Comparing three types of real-time data collection techniques: counting cameras, Wi-Fi sensors and GPS trackers

Daamen Winnie[1]*, Yuan Yufei[1], Duives Dorine[1], Hoogendoorn Serge[1]

[1]Delft University of Technology, Department of Transport & Planning Stevinweg 1, 2628CN Delft, The Netherlands

w.daamen@tudelft.nl; y.yuan@tudelft.nl; d.c.duives@tudelft.nl; s.p.hoogendoorn@tudelft.nl

Abstract: Nowadays, many large scale events are organised in urban areas that are not designed to accommodate these large visitor flows, as not only the sheer number of people is larger than anticipated, but also their behaviour is different. For a timely deployment of crowd management measures, continuous information on the real-time traffic state on the event terrain is necessary. In this paper, we focus on local and global real-time measurements, collected during the SAIL Amsterdam event using counting cameras, Wi-Fi sensors and GPS trackers. The counting cameras gave accurate local counts, but did not have route information. The latter was given by Wi-Fi sensors and GPS trackers. However, the penetration rate for Wi-Fi sensors was low, so to get accurate data these sensors can only be used with large visitor numbers. GPS trackers gave detailed, but not continuous information. Based on the joint data sources, counts, travel times, density and flow are estimated. This way, we could provide the real-time information necessary for crowd management.

Keywords: Real-time crowd monitoring, Counting camera, Wi-Fi sensor, GPS tracker, Travel time, Density, Flows

1 Introduction

During large scale events in urban areas, many visitors gather in areas that typically have been designed to accommodate large pedestrian flows in other circumstances, e.g. rush hour. Moreover, the movement behaviour of visitors during events is different from e.g. commuters (the most widely investigated type of pedestrians), and depends on the characteristics of the event. Dedicated traffic management plans (for pedestrian traffic, but also for other traffic flows to cope with arrival and departure of visitors) are developed in order to keep the city functioning during such large scale events. These crowd management plans focus mostly on the event terrain and consist of lists of measures to alleviate potential (capacity) bottlenecks. The question is when to deploy which measures. Often, event organisers and crowd managers have a lack of information on the real-time traffic state. Typically, information is gathered by staff 'on location', but this information is scarce and not always objective. Clearly, this lack of information hampers efficient deployment of crowd management measures. In this paper, we present three ways to collect information on crowds in real-time, and assess the quality of the gathered data.

Data categories can be distinguished according to different criteria [1]. Local measurements cover traffic dynamics at specific locations in the network (e.g. an entrance), while global measurements refer to the traffic flows in the network. In addition, traffic measurements can be classified into microscopic and macroscopic perspectives. The microscopic perspective relates to movements of individual persons, i.e. trajectories, while the macroscopic perspective deals with movements of traffic flows, in which individuals cannot be distinguished. Examples of data for each category are shown in Table 1. Table 2 shows an overview of the corresponding data collection methods as described in literature.

Table 1 Examples of data for the distinguished categories, based upon [1]

		Measurement objective	
		Local	**Global**
Measurement perspective	**Microscopic**	Trajectories	Routes Travel times
	Macroscopic	Counts Speed distribution Flow patterns over time	Densities (distribution) over network OD matrix Travel times

Table 2 Classification of data collection methods for pedestrians, based upon [1]

		Measurement objective	
		Local	Global
Measurement perspective	Microscopic	Video [5] Time-lapse [5] Infrared [6] Laser [7]	Stalking [4] Questionnaires [4] GPS [8] Bluetooth, Wi-Fi [6] Mobile phone data [9]
	Macroscopic	Manual counts [2] Video [5] Time-lapse [5] Infrared [6] Laser [7]	Aerial observations [3] GPS [8] Bluetooth, Wi-Fi [6] Mobile phone data [9]

In literature, different data collection techniques have been applied for various pedestrian research purposes. These techniques can be further distinguished between manual and automated data collections. For examples, manual counts [2] and aerial observations [3] are typically used to obtain macroscopic measurements. However, for microscopic measurements, manual methods are a challenge, since it is difficult to track individual movements. Human observers need time-lapse and video for post-processing, or rely on stalking and questionnaires [4]. Manual data collection techniques are usually labour-intensive and time-consuming. Moreover, human errors might be embedded in the collected datasets. Technical advancement enables automated data collection for pedestrians. In this category, both image-based sensors (time-lapse and video) [5] and infrared or laser sensors [6,7] are spatially-fixed, and thus they can provide local measurements. With respect to global measurements, technological progress has resulted in several types of methods, including GPS [8], Bluetooth and Wi-Fi sensors [6] and mobile phone data [9].

