Int. J. ITS Res. (2015) 13:17–27
DOI 10.1007/s13177-013-0076-6

The Use of Smartphone Applications in the Collection of Travel Behaviour Data

Sven Vlassenroot · Dominique Gillis · Rik Bellens · Sidharta Gautama

Received: 16 March 2013 / Revised: 26 December 2013 / Accepted: 26 December 2013 / Published online: 4 February 2014

Abstract The MOVE project deals with the collection and analysis of crowd behaviour data. The main goals of the project are to collect data through the use of mobile phones and to develop new technologies to process and mine the collected data for crowd behaviour analysis. This paper describes the different steps in the development of tracking applications for smartphones that make use of advanced data mining. The results on data collection, analysis, and reporting have led to the development and operation of an advanced urban data monitoring system.

Keywords Travel behaviour · Data collection · Urban mobility · Crowd behaviour data

1 Introduction

Cities and urban environments contain various functions located on the relatively small surface of the city centre. These functions are both numerous and varied, resulting in the attraction of diverse groups of "visitors". In contrast to the traffic situation outside city centres [1] that are characterized by high car usage walking and biking are preferred modes of transport in city centres. These non-motorized mode users have their own way of making modal and route choices. This in return has strong implications on the urban economy, tourism, planning and development. Currently research and modelling efforts have neglected these non-motorized movements in urban environments and have focused almost exclusively on motorised transport modes instead. A possible explanation is that movement behaviour in cities is very complex and that data regarding non-motorized movements is difficult and time-consuming to collect.

Tracking movement in cities offers many opportunities for different domains in urban policy (e.g. spatial planning, tourism, mobility studies, social affairs and economy). Different actors in both the private and public sector have acknowledged that there is a need for monitoring instruments of indicators on how route choices in urban environments are made and this for all types of movements and modes. Modelling and simulating all urban movements is needed for various applications including:

(i) Estimating the required capacity of infrastructure,
(ii) Estimating the feasibility of new facilities (shopping, tourism) that depend on the number of passing pedestrians,
(iii) Assessing the impact of policies on non-motorised and motorised movements and therefore on changing temporal and spatial demand in the environment, and
(iv) Measuring and predicting consumer response on provided information.

Once they have been developed, these models can be key to a wide range of applications in the field of tourism, mobility, eventing and urban economy. Recent technological developments have produced a range of digital tracking technologies

S. Vlassenroot (✉) · D. Gillis · R. Bellens · S. Gautama
Department of Telecommunication and Information Processing, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
e-mail: sven.vlassenroot@ugent.be

S. Vlassenroot
e-mail: svenvlassenroot@vim.be

D. Gillis
e-mail: dominique.gillis@ugent.be

R. Bellens
e-mail: rik.bellens@ugent.be

S. Gautama
e-mail: sidharta.gautama@ugent.be

S. Vlassenroot
Flanders Institute for Mobility, Wetenschapspark 13, 3590 Diepenbeek, Belgium

that offer insight into the movements of users. This has given rise to location-based services (LBS). Tracking technology can be integrated with current mobile phones and PDA's. The simple and standard solution is GPS-based devices. However, the popularity of mobile phones and their increasing capabilities also offer the possibility to continuously track people.

Operating on a phone network requires the network operator to be able to continuously detect the subscriber's proximity to a specific antenna, even when no calls are made. In general, the accuracy of tracked mobile devices is lower than GPS-devices, ranging from 50 to 100 m. Projects such as MIT's Senseable City have investigated behaviour patterns through cell phone activity. The analysed activity is still limited to presence detection within a single cell tower range (typical resolution 100 m) and does not take into account dynamic spatial movement patterns.

In 2009, Ghent University started the MOVE project on the collection and analysis of crowd behaviour data through mobile phone signals. In a first phase, different technologies for the collection of data have been developed and tested. This paper describes the technology and insights on the deployment of mobile applications. We describe the technology in relation to other methods applied in the collection of travel data and will answer the following four questions:

(i) Can the technology gather data that is usable for further analysis?
(ii) Is it possible to capture data of different transport modes (foot, bike, tram, bus, train and car) and what is the quality of this data? Can the mobile application adequately detect different modes during one trip?
(iii) What is the acceptance of different user segments on using the mobile application?
(iv) Is the quality of the collected data sufficient to be used in different study domains?

2 The use of Mobile Phones in the Collection of Travel Data

2.1 Different Methods for the Collection of Individual Trip Behavior

Traditional methods for the collection of data on individual trip behaviour are "self-reporting", which means that people are asked to report their daily trips during a survey period of several days. This can be achieved by means of paper reports ("trip diaries") or telephone surveys. For each trip, the most important variables are registered. These typically include the travel purpose, mode, origin/destination, time of departure/arrival, trip length, etc.

These methods have important drawbacks. An important issue is that part of the information is most likely to be lost due to the delay between the actual trip and the time of completing the trip survey. Typically trip surveys are recorded at the end of the day. This can heavily influence the accuracy of reported departure/arrival times and trip distances. Time durations are typically rounded to 5 or 15-min values and distances are rounded to 5 or 10-km values [2]. Another issue is the difficulty to accurately report locations (origin, destination) and routes followed during trips, due to the limitations of the survey instruments. Finally, these survey methods require a lot of effort from the respondent, causing undesired side-effects such as possible underreporting of short trips [3, 4] and a loss of motivation, typically resulting in a decrease of the reported number of daily trips during the survey period [5, 6]. As a consequence, the traditional survey methods use relatively short survey periods of several days [7, 8].

The introduction of GPS started an important revolution in the collection of trip data. By logging GPS-track data, an enormous amount of detailed information on the position of devices together with the corresponding time stamps became available, offering a massive potential to fill some of the gaps that were not covered by the traditional data gathering methods. The use of GPS enabled the reporting of the exact origins and destinations and of the exact times of departure and arrival. In addition the route itself was logged including detailed information on intermediate points, enabling the monitoring of travel speed. On top of this, data was collected without any interference by the respondent, allowing round-the-clock monitoring, even for longer periods. This type of logging became known as "passive logging".

