Collection and Analyses of Crowd Travel Behaviour Data by using Smartphones

Rik Bellens[1]
Sven Vlassenroot[2]
Sidharta Guatama[3]

Abstract: *In 2010 the MOVE project started in the collection and analysis of crowd behaviour data. The two main goals of the project are first, the collection of data through the use of mobile phones. The second goal is to develop new technologies to process and mine the collected data for crowd behaviour analysis. The technology will allow to make advanced interpretations of historic and dynamic mobile crowd data coming from GSM/GPS and from different classes of users (vehicle, pedestrian, indoor/outdoor). Fusion will be made between data coming from different sources (smartphone, navigation device) and external map data. The interpretation will allow the mining of advanced features/geometry from the crowd data as well as the dynamic (travel) behavior of the population.*

Keywords: a maximum of 6 keywords separated by ","

1. <u>Introduction</u>

Cities, and urban environments in general, incorporate myriad functions located on the relative small surface area of the city centre. These functions are both numerous and different, which results in a great attraction of different groups of "visitors". In contrast to the traffic situation outside city centres, which are characterized by high car use, walking and biking are the preferred transport modes in city centres (Brög&Erl, 2001).These non-motorized mode users have their own reasons to make mode and route choices, to decide where they will stay longer, etc., and all this will have strong implications on urban economy, tourism, planning and development.

It is therefore somewhat remarkable that research and modelling efforts seems to neglect these non-motorized movements in urban environments, and has instead largely focused exclusively on analysing motorised transport modes. A possible explanation for this bias is that this behaviour in cities is very complex and data collection is difficult and time-consuming.

However, modelling and simulating all the city movements is relevant for a variety of applications, including (i) estimating the required capacity of infrastructure, (ii) estimating the feasibility of new facilities (shopping, tourism) that depend on the volume of passing pedestrians, (iii) assessing the impact of policies on non-motorised and motorised movement and therefore on changing temporal and spatial demand in the environment, and (iv) predicting consumer response to information provision. Once they have been devised, the(se) model(s) will be the key to a wide range of applications in the field of tourism, mobility, eventing and urban economy.

[1]Affiliation of corresponding author
[2]Affiliation of coauthor 1
[3]Affiliation of coauthor 2

Recent technological developments have produced a range of digital tracking technologies that offer a view on the movement of users, which 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. Less known is the fact that mobile phones also offer the possibility to track people constantly. Operating on a phone network requires the network operator to be able to constantly 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 100m. Projects like MIT's Senseable City investigate behaviour patterns through cell phone activity. The analysed activity is still limited to presence detection within a single cell tower range (typical resolution 100m) and does not take into account dynamic spatial movement patterns.

The combination of tracking mobile phones and movements in cities offers a lot of opportunities for different domains in urban policy (e.g. spatial planning, tourism, mobility studies, social affairs, economy). Different actors in the private and public sector have acknowledged (for a host of reasons) the need for a monitoring instrument of (especially non-motorised movements) indicators on how route choices in urban environments are made.

In 2010 the MOVE project started at Ghent University in the collection and analysis of crowd behaviour data. The two main goals of the project are first, the collection of data through the use of mobile phones. The second goal is to develop new technologies to process and mine the collected data for crowd behaviour analysis. This paper will describe the developed technology and some first results of the project.

2. The developed technology

2.1. Smartphone Application

UGent has developed a smartphone application for monitoring crowd behavior through cell phone localization and activity. In order to collect data for crowd behavior analysis, we have developed a mobile software application running on Android phones and Java Micro Edition. This application collects different kinds of valuable information and sends it to a central server. If the cell phone contains a GPS chip, accurate locations of the phone can be collected. However, because GPS is very demanding on the battery and does not work inside buildings, the GPS is only used by the software when most appropriate. At other times, other information is used to derive the location of the phone as accurately as possible, such as the current and neighboring cell towers, the Wi-Fi stations in sight and their signal strengths. On top of that, we also collect measures of the accelerometer of the phones to help us distinguish between pedestrians, cyclists and cars.The user has full control over which data is sent. Because of the diverse positioning modes, tracking can be done indoor as well as outdoor, and the battery life can be preserved better than in standard GPS-mode. The application is shown in Figure 1.

