Correcting GPS Readings from a Tracked Mobile Sensor

Richard Milton and Anthony Steed

Department of Computer Science, University College London,
Gower St., London, WC1E 6BT, United Kingdom
{R.Milton, A.Steed}@cs.ucl.ac.uk

Abstract. We present a series of techniques that we have been using to process GPS readings to increase their accuracy. In a study of urban pollution, we have deployed a number of tracked mobile pollution monitors comprising a PDA, GPS sensor and carbon monoxide (CO) sensor. These pollution monitors are carried by pedestrians and cyclists. Because we are operating in an urban environment where the sky is often occluded, the resulting GPS logs will show periods of low availability of fix and a wide variety of error conditions. From the raw GPS and CO logs we are able to make maps of pollution at a 50m scale. However, because we know the behaviour of the carriers of the devices, and we can relate the GPS behaviour and known effects of CO in the environment, we can correct the GPS logs semi-automatically. This allows us to achieve a roughly 5m scale in our maps, which enables us to observe a new class of expected environmental effects. In this paper we present the techniques we have developed and give a general overview of how other knowledge might be integrated by system integrators to correct their own log files.

1   Introduction

In this paper we describe a study of carbon-monoxide (CO) pollution that uses a small set of mobile CO sensors. Our aim is to make a map of CO pollution in a small area of a city at a scale fine enough to identify per street variation. The sensors are tracked using a GPS sensor. However, because the availability and quality of a GPS signal will vary greatly in an urban area, we have to be careful how we process the GPS log files.

The quality of GPS varies in the urban environment because of the urban canyon effect. A GPS sensor needs to be able to see three or more GPS satellites in order to get a reliable reading. However, buildings around the sensor cause the view of the sky to be obscured, and as the sensor moves the view of the sky will change quite rapidly.

Without processing the GPS logs, we are able to generate maps of CO at a scale of roughly 50m. Although this allows us to see gross variations in CO, such as CO being lower in a relatively traffic-free square, it is not sufficient to see per street variations. Errors in GPS reading will often put the sensor inside buildings, next to the pavements and streets, or will even put the sensor in

adjacent streets. Obviously this prevents us identifying pollution variations that exist between streets.

However, in our trials, we know quite a lot about the behaviour of the carriers of the pollution monitors. We can use knowledge about the behaviour to process the GPS logs to increase the accuracy. This has enabled us to identify features at the 5m scale including CO hot spots such as bus stops and road crossings. We discuss five different levels of knowledge of the situation that we can apply to exclude or filter GPS readings: knowledge of geographic extent of trial, knowledge of the building footprints, knowledge of the route undertaken, knowledge of the behaviour of the carrier of the GPS receiver and knowledge about expected sensor readings. We then describe tools that we have built to help us do such corrections semi-automatically.

In the following section we will present some related work on environmental sensing and GPS correction. Then in Section 3 we will outline the sensor package and the trials we have undertaken. In Section 4 we give an overview of the type and quality of data that we can capture. Then in Section 5 we discuss how we can refine the GPS readings to give us more focussed data. In Section 6 we will then generalise and discuss tools that should be more applicable to processing tracked data logs. Finally we conclude and present plans for future work.

2   Related Work

The background to the work in this paper is an ongoing study of urban carbon monoxide. The Air Quality Site contains archive data from over 1500 UK monitoring stations going back in some instances to 1972 [1]. Such data sources give a good picture of variation from urban to rural areas. In urban areas some sense of potential variation is conveyed by the difference in readings between kerbside sensors and sensors placed in background areas away from pollutant sources. However they don't capture the detail of per street variation.

CO disperses over a matter of hours, but Croxford et al. have shown that this is affected by local street configuration [2]. Croxford's study used a cluster of sensors in fixed placements in a small area around University College London, UK. The Air Quality Strategy for England, Scotland, Wales and Northern Ireland [3], suggests a standard of 10ppm (11.6mg/m^3) running 8-hour mean. In the vicinity of UCL, the Croxford study found a peak CO concentration of 12ppm, but nearby sensors reported much lower values near the background level for CO. Thus, moving pedestrians or vehicles would probably not experience this peak for a long period.