However, GPS logging had two major shortcomings. Firstly, since GPS devices are typically installed in vehicles, the gathered data only reported on the use of that specific vehicle, covering only a unimodal part of an individual's trip behaviour. Vehicles could also be used by more than one individual, distorting the one-on-one relationship between the individual user and the logged travel data. The use of portable handheld GPS-devices offered a solution to this issue [9], but required the effort and discipline from the respondent to continuously carry the device with him. Forgetting the device by mistake or out of convenience results in unreported gaps in the trip data. Secondly, a drawback of passive logging was the lack of information on trip purpose and travel mode. This introduced new challenges to reconstruct this information in retrospect, either by means of additional surveys [10] or by interpreting the data using logical rules, e.g. using speed information [11] or GIS-information [12]. The interpretation of the data is divided into several partial problems that were approached using different methodologies [11, 13–18]. Splitting the continuous GPS-logging into separate trips by detecting origins and destinations is based on dwelling times at a location. The determination of the travel mode is primarily based on speed characteristics during the trip, which can be further regularized with additional GIS-data e.g. about public transportation networks and rail networks or accelerometer data [e.g. 19]. The trip purpose is estimated using land use maps or by analysing the individual trip chaining. These methodologies yield good results for determining trip ends and travel modes, but the quality of the estimation of trip purposes remains poor [20].

New horizons open up with the popularity of the smartphone, offering similar features as the portable GPS-device but with additional sensors (accelerometer, Bluetooth, Wi-Fi, NFC, UI), which offer a more solid base for travel mode determination [21]. In addition some of the disadvantages mentioned apply to a lesser degree. Carrying a smartphone has become a habit and is therefore less considered as a burden, reducing the risk of non-reported trips. In addition, smartphones offer the opportunity of running dedicated mobile apps, where respondents can report additional trip data on trip purpose, travel mode or on the number of persons undertaking the same trip through the application user interface. Although the mobile apps do require a manual intervention from the respondent ("active" or "interactive logging") [22], the burden remains limited [23] when reporting is restricted to short entries at the very moment of departure and arrival. In addition, time and location of the departure and arrival can be accurately detected.

2.2 The MOVE Mobile Applications

2.2.1 Basic Platform for Crowd Sourcing Mobile Applications

Within MOVE, different mobile applications have been developed for the purpose of monitoring crowd behaviour through cell phone localization and activity. In order to collect data for crowd behaviour analysis, a mobile software application running on Android phones and Java Micro Edition has been developed. This application collects different kinds of valuable information and sends it to a central server. If the smartphone contains a GPS chip, accurate locations of the phone can be collected. Since a GPS chip can severely shorten battery life span and since it does not function well in enclosed spaces such as subway stations or in between large buildings, the GPS signal is not the only source to determine the exact location of the smartphone. Other information sources can be used to estimate the location of the smartphone, such as cell transmission towers in the vicinity or the detected Wi-Fi networks and their signal strengths. In addition, raw accelerometer data if present is collected in order to distinguish between possible modes of transport. Because of the diverse positioning modes, tracking can be done indoor as well as outdoor and the energy consumption of the mobile application is kept under better control than in standard GPS-mode.

A central database accepts the data sent from the smartphones and ensures data homogeneity. Different web services have been built around the database for easy access to the data through spatial queries. Within MOVE, a web-based map service has been built to facilitate the access to the underlying spatial database. This allows easy visualization on map interfaces such as Google Earth, to enable continuous monitoring of the ongoing tests.

2.2.2 Different Variants of the Mobile Application

MOVE offers different mobile apps for both passive and active trip logging. In addition to the technological development looking into architecture and processing, attention has been given to the user interface in order to minimize the burden on the test person and to ensure an attractive "look and feel".

Mobile apps developed for passive trip logging run in the background when the smartphone is switched on. They continuously log data from the GPS, signals from nearby cell towers and Wi-Fi networks and data from the accelerometer. These apps try to minimize impact on battery life span and require no intervention from the user.

Mobile apps developed for active trip logging use a specific interface that is built as a front-end application, allowing manual interventions by the user. This way the user can actively start or stop the data logger in order to restrict the data registration of the actual trips, thereby avoiding unnecessary battery use when the user is not travelling (e.g. walking in the office not considered as travel) or for privacy reasons. Additionally, the user can specify additional trip characteristics such as the travel mode or trip purpose. With active logging, the mobile app function like a travel diary and is used to support mobility studies.

These tools have been developed and tested in a smartphone app 'MOVE', as illustrated in Fig. 1.

Based on the first test results and user findings, a second more advanced mobile application 'CONNECT' has been developed (see Fig. 2), including some technical improvements and a more intuitive interface. An additional feature has been added, namely the possibility to send survey questionnaires to the test person. The activation of a specific questionnaire can be triggered by definable events, like the use of a certain transport mode or trip purpose (e.g. surveys focussed on bicycle trips or on shopping trips) or when passing by a certain location. All registered trips can be consulted in a calendar by the user, allowing him/her to check logs, to correct errors or add missing information.

3 Experimental Results

3.1 Data Collection for EV Living Lab 'Electric Vehicles in Action'

In 2013, the mobile application has been introduced in the context of a living lab on the use of electric vehicles supported

Fig. 1 Illustration of the MOVE mobile application in (*i*) the passive logging mode and (*ii*) the active logging mode, with a diary application

by the Flemish region called 'Electric Vehicles in Action' (EVA). With the purpose of testing the public opinion on electric vehicles, a test group of users has been given the free use of an electric car for 2 months. In order to analyse the role of the electric car within the test group's travel behaviour, the test persons agreed to log their trips during a period of 2 weeks using the smartphone application.

The total period during which data was gathered covered 8 months, from February up to and including August 2013. During this period 23 test persons used the application, logging a total of 1.758 trip segments representing over 10.000 km in distance The test persons have been selected ad random throughout Flanders and represent several behavioural (e.g. travel modes and purposes) and environmental segments (e.g. density, urbanisation, buildings). A smartphone was given to each test person. The Connect app was implemented for active logging, which required the user to manually register the start and end of each trip.

Because of the type of project, the test group is not representative for the general travel behaviour of the population, as it consisted solely of car users with access to a free electric car. For this reason this paper will only discuss the quality of the obtained data and not the registered travel behaviour. There is an underrepresentation of trips using public transportation, bike or foot. To compensate for this bias, the data set has been expanded with the data of one test person who logged his travel behaviour for the same period, including a larger share of non-car trips. These data account for 485 trip segments for a total of nearly 5.000 km.