The application currently runs on Android and Java Micro Edition. It has two modes of operation. The first mode is a passive mode where the application runs as a data logger in the background on the smartphone. The logger can track cell towers, wifi, GPS and/or accelerometers and will do so with minimal impact on battery. The user retains control and can at any given time pause tracking. He has a full view on the data being transmitted. The server can communicate with the application to notify the user for software updates, news on the trial or other messages.

Figure 1. Illustration of the MOVE smartphone in (i) passive mode and (ii) diary application

A second mode is a diary mode aimed towards specific mobility studies (see Figure 1). For these trials, sampling is done through a dedicated test audience carrying a digital diary. This is the same application but with a specific interface, which additionally records specific activity for a given mobility study (e.g. mode of transport, aim of transport, number of passengers etc.). Data is directly transmitted to a central server, allowing direct monitoring of the test audience. Messages can be sent to specific test persons. The diary interface can be changed on-the-fly to refine or even change the study.

2.2. MOVE server

One central database will store data from the smartphones ensuring data homogeneity. Different web services will be built around the database for easy access to the data through spatial queries as described in Figure 2. Within the MOVE project a web feature service based on GeoServer has been built to access the underlying PostGIS database. This allows easy visualization on map interfaces like Google Earth. Continuous monitoring of the ongoing trials is possible.

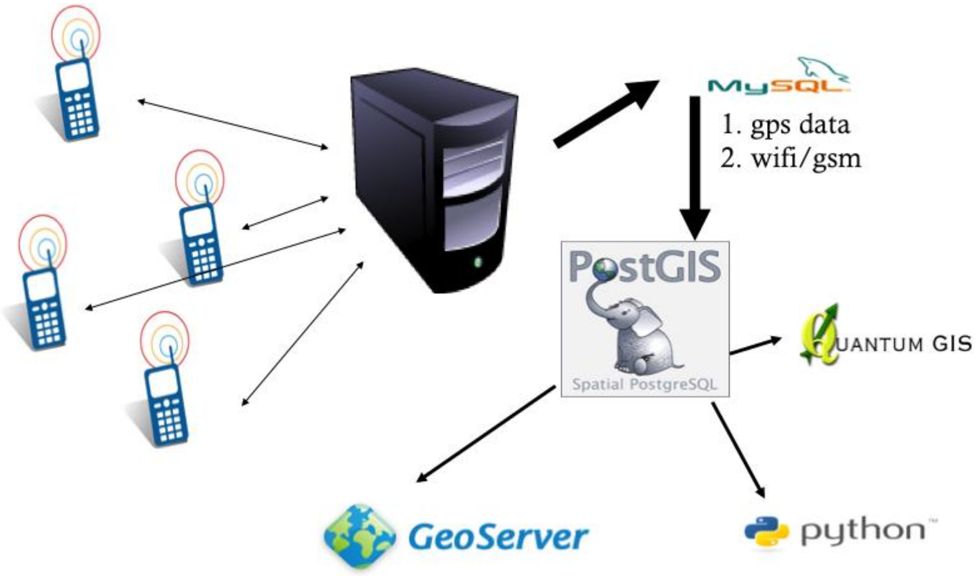


Figure 2. Illustration of the server and server applications

3. Data Processing Technology

The location of individual phones in itself is not very useful. It only becomes interesting when the locations of many users together can be analyzed. However, going from these individual locations to crowd behaviour information, is still a large gap to bridge. Simple statistical analysis of the data can already be interesting, however, the truly interesting information is hidden and sophisticated technologies are needed to analyze the data. Since, the positional data derived from wifi and cell information is very inaccurate, the algorithms have to take into account this inaccuracy.