For large scale events in urban areas an overview of the traffic state is required, which can be based upon global measurements. However, to get detailed information on the location of bottlenecks and to get an idea of the heterogeneity of the pedestrians, local, microscopic measurements are a welcome addition. In this paper, we therefore focus on both types of data, collected during the SAIL Amsterdam event [10] using counting cameras, Wi-Fi sensors and GPS trackers.

The paper starts with an overview of the data collection during SAIL 2015. In section 3 the raw data collected using counting cameras, Wi-Fi sensors and GPS tracks are described. Section 4 shows a comparison between the traffic states predicted by the different data sources, based on flows, travel times and densities. The paper ends with conclusions.

2 Data collection at SAIL Amsterdam

SAIL Amsterdam is the largest free nautical event in the world. Every five years, more than 600 ships navigate along the North Sea Canal before mooring in and around the IJ-port in Amsterdam. In the edition of August 2015, taking place from August 19 until August 23, more than 2 million people enjoyed the event at the SAIL terrain. Although many activities have been organised in and around the city centre of Amsterdam, we focus on the IJ-port area, where most tall ships were moored. Fig. 1 shows this area and the so-called orange route, a walking route along the tall ships that was actively advertised by the municipality.

Data have been collected using 8 counting cameras and 20 Wi-Fi sensors; their locations are shown in Fig. 1. In addition, 324 GPS trackers have been distributed among visitors to register trajectories. In the following, more details are provided regarding the raw data from the individual data sources.

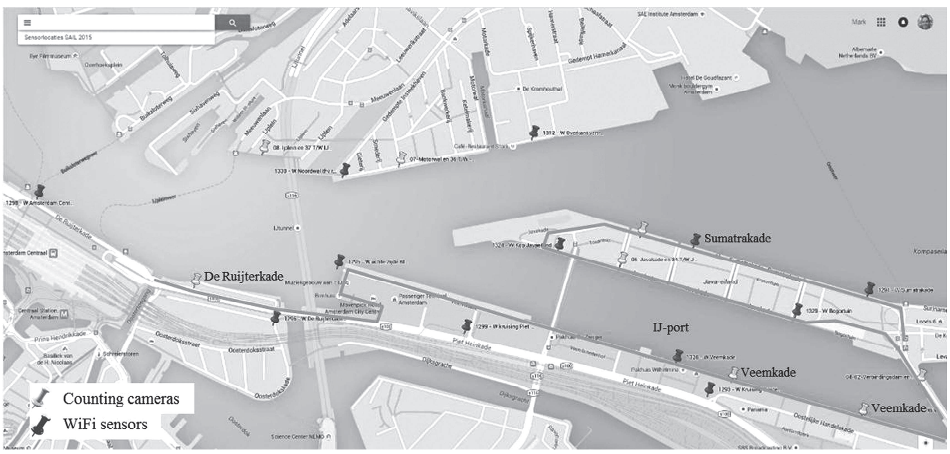


Fig. 1 Overview of the SAIL area. The orange line indicates the walking route along the tall ships. Yellow and blue pushpins indicate counting cameras and Wi-Fi sensors respectively. At each counting camera location also a Wi-Fi sensor has been mounted

3 Overview and comparison of collected raw data

This section coversan overview of the data collection techniques applied during SAIL 2015. Counting cameras have been used to derive flows at cross-sections, Wi-Fi sensors to derive flows as well as travel times and GPS trackers to identify routes and travel times. This section gives some examples of the data to show the applicability of each data source for real-time crowd monitoring.

3.1 Counting camera data

Counting cameras count the number of moving objects (including pedestrians) passing a cross section in two directions based on anonymized videos. The aggregation period of the counts is 60 seconds. The average counting accuracy is 95% [12], though it decreases for increasing density. As an example, Fig. 2 shows the inflows from the left counting camera at the Veemkade. This information can reflect the difference between the days of SAIL as well as the visitor demand pattern during the day.