3.2 Data Extraction and Usability of the Data Gathered from Trips in Urban Areas

Simple statistical analysis of the data already yields interesting results. Part of the information however is hidden and sophisticated techniques are required to analyse the data.

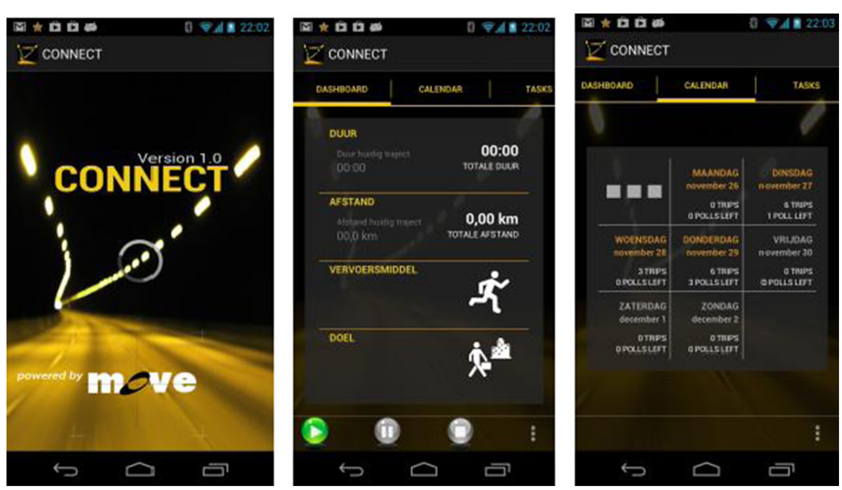


Fig. 2 Screenshot of connect Smartphone application: the opening screen, trip registration screen and weekly overview

3.2.1 Accuracy Enhancement and Geometry Extraction

Individual tracing of mobile devices is relatively inaccurate. The actual accuracy depends on a number of factors: Is the location based on GPS or Wi-Fi/transmission tower data? How many satellites were used to calculate the location? How many Wi-Fi stations or transmission towers were used? Since the positional data derived from Wi-Fi and transmission tower information is detailed only to 50–100 m, the algorithms used to analyse the data had to take this level of inaccuracy into account.

Although the location of individual tracks contains errors, the overall data accuracy could be enhanced by collation data from multiple trips on the same route. Defining which trips had overlapping routes was difficult given the low degree of accuracy. Therefore, a sophisticated and patented technique based on fuzzy voting and image processing [24, 25] was developed.

This technique consists of two stages. In the first stage tracks are plotted in a multi-dimensional grid with a fuzzy buffer. We use fuzzy buffers to take into account the inaccuracy of the locations of the tracks. If a two-dimensional grid would have been used, tracks on adjacent roads would cause interference with each other and it would not be possible to distinguish between those two individual roads. This is why a multi-dimensional grid was used, whereby the extra dimensions allow identification of separate tracks on different routes. For example, the orientation of a trip or a future/past location can be used for this purpose. This process can also be seen as a fuzzy voting in a parameter space and therefore is related to for example the Hough transform [26, 27].

Figure 3 compares the plotting process with a two-dimensional grid, in which case neighbouring roads are not always distinguishable, and a three-dimensional grid, where more details of the underlying road network are visible. The two-dimensional projection in the second case is optimized to visualize the ability to separate roads in three dimensions.

In the second stage of the process, the multi-dimensional grid is processed with morphological image processing techniques [28, 29] adapted to three dimensions and fuzzy values, in order to extract the skeleton of the road network.

Figure 4 shows the result of this process when applied to the two-dimensional grid and the three-dimensional grid of Fig. 3. When using the three dimensional grid the geometry of the roads and the roundabout is very well preserved.

3.2.2 Error Tolerant Graph Matching

The result of the previous process is a skeleton network of the travelled routes. In many cases, one wants to compare this with existing vector data like road network data. This is done by means of error tolerant graph matching [30], taking into account the topology of the graphs. Since not necessarily all roads are present in the probe data and possibly some tracks use routes not in the vector data, it is very important to use graph matching techniques which can handle these kind of errors. Techniques that were originally developed for GIS change detection but which were suitable for this application as well were used. As a result, the position data can be matched with existing vector data, which makes it suitable for further data mining and statistical analysis.

By doing intelligent filtering and recombination of the tracks, one can already make interesting visualizations, which can be used as a starting point for further analysis. For example, Fig. 5a is a visualization of the "local activity". It highlights (in red) those areas where a lot of short trips (<500 m) were made. Figure 5b is a visualization of origin–destination relations. Combinations of origin and destination that are very frequent are shown in red, while less frequent combinations are in green and blue.

Besides the density of tracks in certain areas (Fig. 6a), one can also calculate other interesting features of (local) behaviour. For example, Fig. 6b shows the density of vehicles or users standing still during a certain amount of time. This can be useful to, for instance, detect possible bottlenecks in the road network. In analogy, Fig. 6c shows the average speed on roads.

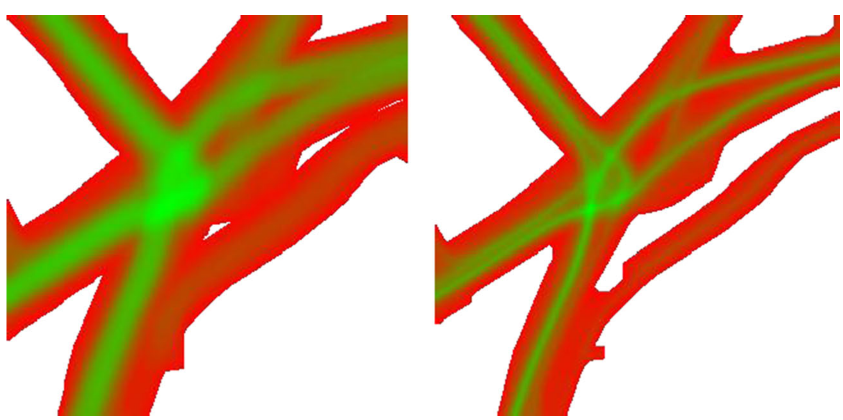


Fig. 3 *Left*: plotting of tracks with fuzzy buffers in a two-dimensional grid. *Right*: two dimensional projection of plotting in a three-dimensional grid with orientation in third dimension. In the image on the right, more details of the road network are visible