Our approach to bridging this gap between large, sparse and small, dense studies is to use tracked mobile sensors that are carried by pedestrians or mounted on vehicles. There are several related works in the UK that aim to increase understanding of pollution variation. These include the dense sensor network of the Discovery Net e-Science project [4], and the combined sensing and air flow modelling approach of DAPPLE [5].

In this paper we apply knowledge from records of trials and map data to correct GPS readings. GPS is of course a very significant area of research and there are many existing and upcoming enhancements to GPS that can improve its accuracy [6]. Most relevant for this work is fusion of velocity and acceleration data with the GPS position [7]. We wanted to focus on consumer-level GPS because we anticipate sensor package cost being a primary concern when deploying large collections. We also didn't want to weigh down the carriers, nor burden them with any device configuration tasks. Indeed, the user should simply have to power on the devices before going outside. In an urban environment, especially in central London where the buildings are typically close to or higher than the road width, we do not expect to get a GPS fix 100% of the time [8].

3   Trial Description

3.1   Monitor Overview

We will briefly describe the trial; a longer description with more background on the task and scientific justification from an environmental science point of view can be found in [9]. We collected data over two weeks, in collaboration with the 'Dispersion of Air Pollution and Penetration into the Local Environment' (DAPPLE) project. The data collection area was centred around Marylebone Road and Gloucester Place in central London. The local area is densely packed with buildings up to roughly eight stories high, with a wide variety of street widths. We collected data from both pedestrians and walkers.

Fig. 1. The ICOM carbon monoxide sensor on the left, the PDA in the middle and the GPS aerial on the right. The HAICOM GPS unit can just be seen sticking out of the top of the PDA

The equipment pack consisted of an HP Jornada PDA running our own data logging software, an ICOM device from Learian Environmental to measure carbon monoxide and a GPS unit. Two types of GPS units were used: the HAICOM

type, or a self-built device based around the GARMIN 15 package. All the equipment was colour coded and divided into separate coloured packs, which were not mixed, to allow for easy tracking of the equipment used for each sample run.

Data is sampled once every second from the ICOM carbon monoxide sensor, combined with the last valid GPS fix and written to a log file on the PDA. Every time the software on the PDA is run, both a GPS log file and a pollution log file are created with a filename that includes the current timestamp. After data collection, these files are uploaded to a database server for analysis.

3.2   System Overview

All the data is stored in a Postgres database, with the reported and corrected GPS positions in separate fields. Correction of the GPS data can take one of two forms: either the data is fitted to the point closest to a route drawn on a map, or an offset is calculated from the position of a known point and applied to a section of the data. With data collection taking place on a limited number of fixed routes, idealised routes have been created to ensure that the corrections are coherent.

3.3   Routes

In these trials, we studied a small number of walkers and cyclists in a small area to prove the concept. To cover larger areas would require more carriers, and eventually we hope to distribute 100s or 1000s of devices to commuters.

Two routes along the length of Marylebone Road were used for most of the walking, in addition to both walking and cycling around the cycle route. On other occasions, different routes around the area were tried (Figure 2).

4   Quality of Captured Data

During the data collection period, up to 3 carriers, either walkers or cyclists were out for up to four 45 minute sessions a day. 87 log files containing GPS positions and carbon monoxide data were collected, resulting in 227,496 separate data points. During the two-week sampling period, a small number of equipment problems were encountered, resulting in the loss of 10 sample runs. Figures 3 and 4 show an example log file.