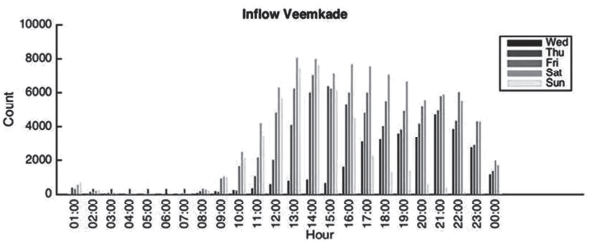
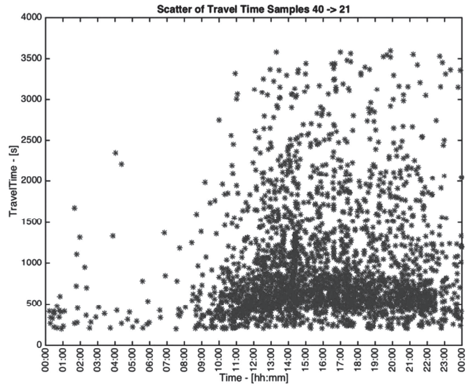


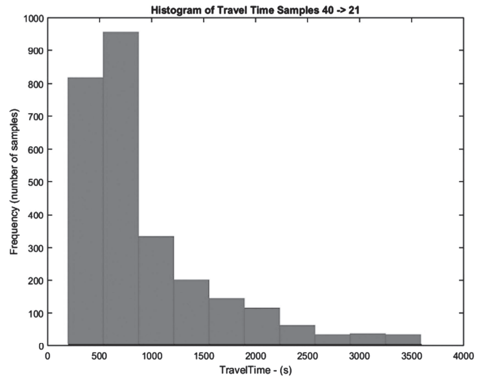
Fig. 2 Inflow at the Veemkade

3.2 Wi-Fi sensor data

Wi-Fi sensorsrecognise the MAC-address of a mobile device with an active Wi-Fi connection (such as smartphones, tablets and smartwatches), and anonymise this before sending it to the server (so-called hashing). We have developed a filter to follow hashed MAC addresses of mobile devices from sensor to sensor to derive routes, route splits and travel times between sensors. Moreover, Wi-Fi counts can be obtained at each sensor. However, static devices and other noise need to be filtered to get accurate results. After filtering, only a small fraction of the hashed MAC-addresses is matched. The number of counted devices therefore needs to be translated into the total number of pedestrians. Yet, a substantial amount of information exists between two sensors due to the large amount of visitors. For instance, there are over hundreds of travel time samples from the left sensor to the right sensor at the Veemkade (see Fig. 1, unidirectional) at an hourly basis during daytime, as shown in Fig. 3.



(a) Travel time scatter



(b) Travel time histogram

Fig. 3 Travel time sample on August 22nd between the two sensors at the Veemkade

3.3 GPS data

GPS trackers are used toderive route splits, that is, the distribution of the total flow at one location over neighbouring locations. These route splits have been based on the detailed routes of individual visitors, as GPS locations are recorded each 20 seconds. While the level of detail in the route information is much higher than from the Wi-Fi sensors, the amount of GPS trackers is much smaller: in total 324 trips have been observed during 5 days, though only 155 trips took place in the studied area.

Unfortunately, GPS-trackers did not synchronize data points real-time to the data server due to occasional poor communications or technical constraints. Moreover, due to the vicinity of water and high building, the GPS measurements were not always accurate. Data filtering and cleaning algorithms have been applied to process and clean the raw data to deliver a smooth trajectory dataset for analysis purposes to compensate as much as possible for missing data (no map-matching). Fig. 4 shows the cleaned GPS tracks on Friday 21 August 2015.

4 Comparison of data collection techniques

To derive the traffic states (and more specifically, to calculate travel time (speed), density and flow) from the collected data, three estimation methods have been developed:

- Travel time estimation method based on the Buckley method[13].
- NWi-Fi method, using Wi-Fi counts to determine density.
- Flow estimation using the BPR method [14].

For a detailed description of each method, we refer to[15]. Table 3 shows the relation between the data sources, the traffic state estimation methods and the traffic state indicators.

Fig.4 GPS tracks on Friday 21 August 2015 (50 trips) [11]

Table 3 Relation between traffic state estimation methods, data sources and traffic state indicators. The roman numbers indicate the order in which the indicators are estimated (I indicates that the state is directly derived from the data). L is the section length, *NWi-Fi* the count from the Wi-Fi sensors, *P* is the Wi-Fi penetration rate, *FD* stands for fundamental diagram