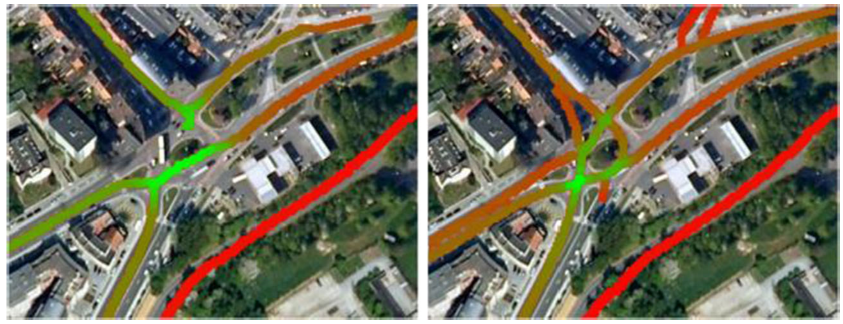


Fig. 4 Examples of deriving geometry from GPS location data. *Left image* based on two-dimensional plotting: the detailed structure of the roundabout is lost. *Right image* based on three-dimensional plotting: structures are well preserved

3.3 Multimodal Travel Behaviour

Using smartphones in the collection of travel behaviour makes it possible to collect data from multimodal trips. A trip becomes multimodal when at least two transport modes are used to get to a certain destination. The quality of the collected data was tested for different transport modes such as walking, cycling, train, trams and buses. The quality of the data depends on the frequency of two types of errors.

On one hand there are user-determined factors, depending on the correct manual registration by the test person. In the current use of the application, the test person manually starts and stops the registration of each trip he makes, indicating the transport mode and purpose for each trip. Such human errors involve, for instance, forgetting to start a trip registration - hence not registering (a part of) the trip - or failure to register a modal change during a trip. Forgetting to stop a trip registration is also a possible error but it is one easy to correct based on the zero speed after arrival at the destination. A last typical problem is forgetting to charge the battery of the smartphone, causing a sudden interruption of the trip registration because of a flat battery. These factors will be discussed in paragraph 3.4.

On the other hand, the quality of the data depends on technical, device-related factors, concerning the accuracy of logged trips. The availability of positioning data during the (complete) trip is crucial. Possible problems in this domain are a quality loss of GPS positioning (e.g. an insufficient number of 'visible' satellites, reflection by buildings, etc.) or loss of data communication. These issues lead to gaps in the registered trip data. The impact of these gaps will be discussed in the following paragraphs.

As a first indicator for the data quality, the data coverage is reported and calculated as the ratio between the number of reported GPS locations and the theoretically expected number of registrations. This theoretically expected number is determined by multiplying the trip duration by the sample frequency (e.g. for a 1-h trip logged at 1 Hz, a total number of 3,600 locations is expected). In practical applications, the reported number of registrations is shown as a percentage of this theoretical number.

The resulting data coverage is summarized in Table 1. Split according to transport mode, the table shows the total number of trips reported, the theoretical number of GPS registrations according to the total trip durations, the reported number of GPS loggings and the resulting data coverage (%).

Over all 2,243 trips, the total registered trip duration is 2,318,469 s or more than 644 h. During these trips, a total of 2,148,542 GPS registrations have been logged, resulting in a global coverage of 92.7 %. Important differences appear between the different transport modes. The data coverage is clearly best for car trips (96.2 % for trips as a car driver, 89.7 % for trips as a car passenger). For walking (80.6 %) and biking (79.9 %) the data coverage is lower, which can be explained by the shorter distances as short gaps have a larger relative impact) and the generally more urban context (concentration in a densely built or urban environment, resulting in a larger risk of shielding or reflection by buildings). A large share of the car trips occurs on highways or roads outside the urban environment, where these distortions are less likely to occur. The data coverage for train tips is the lowest (36.0 %). This is a consequence of the bad GPS localization during train trips, which is a known issue, as reported in e.g. [17].

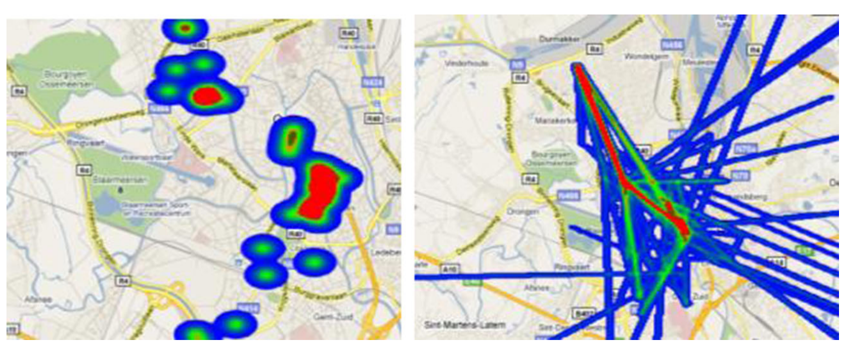
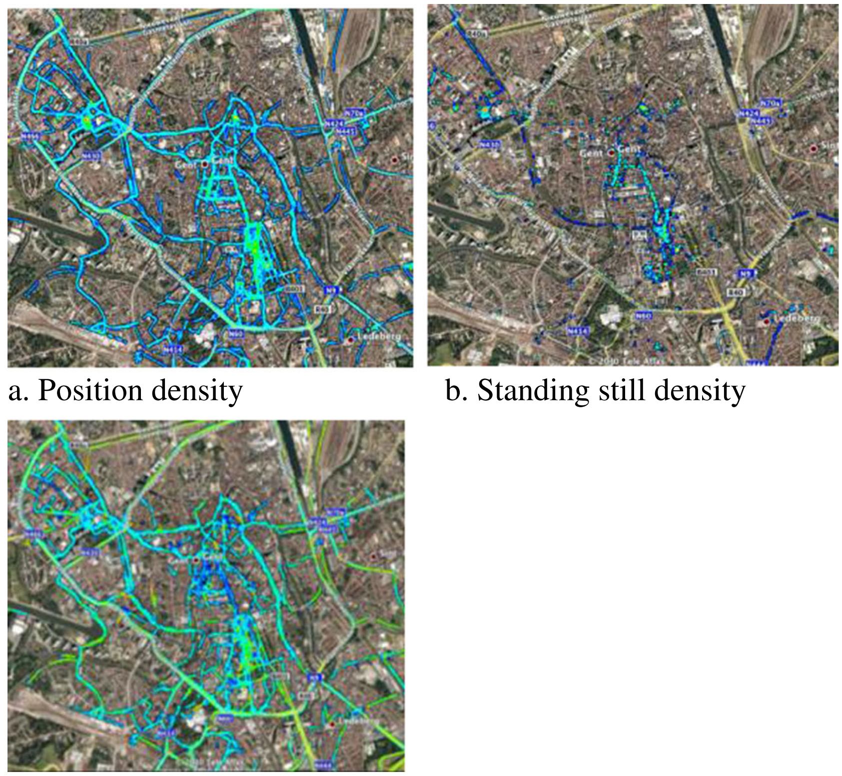