### 4.1   Reliability of GPS in an Urban Environment

Overall, the percentage of data points with corresponding GPS positions was 42.8% of the total walking and cycling time, corresponding to 34.6 hours of collection time over two weeks. By comparing daily GPS data for the pedestrians and cyclists, it can be seen that the cyclists have a marginally higher percentage of valid GPS positions. These levels of GPS reliability might sound low, but are not atypical for an urban environment. In [10], a car-mounted consumer GPS unit driven around central London achieved 60% reliability. Given that in this case the aerial placement and vehicle behaviour would probably have afforded a more stable view of the sky, this suggests our availability results are not low.
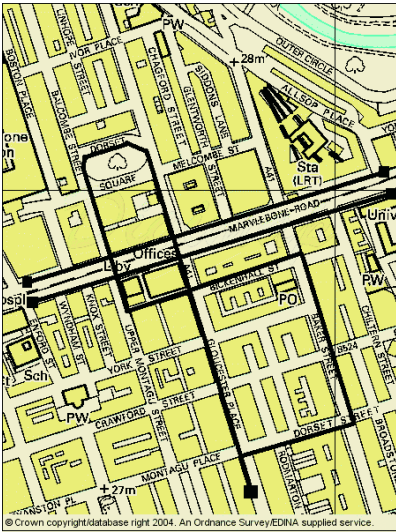
Fig. 2. The two walking routes along Marylebone Road and the figure 8 cycle route


Fig. 3. GPS trace for the cycle route on 19 May 2004 at 11:48
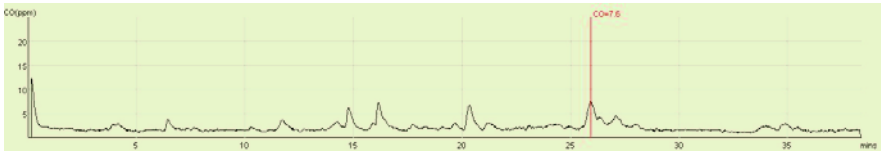
Fig. 4. Carbon monoxide trace for the cycle route on 19 May 2004 at 11:48. The maximum CO reading, 7.6ppm is indicated

When looking at the GPS traces, it is apparent that the GPS is not reliable enough to place a sensor on the correct side of the road. Occasionally, a log file will contain a distinct trace that sticks to the correct side of the road, but the majority meander along the centre of the road, or drift from side to side. For the Marylebone Road walking routes, without looking at the recorded route notes, it is not usually possible to be certain whether the sensor was on the north or south side of the road, unless additional features are present. For the cycle route this is less of a problem as the cyclists are always in the road and the pedestrians always take the same route, although the problem of detecting junctions accurately remains.

The method of standing in a fixed location and averaging the GPS position for a period of time is a standard GPS surveying technique used to increase accuracy, but this is not applicable to mobile CO sensors. Using differential GPS would be more appropriate, but again, this method requires expensive equipment and does
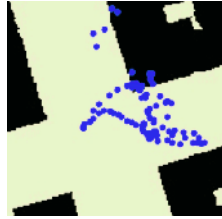
Fig. 5. Reported GPS location while standing at the bottom right corner of the junction for 30 minutes

not lend itself to ad-hoc data collection by people during their normal daily lives. Using inexpensive commercial GPS devices, post processing of the data is not possible due to the pseudo range information not being stored, so the accuracy is limited to about 15m. As was shown previously, the carbon monoxide level can vary significantly from the kerb side to the building, so a more accurate method of positioning samples using standard GPS devices is required.

By analyzing the collected GPS data, there is additional information present in the trace that can be exploited to perform further correction. There is little that can be done about multi-path reflections other than detecting when low elevation satellites are used in the solution or using the GPS unit's own dilution of precision figure to assign a confidence measure to the position.

4.2   ICOM Accuracy

The quoted tolerance of the ICOM CO sensors is +/- 5%, but the nature of the carbon monoxide levels being measured is such that they can change very rapidly. The range of values of most interest is between 0.3ppm and 5.0ppm. Although much larger values do occur, they are very rare and their location is of more interest than their value. A base level of 0.3ppm corresponds to the global background carbon monoxide level.