	Raw data	Travel time TT	Speed v	Density k	Flow q
Travel time estimation based on Buckley method (M1)	Wi-Fi travel time samples	Buckley Method-Composite TT model $^{\mathrm{I}}$	$v = \frac{L}{TT}$ $^{\mathrm{II}}$	$k = FD(v)$ $^{\mathrm{III}}$	$q = k \cdot v$ $^{\mathrm{IV}}$
NWi-Fi method to estimate density (M2)	Wi-Fi count matches and camera counts	$TT = \frac{L}{v}$ $^{\mathrm{III}}$	$v = FD(k)$ $^{\mathrm{II}}$	$k = \frac{NWi - Fi}{P \cdot A}$ $^{\mathrm{I}}$	$q = k \cdot v$ $^{\mathrm{III}}$
BPR method and flow estimation (M3)	Wi-Fi count matches and camera counts	$TT = BPR(q)$ $^{\mathrm{II}}$	$v = \frac{L}{TT}$ $^{\mathrm{II}}$	$k = \frac{q}{v}$ $^{\mathrm{IV}}$	$q = \bar{q}$ $^{\mathrm{I}}$

In this section, we use the results of these three methods to compare the quality of the data sources. Unfortunately, we do not always know the so-called ground truth, in order to determine which dataquality is sufficient. However, the footage of a high-resolution camera connected to an air balloon has been used as ground-truth in the comparison.

The comparisonhas been done based on three of the four indicators indicated in Table 3, namely flows, travel times and density. Here, we discard speed, as this can be directly calculated from travel times. In the following subsections each indicator is described in more detail.

4.1 Flows

One of the most important local measurements are flows, or counts. Clearly, the counting cameras readily provide these counts. To see if also the Wi-Fi data can be used as input source for flows, we need to derive the penetration rate of Wi-Fi detections and counts. The two data sources are compared for the most right sensor at the Veemkade, where both a Wi-Fi sensor and a counting camera have been mounted. Fig. 5 shows the comparison between the data from the counting cameras and the Wi-Fi sensors for the last four days (Thursday to

Sunday). Here, we have compared three types of Wi-Fi counts: the total number of counts at the Wi-Fi sensor, the number of unique counts at the Wi-Fi sensor and the number of matches between this Wi-Fi sensor and the adjacent Wi-Fi sensor, in this case the Wi-Fi sensor in the middle of the Veemkade. Note that both the total counts and the number of unique counts might contain detections from mobile devices and local standstill devices.

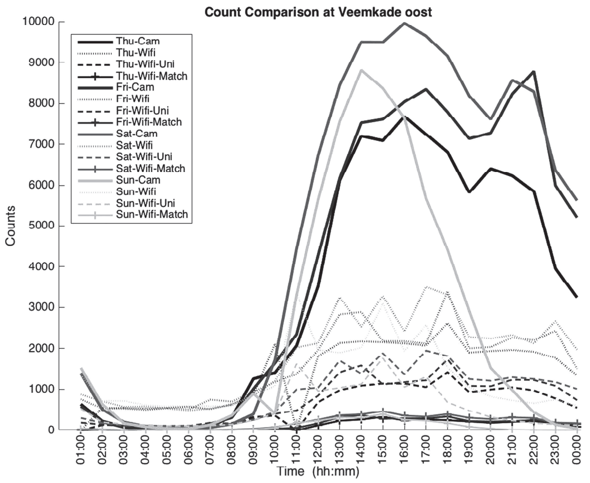


Fig. 5 Comparison of counts from counting cameras and Wi-Fi sensors from Thursday to Sunday

From Fig. 5, it is noticed that about one third (1/3) of the counts from counting cameras can be identified by local Wi-Fi sensors, while about half of these counts are unique. The matching rate between two adjacent Wi-Fi sensors is only 3～4 % of the total flow at the cross-section. As the flows during a large scale event are large, this still leads to a considerable amount of observations. However, for smaller events or regular situations, this percentage might be too small to derive accurate real-time flows.

Secondly, we compare the counts from the Wi-Fi sensors to the GPS tracker counts (Table 4), as we have seen that the GPS trackers provide rich and reliable information on individual route choice behaviour [11], but their penetration rate is rather low.