Fig. 5 Visualisations of the data: **a** local activity: areas where a lot of short trips were made are shown in *red*. **b** origin–destination relations: frequent combinations of origin and destination are highlighted in *red*



a. Position density

b. Standing still density

c. Average speed

Fig. 6 Visualizations of different calculated features, based on the probe data. **a** Position density **b** Standing still density **c**. Average speed

This global data coverage is only a partial indicator for the problem, as also the gap duration is a crucial factor. The missed GPS registrations may occur as one long gap, or as a large number of smaller gaps. In the first case, the gap may cause longer unreported segments within a trip, while in the second case a sufficient data quality may be reached by simple interpolation between the last known locations. Therefore a further analysis on the data gaps length has been added. Gaps shorter than 5 s are not analysed as they can be easily solved by simple interpolation. Also, the importance of these short gaps should not be overrated, as the 1 Hz logging puts a very high quality standard. We note that in practical applications a frequency of one GPS location per 5 s is common.

Table 1 Data coverage (%) accor. to the transport mode

Transport mode	Number of trips	Theoreticnumber of loggings	Reported number of GPS loggings	Data cover. (%)
Bike	115	100.991	80.710	79,9 %
Driver	1.700	1.908.665	1.837.022	96,2 %
Foot	338	189.190	152.479	80,6 %
Passenger	43	48.861	43.804	89,7 %
Train	31	55.512	19.957	36,0 %
Other (tram, motor)	16	15.250	14.570	95,5 %
Total	2.243	2.318.469	2.148.542	92,7 %

The remaining data gaps fall into two groups. The first type consists of gaps at the beginning of a trip leg. A trip leg is defined as any unimodal part within a multimodal trip. For example, a trip to work may consist of a trip leg by bike to the train station, a trip leg by train to the next city and a trip leg walking to the office. When starting a registration of a trip leg, the smartphone needs to determine the current position, which may take typically up to half a minute. This causes a delay between the manual start of the trip logging and the effective registration of the first GPS data.

The second type consists of gaps during a trip leg, when temporarily no data is registered. As described above, these gaps may be caused by poor quality of loss of data communication or GPS positioning.

In 60 % of the trip legs, the delay at the start of the registration is shorter than 5 s. The gaps are not included in the further discussion. For the other trips, Fig. 7a shows the distribution of the gap lengths. In 78 % of these trip legs, the duration is shorter than 30 s and in 85 % of the trip legs it is shorter than 60 s. These gaps are sufficiently repairable by interpolation or by shortest path calculation using the real road network. Gaps longer than 1 min mostly occur for trips on foot, as a passenger in a car or (mainly) by train. In the last case, this is due to signal shielding by the train vehicle and its infrastructure.

For the other modes no clear reason can be pointed out. We note that for trips on foot, longer gaps can be restored because

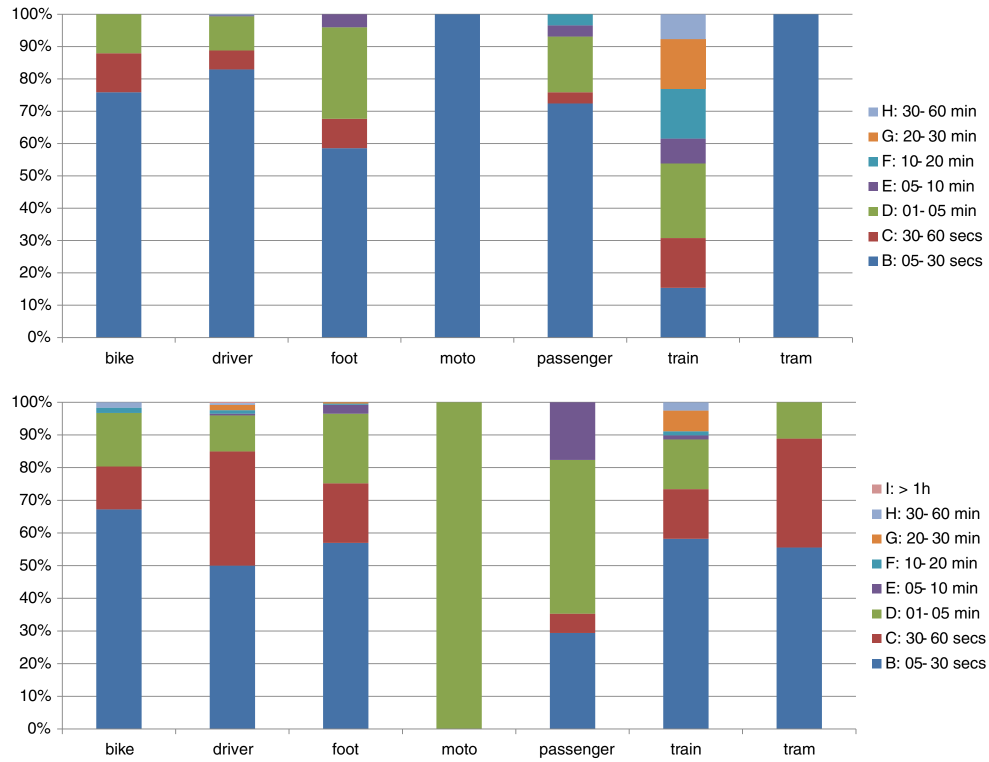


Fig. 7 Distribution of gap lengths, according to the transport model, for gaps at the start of a trip leg (**a**, *above*) and for gaps during a trip leg (**b**, *under*)

of the lower speed: a similar gap duration corresponds with a shorter distance, allowing easier interpolation between the points just before and after the gap. Therefore the gaps are mostly problematic for the train trips.