5   Correction Methodologies

In this section we describe features in the data that can be used to correct GPS. These come from two main sources: knowledge about the behaviour of the carriers and knowledge about expected CO effects. Using such knowledge allows us to correct our GPS logs so that we can discriminate between the readings more accurately, and thus make maps at a finer scale.

In the following figures, raw trace logs (Figures 6, 7, 10 & 11, and also Fig 3) are drawn with buildings in black, background in white and a red-scale for pollution. Summary maps (Figures 8 & 9) are drawn with a similar scale but on a grey background. The method of plotting summary maps was to divide the area into grid squares of either 20 metre or 5 metre spacing and to place data points into the correct square based on GPS position. The mean and variance
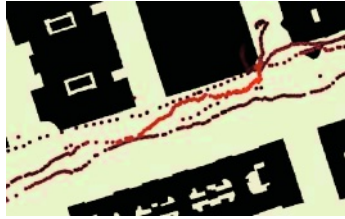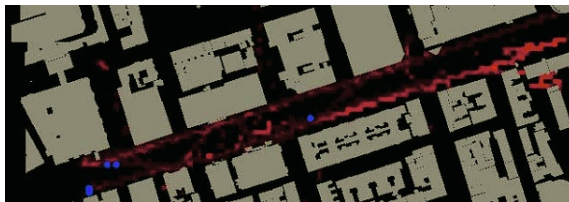
Fig. 6. An uncorrected GPS trace on Marylebone Road

Fig. 7. The trace of Figure 6, fitted to a line drawn along the kerbside using the GPS correction tool

Fig. 8. 5 May 2004 09:00-17:00, 5 metre mean boxes without GPS correction

Fig. 9. 5 May 2004 09:00-17:00, 5 metre mean boxes with corrected GPS positions

for each square can then be calculated and plotted. By altering the size of the grid squares, the data can be plotted to within the accuracy limits imposed by the GPS positional error. We use a simple grid-based visualisation because the mean and variance vary at a fine geometrical scale and don't lend themselves well to smooth representations as, say, surfaces.

Fig. 10. Identifying a south side route from a pedestrian crossing

Fig. 11. GPS trace crossing Marylebone Road

5.1 GPS Logfile Manipulation

The Postgres database that stores all of the information in the collected log files is structured so that data for an individual path can be extracted. Each path can be analyzed separately using a tool that plots a colour coded carbon monoxide value over an aerial map using the GPS position logged at the time the sample was taken. Traces can be played back in real-time, or faster, allowing the velocity of the sensor to be tracked. The GPS positions for a path can be corrected by drawing a polyline representing the idealised route on the map and fitting the GPS points to the nearest point on the line. When correcting the data, only points which lie a user-defined distance from the line are corrected, allowing a limit to be set on how far points will be moved. Typically, 20m has been used as this is the GPS accuracy. During the correction process, points that have no valid GPS position, but lie between two valid GPS positions, are linearly interpolated. See Figure 6 and 7 for an example of before and after correction.

Using the viewer tool, GPS tracks that cross into buildings can often be seen, Figure 3 is a good example of this. This type of error is harder to correct, as using the line correction tool will result in a trace with a discontinuity at the corner of the building. The trace needs to be stretched around the building, but this has not been attempted due to the majority of the data fitting the line correction model more closely.

5.2 Fitting GPS Traces to Routes

As an example, data for 5 May 2004 was corrected using the GPS correction tool to see what physical features could be observed. This day was chosen as a good example for wind variation on opposite sides of the road as the wind only varied between 170 degrees and 260 degrees all day, starting at 5 knots and increasing to 16 knots. The difference between the corrected and uncorrected data is clearly visible.

The data in Figure 9 shows a marked difference in carbon monoxide levels between the north and south sides of the road at the east end. Compare to the raw data in Figure 8. On the north side the levels are 1.5 ppm, while on the south side they are 3.5ppm. A factor of 2-3 difference is something that would be expected with the wind coming from the southwest quadrant, dispersing the carbon monoxide on the north side of the road where the wind speed is highest and causing levels to build up on the south side where the air is static.

