

Secondly, Bluetooth tracking has a clear advantage over other technologies because of its ubiquitous applicability in indoor as well as outdoor environments. Because of the freedom with which Bluetooth scanners can be placed in indoor environments, it is possible to follow individuals at room level inside buildings. Signal deterioration makes this practically impossible with GPS, and very challenging with mobile positioning.

Finally, the limited detection range of Bluetooth sensors makes the resulting trajectories more accurate than those in mobile positioning datasets. As mentioned above, the actual range of our Bluetooth scanners varied from around 10 to 100 meters, depending on the used sensor. Location estimations in mobile positioning data typically have an accuracy of a few hundred meters in urban settings (Ahas et al., 2007), which is insufficient for distinguishing between locations that are less than 100 meters apart (such as locations 8 and 9 in the festivities zone).
6.3. Remaining issues and suggestions for future research

Although we have demonstrated that Bluetooth tracking is able to deliver valuable information to stakeholders of mass events, there are some remaining issues that need to be addressed. One of the prime issues is the possibility of biased results by oversampling certain segments of the total population of individuals (Rice & Katz, 2003). Adolescents with a higher education might indeed carry more Bluetooth-enabled devices than elderly people, while young children will probably never be detected. The potential difference in Bluetooth usage among different audiences might significantly influence generated insights. Accordingly, more research is needed into the use of discoverable Bluetooth-enabled devices by different population segments in order for Bluetooth tracking to evolve into a technology delivering accurate and reliable information to policy makers. The estimated ratio of 11% is sufficient for making rough extrapolations to the entire population of visitors, but a more systematic way of calculating the percentage of the population being tracked will be necessary for more reliable extrapolations in the future. Additionally, the possible influence of time and space on the detection ratio needs to be investigated.

Focusing on this case study, several analyses have exposed the existence of multiple profiles of visitors. More research is needed into which profiles exist, and how they can be distinguished from each other. This could either be done with the current dataset containing only locational information (e.g. the time that a visitor spends on a certain public square might be used to define different visitor profiles), or with a new dataset enriched with socio-economical information about the tracked individuals. In the latter case, we might statistically prove the likely assumption that there are two main visitor profiles (adult visitors in the afternoon and adolescents during the night) and study the influence of socio-economical attributes on movement behaviour.

Transportation mode detection was adequate for train and tram users, but the scope should be extended to include other transportation modes as well (cars, pedestrians, cyclists). More research is needed into the influence of the transportation mode on visitor behaviour, such as the duration of a visit or the movement patterns between the public squares.

The tentative flow analysis already suggested certain patterns, but further research should be devoted to finding representative patterns in the order in which the different squares are visited. Sequence alignment methods seem the most likely candidate for extracting these patterns (Shoval & Isaacson, 2007b; Wilson, 1998) but they might need further modifications to handle the spatiotemporal complexity of Bluetooth tracking data gathered at large mass events such as the Ghent Festivities.

Because of the difficulty of directly correlating the RSSI of a detection with the distance between the sensor and the detected device, we have used the proximity principle to generate trajectories from the detection data. Fine multilateration of the signal strengths registered on different sensors to calculate an accurate location seems unrealistic, but a rough multilateration might be possible under certain circumstances. In this way, a continuous crowd density over a public square might be calculated while this is not possible using the proximity principle.

References


Fig 1. Bluetooth hardware used in the tracking experiment: computational unit (1), power source (2), USB cable (3), class 2 Bluetooth dongle with internal antenna (4), class 1 Bluetooth dongle (5), and 3 dBi (6) and 5 dBi (7) omnidirectional antennas.
Fig 2. Overview of the study area and location of Bluetooth scanners.
Fig 3. Extract of logged data showing the raw time-point detection data (top) and the compressed time-interval data (bottom). This example shows one Bluetooth device (MAC address 20:21:A5:45:40:40) being detected 5 times on 20/07/2010 between 17:53:38 and 17:53:55 (CEST: Central European Summer Time). The buffer time of 10 seconds causes the raw data to be split into two separate detection time intervals (in/out). The COD code of the device (5898756) shows that this was a regular cell phone.

Fig 4. Aggregated number of detected phones on the third event day of the Ghent Festivities (19/07/2010 11 am until 20/07/2010 07 am) at the 11 public squares and the main access point (12).
**Fig 5.** Daily (event days starting and ending at 7 am) and hourly number of detected phones over the entire festivities zone as an indicator of crowdedness. Solid vertical gridlines point to midnights, dashed vertical gridlines are plotted every 4 hours.

**Fig 6.** Distribution of the detected crowd over the different public squares in the festivities zone over time (hourly aggregates, summated over the 10 festival days).
Fig 7. Share of detected phones in function of the number of visit days.

Fig 8. Share of phones detected on more than one day at the different public squares inside of the festivities zone.
Fig 9. Relative share of train and park and ride users on the different days of the event.

Fig 10. Histogram of the duration of a visit to the Ghent Festivities (class-width: 15 minutes, sample size: 9,648). The average value is depicted by the solid line (3 hours, 53 minutes and 58 seconds), the median value by the dashed line (3 hours, 32 minutes and 42 seconds).

(a) 19/07/2010 4 pm – 4:30 pm
(b) 19/07/2010 10 pm – 10:30 pm
(c) 20/07/2010 3 am – 3:30 am
(d) 20/07/2010 4:30 am – 5 am
Fig 11. Four snapshots of visitor flows in the festivities zone during periods of 30 minutes. The direction of the arrow indicates the direction of the flow, the width of the arrow indicates its size. The widths of the arrows are normalised to the size of the largest flow during each time period separately (indicated by the number next to the widest arrow). The numbers in the circles refer to Fig 2. Locations 8 and 9 were generalised into one zone ‘89’ for easier representation. Movements taking less than 30 seconds or longer than 30 minutes were discarded, and the remaining movements were allocated to the time period containing their departure time.