Similarly, for the gaps during a trip registration the gaps shorter than 5 s are not taken into consideration. For the rest of the gaps, the distribution of the gap durations is reported in Fig. 7b. For 54 % of these gaps the duration is below 30 s, and for 79 % it is below 60 s. Like for the first type of gaps, these gaps can be repaired by interpolation or shortest route calculation. Gaps longer than 60 s mostly occur during trips on foot, as a passenger or by train. The results for trip by motorcycle are not relevant because of the low number of registered trips. There is no clear reason for the poor quality of the results for car passengers, as opposed to the results for car drivers. For the trips on foot, we repeat the remark that larger gaps are acceptable because of the low speed, which means that gaps correspond with shorter distances. Globally less than 5 % of the gaps during a trip take more than 5 min, which mostly happens during train trips. Therefore, also during the trips, gaps with a problematic duration mainly concern train trips. For these trips, a reconstruction of the trip, considering the timetables, is advised, as described in [17].

In conclusion, it can be stated that the application is able to track multimodal transport use. Future studies will also focus on the other smartphone sensors such as Wi-Fi, Bluetooth or accelerometers to cover the gaps or to trace the users where GPS is not available.

3.4 Acceptance and User Friendliness of the Application

When using smartphones as a tool to collect travel data, the aspect of user acceptance and user friendliness plays an important role. Nitsche et al. [21] state that it should neither distract users in their daily phone activities nor cause any limitations of the phone performance. Ideally, data collection should run as a background task, while the effort for the participants should be kept to a minimum with manual entries or questionnaires being unnecessary. An obvious problem of the current generation of smartphones is the battery performance that can dramatically decrease when multiple tasks and sensors are active.

The passive MOVE application has been developed to require minimal intervention by the user and to take into account the energy consumption of the sensors. But again, it was noted that the application was not suited for use in every study domain, since no additional trip characteristics such as travel mode or trip purpose are available unless these are indicated by the test person through manual entries or additional ex-post queries, which would increase the manual effort – contradicting the app's aim to minimize intervention.

The active MOVE application requires intervention by the user whilst trying to avoid complete battery depletion. During a small-scale test twelve colleagues tested the user friendliness of the app, showing that the app is easy to use and to understand. However, some complained about the energy consumption when GPS is activated, as they had to recharge the smartphone battery at least once a day. Users with (older) smartphones with smaller screens reported the readability of the icons as a point of attention. A more in-depth evaluation of the user acceptance of the active MOVE app will be based on the ex-post evaluations by the test persons within the EV Living Lab EVA.

As discussed earlier, the quality of the logged data depends on the correct manual use of the application by the test persons. An attractive and easy-to-use app is a first condition for correct use, in order to avoid frustration and annoyance for the test user and to keep the user motivated to register his trips correctly. To enhance the correct use by the test person, further improvements of the smartphone application are being developed with the aim to reduce the need for manual interventions by the user. This means that the active trip diary will evolve towards a more passive trip logger, requiring automatic detection of trip starts and stops, automatic travel mode recognition (and modal changes during the trip) and estimation of the trip purpose. A methodology is considered where the app proposes its estimations to the user, who can either approve them, either correct them to match the real situation.

3.5 Different Types of Transport and Mobility Behaviour Applications

The collected data offer opportunities for several types of studies. Currently, GPS-data is commonly known as a source of traffic information. Actual speed and travel time data give a direct impression of queues and incidents on the network, which is useful traffic information to be broadcasted towards (potential) road users.

This floating car data is crucial for the calculation of route advice corresponding to the current traffic situation in navigation systems. However, this real-time information is not suitable to give route advice for planned trips, as the current situation may not be relevant anymore, or for long distance trips as bottlenecks may be solved or new incidents may have occurred by the time the driver reaches a specific location. This problem can be tackled using historic speed information, giving statistical insight in the expected (minimal, mean or maximal) speeds on sections of the road network for a given type of road and time of the day or the week (e.g. morning peak, evening peak, off-peak, weekend). With this information route advice can be calculated, aiming at the expected fastest route for the given circumstances, but even taking into account the network robustness i.e. risk of time delays on the proposed route.

As the previous described applications are commonly used in commercial navigation systems, we will not discuss them further. However, collecting similar tracking data by smartphone offers new opportunities because of the multimodal travel data, whereas current navigation systems strongly focus on motorized road traffic. Similarly, smartphone data can help to detect (possible) delays on the public transport network, in order to give users more reliable routing information.

More novel opportunities are found in the field of travel behaviour, both on the individual and on the aggregate level. For individual travel behaviour, smartphone GPS data offer the most direct observation of a person's trips, which is fundamental information for analysing and understanding the rationale behind individual transportation choices. Comparing the observed behaviour to the expected behaviour shows how rationally predictable the behaviour is and which additional factors may influence the individual's choices. This field of research can cover different aspects of trip behaviour, such as transport mode choice [31], route choice [32, 33], departure time choice, etc. Analysis of the observed trip data show how individuals make decisions about their trips and to which extent choices are influenced by personal or household characteristics (age, sex, education level, children, …), or by the supply of transport (car availability, public transport quality, travel time/cost per mode, …) or by external factors (weather, …) [34–36]. A better understanding of individual travel behaviour will also give insight into how this behaviour can be influenced by transport policies or into reasons why transport policy measures do not always results in the expected outcomes.

A specific point of interest in this field is the trip chaining behaviour, combining trips for different purposes into one 'trip chain' [37]. This behaviour differs largely from the classic approach in traffic models, based on home-based trips.