The proposed workflow to process the data consists of three stages:

- In a first stage, the accuracy is enhanced and the geometry is extracted through a multidimensional rastering and fuzzy voting mechanism. This allows us to extract even the fine details of movement flows (see Figures 2 and 3).
- The second stage uses error tolerant graph matching to match the data with existing vector data such as road maps.
- Thirdly, different local features are calculated on the data and data mining techniques, such as neural networks, are used to derive high quality crowd behaviour data.
-

3.1 Accuracy enhancement and geometry extraction

Individual tracks of mobile devices are relatively inaccurate. The actual accuracy depends on a number of factors: Is the location based on gps or wifi/cell data? How many satellites were used to calculate the location. How many wifi stations or cell towers were used? Which algorithms were used? As a result, the accuracy ranges from a few meters to a few hundreds of meters. This limited and variable accuracy make it difficult to analyze and match tracks. Therefore, advanced algorithms are necessary.

Although the location of individual tracks contains errors, the accuracy can be enhanced by using the mean of multiple tracks of the same route. However, because of the low accuracy it is very difficult to find which tracks have overlapping routes. Therefore, we developed a sophisticated and patented technique based on fuzzy voting and image processing. This technique consists of two stages. In the first stage tracks are rasterized in a multi-dimensional grid with a fuzzy buffer. We use fuzzy buffers (Zadeh 1965, Charlier et al. 2010) to take into account the inaccuracy of the locations of the tracks. When a two-dimensional raster would be used, tracks on roads which lie close to each other, would interfere a lot and it would not be possible to distinguish between those roads. Therefore, we use a more-dimensional raster, where the extra dimensions are used to separate tracks on different routes. For example, the orientation 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 (Duda and Hart 1972, Hough 1959).

Figure 3 compares this rasterization process when using a two-dimensional raster, in which case neighbouring roads are not always distinguishable, and a three-dimensional raster, where more details of the underlying road network are visible. The two-dimensional projection in the latter case is optimized to visualise the ability to separate roads in three dimensions.

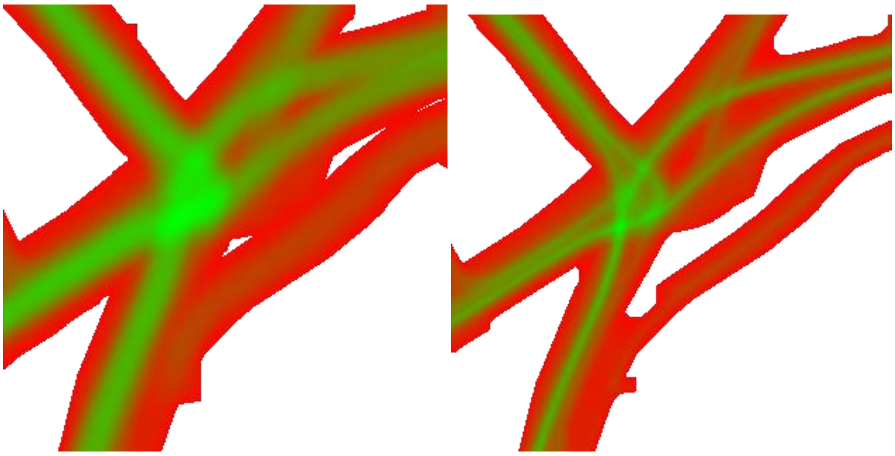


Figure 3. Left: rasterization of tracks with fuzzy buffers in a two dimensional raster. Right: two dimensional projection of a rasterization in a three dimensional raster with orientation in third dimension. In the image on the right, more details of the road network are visible.

In the second stage of the process, the multi-dimensional raster is processed with morphological image processing techniques (Serra 1982, Soille 2003) 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 raster and the three dimensional raster of figure 3. When using the three dimensional raster the geometry of the roads and the roundabout is very well preserved. Figure 5 shows another example of the fine structures that can be extracted with the proposed method.

Figure 4. Examples of deriving geometry from gps location data. Left image based on two dimensional rastering: the detailed structure of the roundabout is lost. Right image based on three dimensional rastering: structures are well preserved.