5.3 Disambiguating Route Choice

In addition to correcting known routes, it is also necessary to be able to work out the route taken from the GPS data. On the Marylebone Road walking routes, which are virtually straight lines, there are points where the walkers move off line and down a side road in order to cross at a crossing away from the junction. The most noticeable instance of this is on the south side of Marylebone Road at the junction with Baker Street, where the pedestrian crossing is a short distance south down Baker Street (Figure 10). By recognising certain features present in the GPS data, it is possible to determine which side of the street the samples were taken and also identify when the sensors reach junctions.

When crossing Marylebone Road at the junctions to the east or west of the council building, there is a characteristic that often appears in the GPS trace due to the position of the crossings relative to the road (Figure 11). The significant deviations to the left are a function of where the crossings are placed in relation to Marylebone Road. Both GPS paths are significantly out of position, placing the pedestrian in the middle of the road while he is still waiting to cross. This fact can be seen when the traces are played back in real-time by noting where the GPS position is stationary while the pedestrian is waiting to cross.

In both these situations, once the behaviour has been identified, appropriate route corrections can be applied. The trace in Figure 10 can be positioned accurately from the shape of the crossing, but the rest of the trace is on the wrong side of the road. This can be corrected by fitting the trace to a line on the correct side of the road. Figure 11 shows a different situation where a strong GPS fix is offset by a constant amount. Matching the shape of the GPS trace to the junction layout and subtracting this from the GPS coordinates can remove the offset.

5.4 Recorded Carbon Monoxide Levels When Crossing Roads

As all the pedestrian routes involve crossing roads at some point, there is a significant amount of data available in the middle of the road. The problem is how to detect when a sensor crosses the road due to the poor quality of GPS positioning information. One route stands out as a good example of road crossings, which was recorded on 10 May 2004 from 15:39, see Figures 12 and 13.

The actual positioning of these points is very hard to do from the GPS position alone and is determined more by when the GPS is stationary or additional satellites suddenly come into view, rather than the actual reported position.

The final peak on Figure 12 is marked as point X as there is no GPS and there is no additional information to identify this point, other than that it is somewhere along Gloucester Place.

It is worth noting that there are eleven crossing points when walking around the cycle route, not all of which produced peaks in this example and where there was a peak at a junction, it might disappear on the next circuit. For these reasons, using CO data to indicate the route taken can only be performed when there is extensive knowledge of the traffic flow patterns. Scientists working in
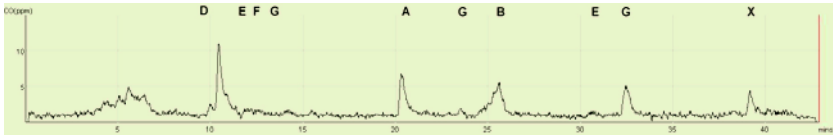
Fig. 12. Carbon monoxide levels while walking around the cycle route

Fig. 13. Map showing positions of carbon monoxide trace

this area can tell from the CO data when one of the pedestrians skips part of their assigned route. This is done using knowledge of the mean CO values for each road, together with where the CO peaks occur.

6   Dealing with Tracking Data

In Section 5 we saw several features in the data that allowed us to determine corrections to GPS log files. To summarise:

- Noting readings that were inside buildings.
- Noting that logs could be constrained to paths.
- Noting the micro-scale behaviour of the carriers when crossing a road allowed us to identify that a crossing was taking place and, in some cases, which side.
- Noting disambiguating paths by noting a particularly high CO reading.

We can generalise these to suggest there are five levels of knowledge that can be exploited when observing GPS logs in order to improve understanding:

- Knowledge of geographic extent of trial. We can discard some GPS readings because they are not in the geographic vicinity. This can be done as a pre-processing stage in the database.