Aggregating these individual trips to the resulting traffic streams will show the global traffic composition in terms of origins and destinations, travel times and distances, mode shares, trip purposes, etc. Further analysis is possible for specific subgroups, either based on geographic selection (properties of traffic in a specific area or using a specific road segment), or trip properties (e.g. deducting typical user profiles per travel mode, or typical travel patterns per trip purpose). This can be valuable information in determining the susceptibility of target groups or profiles for behavioural changes, in order to increase the efficiency of policy priorities or communication strategies, aiming at more sustainable transport behaviour.

An important issue in this field is the representativeness of the monitored population [38]. As smartphones and similar devices are expected to be underrepresented in certain subgroups, depending on age, income and social status, attention [37] should be paid to the selection and/or stratification of a representative test population.

A last research field, which implies a detailed knowledge of individual positioning and dwelling times, is the individual exposure to air pollution, (traffic) noise, traffic unsafety risks, … and the impact of individual travel choices on exposure [39, 40]. This is relevant to estimate the total exposure and the derived health impacts, but also to investigate how a transport policy can steer this travel behaviour towards more healthy and sustainable choices.

4 Conclusion

The combination of tracking mobile phones and movements in cities offers a host of opportunities for different domains in urban policy such as spatial planning, tourism, mobility studies, social affairs or economy. Different actors in the private and public sector have acknowledged the need for a monitoring instrument of motorized and non-motorised movement indicators on how route choices in urban environments are made. Multi-modal use of transport is difficult to monitor. Using trip registration through smartphone mobile applications can help to take the technological barriers down and to gain better acceptance by users (e.g. reducing inaccurate provision of travel data). The data-processing technology can be helpful in giving information to the users but also to provide policy makers with an evaluation of their policies and actions.

In answer to our proposed research questions we can state that (i) the developed tools collect data that is suitable for most types of traffic or mobility studies. We note that gaps occur during the trips registrations, resulting in a temporary loss of data. However, the duration of these gaps is mostly limited. Assuming that gaps shorter than 1 min (longer for slow modes as walking or cycling) can be sufficiently solved by intelligent interpolation between the available registrations, gaps with a problematically long duration are mainly restricted to train trips. Correcting these gaps takes more complex interpolations taking into account the train timetables. For the other modes we can conclude (ii) that the smartphone-based registrations offer good data quality.

Concerning (iii) the user acceptance, the most frequent remark concerns battery life. Users need to charge the smartphone daily, imposing an undesirable burden upon the user. A further large-scale evaluation of the user's satisfaction is planned after completion of the EVA Living Lab.

The collected data certainly offer potential (iv) for travel behaviour surveys. We have described how smartphone registration can solve a large number of the shortcomings of classical travel behaviour surveys, improving the quality and reliability of the data. We did not find comparative studies where both methods were applied simultaneously and results could be compared.

Although the passive tracking application was satisfactory in terms of user friendliness, it is noted that a lot of useful information is missed, like travel mode or purpose. This is why we currently adopt mobile applications with active registration. They will however evolve towards hybrid passive/active logging. Further steps will include the automatic mode recognition and classification and automatic detection or trip starts and stops.

Acknowledgments The authors acknowledge the support of the Agency for Innovation by Science and Technology though the Flemish Living Lab for Electric Vehicles 'Electric Vehicles in Action'.

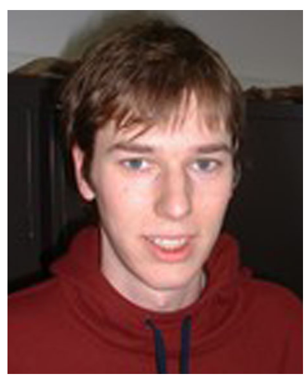
References

1. Brög, W., Erl, E.: "Walking - A Neglected Mode in Transport Surveys", Proceedings of Australia: Walking the 21st Century; An International Walking Conference
2. Wolf, J., Hallmark, S., Oliveira, M., Guensler, R., Sarasua, W.: Accuracy issues with route choice data collection by using global positioning system. Transp. Res. Rec. **1660**, 66–74 (1999)
3. Ettema, D.F., Timmermans, H.J.P., van Vegel, L.: "Effects of data collection methods in travel and activity research", European institute of retailing and services studies. Eindhoven University of Technology, Eindhoven (1996)
4. Wolf, J., Oliveira, M., Thompson, M.: Impact of underreporting on milega and travel time estimates: results from global positioning system-enhanced household travel survey. Transp. Res. Rec. **1854**, 189–198 (1999)
5. Arentze, T., Dijst, M., Dugundji, E., Chang-Hyeon, J., Kapoen, L., Krygsman, S., Maat, K., Timmermans, H.: New activity diary format: design and limited empirical evidence. Transp. Res. Rec. **1768**, 79–88 (2001)
6. Pearson, D.: A comparison of trip determination methods in GPS-enhanced household travel surveys. 84th Annual Meeting of the Transportation Research Board, Washington, DC (2004)
7. Schönfelder, S., Axhausen, K.W., Antille, N., Bierlaire, M.: Exploring the potentials of automatically collected GPS data for travel behaviour analysis – a Swedish data source. In: öltgen, J., Wytzisk, E. (eds.) GI-Technologien für Verkehr und Logis-tik, IfGIprints 13, Institut für Geoinformatik, pp. 55–179. Universität Münster, Münster (2002)
8. Schlich, R., Axhausen, K.W.: Habitual travel behaviour: evidence from a six-week travel diary. Transportation **30**, 13–36 (2003)
9. Steer Davies Gleave and GeoStats, The use of GPS to improve travel data, study report. Prepared for the DTLR New Horizons Programme, Submitted to the London Department for Transport, London (2003)
10. Ohmori, N., Nakazato, M., Harata, N.: GPS mobile phone-based activity diary survey, Proc. East. Asia Soc. Transp. Stud. **5**, 1104–1115 (2005)
11. Tsui, A., Shalaby, A.: Enhanced system for link and mode identification for personal travel surveys based on global positioning systems. Transp. Res. Rec. **1972**, 38–45 (2006)
12. Stopher, P.R., Collins, A.: Conducting a GPS prompted recall survey over the internet. Paper Presented at 84th Annual Meeting of the Transportation Research Board, Washington, DC (2005)
13. Asakura, Y., Hato, E.: Tracking survey for individual travel behaviour using mobile communication instruments. Transp. Res. C **12**, 273–291 (2004)
14. Wolf, J., Schönfelder, S., Samaga, U., Oliveira, M., Axhausen, K.W.: 80 weeks of GPS-traces: approaches to enriching the trip information. Transp. Res. Rec. **1870**, 46–54 (2004)
15. Du, J., Aultman-Hall, L.: Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: automatic trip end identification issues. Transp. Res. A **4**(1), 220–232 (2007)
16. Stopher, P., Clifford, E., Zhang, J., FitzGerald, C.: Deducing Mode and Purpose from GPS data. Working Paper of the Austrian Key Centre" in Transport and Logistics. University of Sydney, Sydney (2008)
17. Bohte, W., Maat, K.: Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys : a large-scale application in the Netherlands. Transp. Res. C **17**(3), 285–297 (2009)