3.2 Error tolerant graph matching

The results of the previous process is a skeleton of the travelled routes. In many case, one wants to compare this with existing vector data like road networks. For this, we use error tolerant graph matching 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. We use techniques developed for GIS change detection (Gautama et al. 2007), but also applicable for this application.

4. Test group

Currently, a group of students of the University of Ghent have installed this software and are constantly collecting data in and around Ghent, Belgium. The received free mobile Internet for six months to participate in the project. Main goal is to test our software and the developed technology.

5. Data mining and first analysis

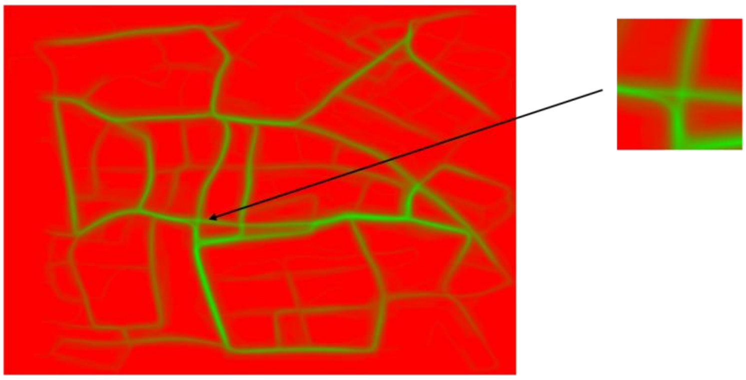
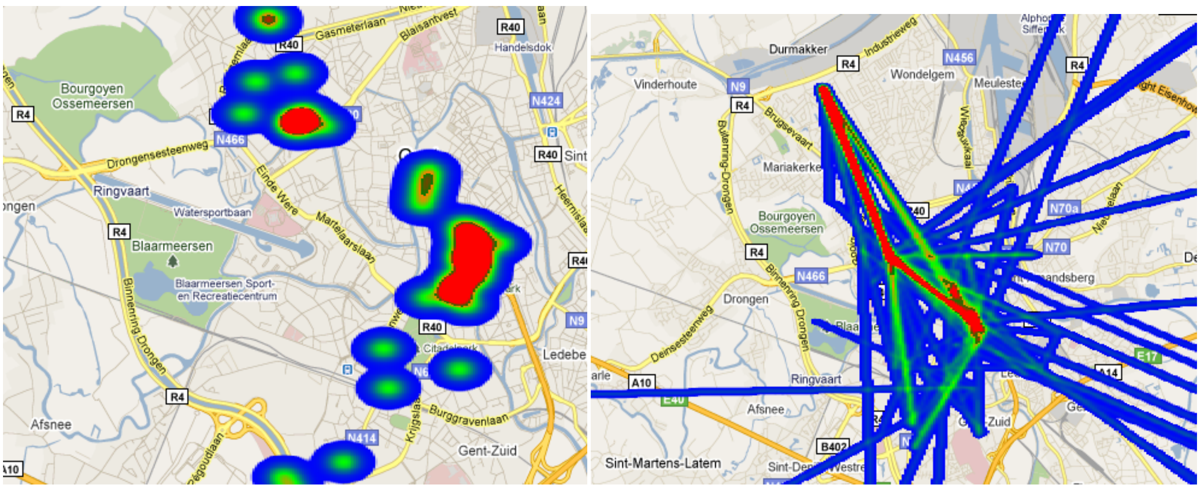


Figure 5. Example of the enhancement of accuracy through multidimensional rastering with fuzzy voting.

With the previous techniques geometry can be extracted from the data and it can be matched with existing vector data. However, in order to derive really interesting crowd behavior information from the data, we have to take one step further. The data can be further analyzed and processed with many data mining and statistical analysis tools. What follows is a small number of example results we achieved with the data.