- Knowledge of the building footprints. By overlaying GPS readings over a bitmap showing places where the carrier could not have received a signal we can easily see which locations need correcting. Note it is not appropriate to simply discard the points, as they may only be a few metres inside a building for example. These points can be corrected semi-automatically by moving them to the closest viable position.
- Knowledge of the route undertaken. Once we have filtered based on the impossible location, we can constrain the route based on route knowledge. We have made the distinction from the above, because at this stage we start identifying probable regions where the device was located, not impossible. There are many potential implementations here. Ours uses proximity to a route centre line, but it could be any region or line defined in a Geographical Information System (GIS).
- Knowledge of the behaviour of the carrier of the GPS receiver. At this stage we employ knowledge that is difficult to capture in GIS systems: behaviour of the users in the space. This requires some human knowledge. Although interesting artefacts such as velocity changes can be flagged in the data, a human observer is really needed to interpret this data. Again, a number of tools could be built, ours allows us to watch scaled-time replays of the event logs to get a sense of the behaviour.
- Knowledge of expected correlation with other sensor readings. At this stage we employ knowledge about other readings that can allow us to reason about positions on paths or corrections to paths. In other cases this could be considered implemented in a semi-automatic manner if framed as a sensor fusion problem. However at this time we are interpreting these observations by hand because the observer needs to know the expected effect and the context in which the readings were taken.

7   Conclusions

In this paper we investigated techniques to correct GPS logs. The need for correction arose because GPS readings are only accurate to roughly 20m, and because the urban environment causes quite significant variation in the behaviour of the readings. Several different types of knowledge have been exploited to correct the logs: geographic region, building footprints, routes and user behaviour. We have shown that by semi-automatically correcting the GPS logs, we can start to observe phenomena at a much smaller scale. In the previous section we discussed how these tools might be more generally applicable to other sensor tracking trials.

CO has proved interesting because it shows variation at a scale we can detect without expending effort in correcting GPS traces, but if that effort can be made, we can also find smaller-scale geographic effects. In our case, correcting GPS logs has allowed us to identify features on a 5m scale. This is sufficient for us to note expected properties of CO pollution: differentials across a street on a windy day and increases in CO near bus stops and road junctions.

Acknowledgements

References

1. Department for Environment, Food and Rural Affairs (Defra): The air quality archive. http://www.airquality.co.uk/ (verified 2005-02-09) (2005)
2. Croxford, B., Penn, A., Hillier, B.: Spatial distribution of urban pollution: civilizing urban traffic. 5th Symposium on Highway and Urban Pollution (1995)
3. Department for Environment Food and Rural Affairs (Defra): The air quality strategy for england, scotland, wales and northern ireland, 1999 (1999)
4. Discovery Net. Web Resource (2005) http://ex.doc.ic.ac.uk/new/index.php (verified 2005-02-09).
5. DAPPLE (Dispersion of Air Pollutants and their Penetration into the Local Environment). Web Resource (2005) http://www.dapple.org.uk/ (verified 2005-02-09).
6. O'Donnell, M., Watson, T., Fisher, J., Simpson, S., Brodin, G., Bryant, E., Walsh, D.: Galileo performance: Gps interoperability and discriminators for urban and indoor environments. GPS World, June (2003)
7. El-Sheimy, N.: Integrated systems and their impact on the future of positioning, navigation, and mapping applications. Quo Vadis International Conference, International Federation of Surveyors (FIG) Working Week, 21-26 May 2000, Prague. (2000)
8. Steed, A.: Supporting mobile applications with real-time visualisation of gps availability, s. brewster and m. dunlop (eds). MobileHCI 2004, LNCS 3160 (2004) 373–377
9. Steed, A., Milton, R.: Making maps of pollution using tracked mobile sensors. Under submission (2004)
10. Ochieng, W., Noland, R., Polak, J., Park, J.H., Zhao, L., Briggs, D., Crookell, A., Evans, R., Walker, M., Randolph, W.: Integration of gps and dead reckoning for real time vehicle performance and emissions monitoring. The GPS Solutions Journal, 6(4) (2003) 229–241