Table 4 Comparison of daily counts between Wi-Fi sensors and GPS trackers

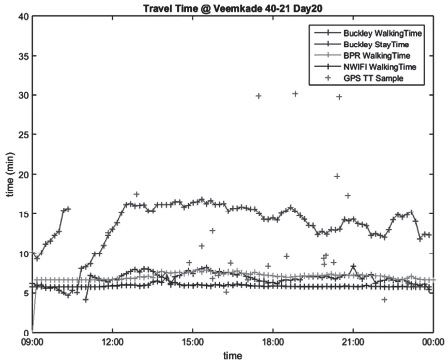
Daily TT counts	Wi-Fi	GPS	Share
August, 19	1644	0	-
August, 20	2848	25	0.88%
August, 21	3393	36	1.06%
August, 22	4351	41	0.94%
August, 23	2338	28	1.20%

Compared to Wi-Fi data, the GPS data derive much less travel time samples regarding the same condition, and the total sample size accounts for approximately 1% of the total Wi-Fi sample size over the five days. Here, we should remark that on the first day (19^{th} Aug 2015), there were problems with keeping the GPS-trackers active, resulting in no useful observations. Obviously, more travel time samples can be obtained during busy periods (afternoon and late evening) compared to non-busy periods (from midnight to morning). During some time slots, travel time samples from GPS data lack completely. This implies that the GPS travel time samples collected during SAIL are too small to provide continuous information for the whole pedestrian population.

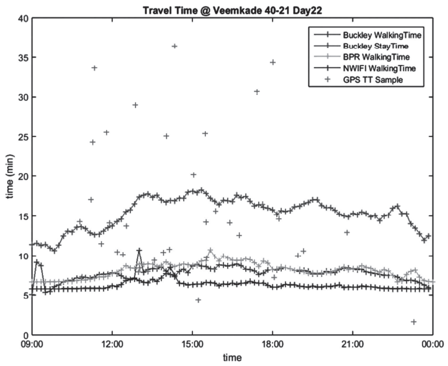
4.2 Walking times

First of all, we need to specify the definition of travel time. Here, travel times consist of two elements: the time needed to walk between two sensors and the time used to perform activities (e.g. watching tall ships, having some food, in the following called area presence time). The travel time estimation method based on the Buckley method provides both walking time and area presence time, where the latter is a combination of walking time and time performing activities. This method requires travel time samples observed from data collection techniques. In each analysis interval (e.g., hourly) there should be a sufficient amount of travel time samples to deliver meaningful results. According to preliminary testing, the travel time dataset should contain at least one free-flow walking time sample and one travel time sample performing activities. The sample size from the GPS dataset does not fulfil this requirement, but the Wi-Fi data set is sufficiently large.

The length of the surveyed section is 400 m. Given an estimated average free-flow speed of 1 m/s, a reasonable free-flow walking time should be about 6.7 min. We can see that, in general, the Buckley method (see the blue curves) provides plausible walking times, within the range between 6 min and 10 min. During afternoon peak hours, the walking time and the area presence time are slightly higher than those in non-busy periods. The GPS travel times are scattered, and the sample is clearly too small to accurately identify the walking time. The counting camera data (used in the BPR method) are used to derive flow, and then turn this flow into travel time using the BPR function. The results for travel time closely match the results from the travel time estimation method. However, when comparing the methods for other areas, it is seen that the BPR method heavily depends on an accurate capacity estimate. The NWi-Fi method derives density from the Wi-Fi counts, and then turns this density into travel time using the fundamental diagram. As the accuracy of the fundamental diagram is still debatable, the resulting travel times are different from those from the other two methods (Fig. 6).



(a) Travel times on August, 20



(b) Travel times on August, 22

Fig. 6 Travel time estimation at the Veemkade using three different traffic state estimation methods

4.3 Density

Fig. 7 shows the densities derived for the De Ruijterkade. For this section, photos from the air balloon camera are available to provide comparison results

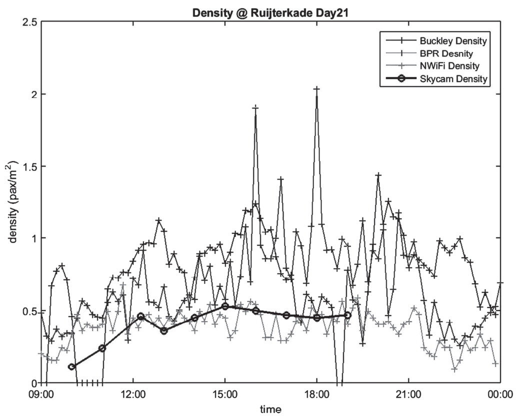


Fig. 7 Density estimation at the De Ruijterkade on August, 21. The air balloon photos (Skycam density, used as a reference) were available from 10:00 until 19:00 on a 15 min basis

Fig. 7 clearly shows that the densities calculated by the NWi-Fi method are closest to the ground truth observations from the air balloon camera. The NWi-Fi method directly uses the Wi-Fi counts as well as the counting camera counts (if these are available). Previously, we found that the counting camera counts were a valuable data source. Here we can see that also the Wi-Fi counts perform well. However, an accurate estimate of the penetration rate of the Wi-Fi devices is essential. The critical point for practical applications is how to calculate these penetration rates, especially at the places where there are no camera counts to couple with Wi-Fi matches.