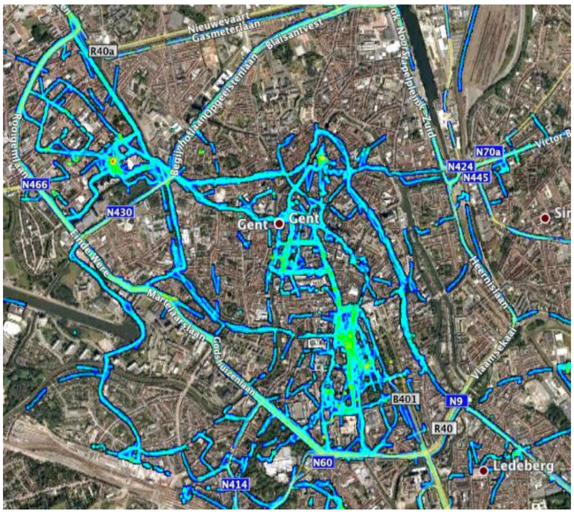
By doing intelligent filtering and recombination of the tracks, one can already make interesting visualizations, which can be used as a starting point for further analysis. For example, figure 6 (a) is a visualization of the "local activity". It highlights (in red) those areas where a lot of short trips (<500 m) were made. Figure 6 (b) is a visualization of origin-destination relations. Combinations of origin and destination which are very frequent are shown in red, while less frequent combinations are in green and blue.



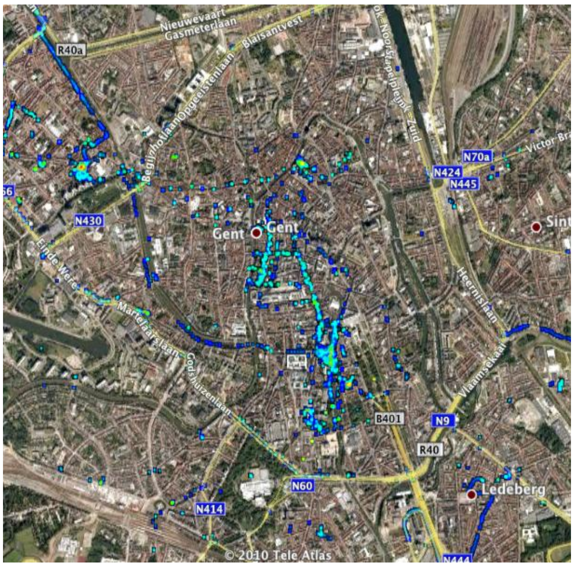
(a) Local activity **(b) origin – destination relations**

Figure 6. 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.

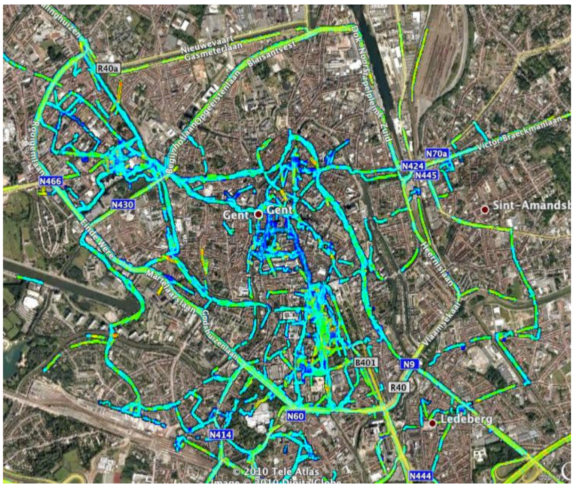
Besides the density of tracks in certain areas, one can also calculate other interesting features of (local) behaviour. For example, figure 7(b) shows the density of vehicles standing still during a certain amount of time. This can be interesting to discover possible bottlenecks in the road network. Analogously, figure 7(c) shows the average speed on roads. Many other features, such as acceleration, stop-and-go behaviour and so on could be calculated in this manner.



(a) Position density



(b) Standing still density



(c) Average speed

Figure 7. Visualisations of different features calculated based on the probe data.