18. Gong, H., Chen, C., Bialostozky, E., Lawson, C.T.: A GPS/GIS method for travel mode detection in New York City. Comput. Environ. Urban. Syst. **36**(2), 131–139 (2011)
19. Hato, E.: Development of behavioral context addressable loggers in the shell for travel-activity analysis. Transp. Res. C **18**, 55–67 (2010)
20. Chen, C., Gong, H., Lawson, C., Bialostozky, E.: Evaluating the feasibility of a passive travel survey collection in a complex urban environment: lessons learned from the New York City case study. Transp. Res. A **44**, 830–840 (2010)
21. Nitsche, P., Widhalm, P., Breuss, S., Maurer, P.: A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys. Procedia Soc. Behav. Sci. **48**, 1033–1046 (2011)
22. Auld, J., Williams, C., Mohammadian, K., Nelson, P.: An automated GPS-based prompted recall survey with learning algorithms. Transp. Lett. Int. J. Transp. Res. **1**, 59–79 (2009)
23. Zhou, J., Golledge, R.: Real-time tracking of activity scheduling/schedule execution within a unified data collection framework. Transp. Res. A **41**, 444–463 (2007)
24. Zadeh, L.A.: Fuzzy sets. Inf. Control. **8**(3), 239–354 (1965)
25. Charlier, N., De Tré, G., Gautama, S., Bellens, R.: Automating actualisation of geographic information using the twofold fuzzy region model. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **18**(3), 247–268 (2010)
26. Duda, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. Comm. ACM **15**, 11–15 (1972)
27. Hough, P.V.C.: “Machine Analysis of Bubble Chamber Pictures”, Proc. Int. Conf. High Energy Accelerators and Instrumentation (1959)
28. Serra, J.: Image Analysis and Mathematical Morphology (1982)
29. Soille, P: Morphological Image Analysis; Principles and Applications, 2nd edition (2003)
30. Gautama, S., Bellens, R., De Tré, G., Philips, W.: Relevance criteria for spatial information retrieval using error-tolerant graph matching. IEEE Trans. Geosci. Remote Sens. **45**(4), 49–54 (2007)
31. Müller, S., Tscharaktschiew, S., Haase, K.: Travel-to-school mode choice modelling and patterns of school choice in urban areas. J. Transp. Geogr. **16**, 342–357 (2008)
32. Papinski, D., Scott, D., Doherty, S.: Exploring the route choice decision-making process: a comparison of planned and observed routes obtained using person-based GPS. Transp. Res. F **12**, 347–358 (2009)
33. Broach, J., Dill, J., Gliebe, J.: Where do cyclists ride? A route choice model developped with revealed preference GPS data. Transp. Res. A **46**, 1730–1740 (2012)
34. Arentze, T., Ettema, D., Timmermans, H.: Estimating a model of dynamic activity generation based on one-day observations : method and results. Transp. Res. B **45**, 447–460 (2011)
35. Kang, H., Scott, D.: Exploring day-to-day variability in time use for household members. Transp. Res. A **44**, 609–619 (2010)
36. Dijst, M., Böcker, L., Kwan, M.: Exposure to weather and implications for travel behaviour: introducing empirical evidence from Europe and Canada. J. Transp. Geogr. **28**, 164–166 (2013)
37. Lee, Y., Hickman, M., Washington, S.: Household type and structure, time-use pattern, and trip chaining behaviour. Transp. Res. A **41**, 1004–1020 (2007)
38. Verkasalo, H., López-Nicolás, C., Molina-Castillo, F.J., Bouwman, H.: Analysis of users and non-users of smartphone applications. Telematics Inform. **27**(3), 242–255 (2010)
39. Boogaard, H., Borgman, F., Kamminga, J., Hoek, G.: Exposure to ultrafine and fine particles and noise during cycling and driving in 11 Dutch cities. Atmos. Environ. **43**(27), 4234–4242 (2009)
40. Dons, E., Int Panis, L., Van Poppel, M., Theunis, J., Willems, H., Torfs, R., Wets, G.: Impact of time–activity patterns on personal exposure to black carbon. Atmos. Environ. **45**(21), 3594–3602 (2011)

Sven Vlassenroot is project manager at the Flanders Institute for Mobility (VIM). He holds a PhD in technology, policy and management (Delft University of Technology). His main interests are sustainable mobility and transport behaviour when using ITS and use of crowd- behaviour data with mobile communication technologies in the domains of transport and electric vehicles.

Dominique Gillis is researcher at the Institute for Sustainable Mobility (IDM) of Ghent University. His research covers the possibilities and applications of tracking data (GPS, smartphone) for travel behavior research, and the interrelation between travel behavior and the individual perception of different types of traffic annoyance (traffic noise, emissions, living quality).



Rik Bellens is researcher at the Department of Telecommunication and Information processing of Ghent University, specializing in the field of processing urban sensor data from several sources (GPS-logging, Wifi- and Bluetooth-scanning, …) into high quality mobility data.

Prof. Sidharta Gautama teaches since 2008 Computer vision at the Department of Telecommunication and Information processing of Ghent University. As Director of i-KNOW, an innovation cluster of Ghent University in the field of intelligent information processing, aiming at the valorisation of research outcomes. Prof. Gautama is the auteur of over 90 publications.