5 Conclusion

In this paper, we have compared three different data collection techniques (counting cameras, Wi-Fi sensors, GPS trackers) to estimate the pedestrian traffic state real-time during a large scale event in an urban area. The comparison has shown that all data sources provide necessary (and reasonable) input for the estimation methods. The counting cameras appeared to give accurate local counts (at an average accuracy level of 95%), but did not have speed and route information. With the large visitor flows in SAIL, the Wi-Fi dataset was sufficient to get information on routes and travel times (speeds). However, an accurate estimation of the penetration rate (ratio between total flow and share of detected mobile devices) is essential for a correct traffic state. Based on the two joint data sources, estimations for counts, travel times, density and flow can be achieved. GPS data provided detailed but not continuous information on individual route choice behaviour, and the sample size is rather small compared to the Wi-Fi data. Though the sample size was too small for the SAIL event, the GPS data (e.g., from mobile crowd sensing networks with a higher penetration) may serve as a promising data source for the estimation algorithms.

Future research will cover the assessment of other data sources as well as an optimisation of the number and locations of sensors needed for an accurate real-time crowd monitoring system.

Acknowledgements

The research presented in this paper is part of the research program “Allegro: Unravelling slow mode travelling and traffic: with innovative data to a new transportation and traffic theory for pedestrians and bicycles”, a Horizon 2020 project which is funded by the European Research Council. We also like to acknowledge the Amsterdam Institute for Advanced Metropolitan solutions (AMS institute) and the municipality of Amsterdam for their support.

References

[1] W. Daamen, J. van den Heuvel, D. Ton, S. P. Hoogendoorn, "Using Bluetooth and WiFi to unravel real-world slow mode activity travel behaviour," under review.

[2] J. J. Fruin, Pedestrian planning and design (Revised Edition), Elevator World, Inc., 1987.

[3] B. Pushkarev, J. Zupan, Urban space for pedestrians: a report of the Regional Plan Association, Cambridge, Mass., MIT Press, 1975.

[4] P. H. L. Bovy, E. Stern, Route choice, Wayfinding in transport networks, Kluwer Academic Publishers, 1990.

[5] H.J.P. Timmermans, Pedestrian Behavior: Models, Data Collection and Applications, Emerald Group Publishing Limited, 2009.

[6] J. Van den Heuvel, D. Ton, K. Hermansen, "Advances in measuring pedestrians at Dutch train stations using bluetooth, wifi and infrared technology," in Proceedings of Traffic and Granular Flow '15, Delft, 2015.

[7] D. Glas, T. Miyashita, H. Ishiguro, N. Hagita, "Laser-based tracking of human position and orientation using parametric shape modelling," Advanced robotics, vol. 23, no. 4, pp. 405-428, 2009.

[8] M. Mamoli, P. Michieletto, A. Bazzani, B. Giorgini, "Venice as pedestrian city and tourist magnet: mass events and ordinary life", in Proceedings of Conference of the International Forum on Urbanism, 2012.

[9] S. Çolak, L. P. Alexander, B. G. Alvim, S. R. Mehndiretta, M. C. Gonzalez, "Analyzing cell phone location data for urban travel: current methods, limitations and opportunities," in Transportation Research Board 94th Annual Meeting, Washington D.C., 2015.

[10] SAIL Amsterdam, Aanvraag evenementenvergunning SAIL Amsterdam 2015, Technical Report, in Dutch, 2015.

[11] I.M. Galama. Route choice behaviour at mass events. Stated versus revealed preferences of pedestrian route choices at SAIL Amsterdam 2015. MSc thesis, department of Transport & Planning, Delft University of Technology, the Netherlands, 2016.

[12] www.visense.eu. Last accessed 20 June 2016.

[13] D. Buckley, "A semi-Poisson model of traffic flow," Transportation Science, vol. 2 no. 2, pp. 107-132, 1968.

[14] Bureau of Public Roads, Traffic Assignment Manual, U.S. Department of Commerce, 1964.

[15] Y. Yuan, W. Daamen, D. C. Duives, S. P. Hoogendoorn, "Comparison of three algorithms for real-time pedestrian state estimation-supporting a monitoring dashboard for large-scale events," submitted to IEEE-ITSC 2016.