Those features are already interesting on their own, but it becomes really interesting when they are combined. Based on the local behaviour of individual vehicles, it is even possible to derive certain properties of the road network. This is illustrated in figure 8. Here, three different features of the local speed profile are visualised in the different colour channels: in the red channel, stop-and-go behaviour, in the green channel, constant speed and in the blue channel slowing down/acceleration. As can be expected, cross-roads with traffic lights are coloured red, while other cross-roads and roundabouts are coloured blue. Straight roads are coloured green.

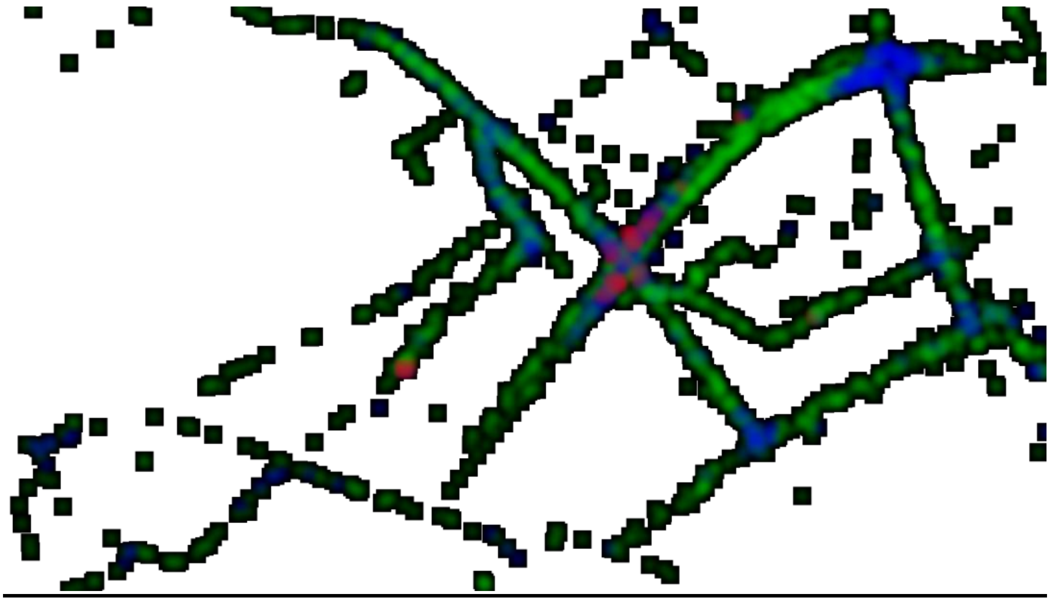


Figure 8. Example of traffic light detection based on features of local behaviour (of speed profile). In the red colour channel, stop and go behaviour is shown, in the blue channel, slowing down/accelerate behaviour and in the green channel constant speed behaviour.

6. Conclusion

Our first results indicated that the application is very interesting and useful in the domain of collecting geo-spatial data.

Bibliography

Brög, W., and Erl, E. (2001) *Walking - A Neglected Mode in Transport Surveys, Proceedings of Australia: Walking the 21st Century; An International Walking Conference.*

Duda, R. O. and P. E. Hart, *Use of the Hough Transformation to Detect Lines and Curves in Pictures*, **Comm. ACM**, Vol. 15, pp. 11-15 (January, 1972),

P.V.C. Hough, *Machine Analysis of Bubble Chamber Pictures,* **Proc. Int. Conf. High Energy Accelerators and Instrumentation**, 1959

N. Charlier, G. De Tré, S. Gautama,and R. Bellens, *Automating actualisation of geographic information using the twofold fuzzy region model,* **International Journal Of Uncertainty Fuzziness And Knowledge-Based Systems**, Vol. 18(3), 2010

J. Serra, *Image Analysis and Mathematical Morphology,* 1982

P. Soille, *Morphological Image Analysis; Principles and Applications, 2nd edition,* 2003
L. A. Zadeh, *Fuzzy sets,* **Information and Control,** Vol. 8(3), 1965
S. Gautama, R. Bellens, G. De Tré, W. Philips, *Relevance criteria for spatial information retrieval using error-tolerant graph matching*, **IEEE Transactions On Geoscience And Remote Sensing,** Vol. 45(4), 2007