Visual analytics of GPS tracks: From location to place to behaviour

Jo Wood, *Member, IEEE*

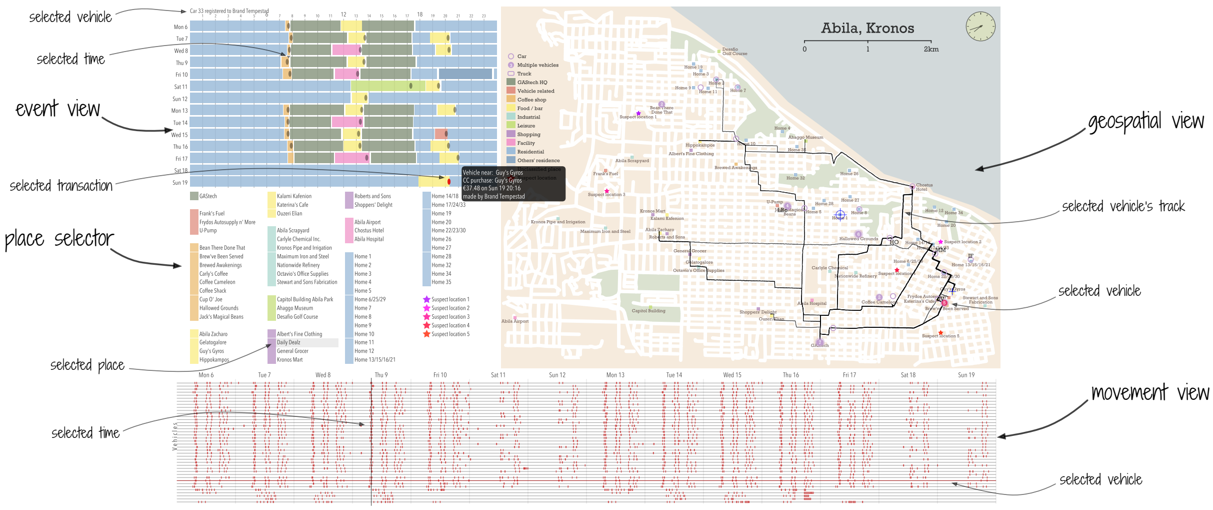

Fig. 1. Overview of the Visual Analytics system used to explore movement-derived behaviour.

Index Terms—Movement visualization, visual analytics, GPS, coordinated views, geospatial.

1 Introduction

The VAST 2014 Mini-challenge 2 [2] provided a set of GPS vehicle tracking and credit card transaction data with the aim of inferring behaviour of those using the vehicles and credit cards. In particular it required designing or applying a visual analytics system to identify typical behaviours with which atypical, suspicious activity could be contrasted. It is an example of the more general case of inferring behaviour from movement records (e.g., from mobile devices [6] or public bikeshare schemes [1]). In this brief paper we summarise the design taken to address the particular VAST challenge while proposing a general approach to movement-based behaviour detection.

2 Facet-oriented views

The primary data can be broken down into three fundamental types: geospatial (locations of GPS points, locations of road segments); temporal (GPS and credit card transaction timestamps) and attributes (vehicle owner and credit card payee). In turn these data types may be represented by primary variables such as easting, northing, day of year, hour of day, vendor type etc.). In total this yielded 11 primary variables to be represented visually. The design challenge became one of mapping these data variables to the appropriate visual variables in a way that assisted behaviour inference.

We followed a *faceted design model* [8] for creating different views of the data, each emphasising different fundamental types. The *geospatial view* mapped location of GPS coordinates to the (x,y) graphical space in a conventional mapping style (after projecting from longitude,latitude to UTM coordinates for easier distance calculations). This mapping can be formalised using the Hierarchical Visualization Expression language (HiVE) [7] : `sHier(/,$vehicle); sLayout(/,CA); sOrder(/,[$easting,$northing]);` This allowed spatial properties of behaviour to be made salient either by showing GPS tracks as lines (see Figure 1 right panel) or as animated movement of vehicle symbols.

The geospatial view supports spatial comparisons well, but is poor for temporal comparison, relying on visual memory of animated vehicle movement. A second faceted view – the *movement view* (bottom of Figure 1) – shows periods of movement for all vehicles over the full time period and acts as a selector both of vehicle (highlighted row) and time (highlighted column). Its HiVE data mapping is `sLayout(/,CA); sOrder(/,[$gpsTime,$vehicle]); sColor(/,$isMoving);` It is coordinated with the geospatial view to afford spatio-temporal comparison.

To link location with behaviour it was necessary to define and identify *place* – location that has some functional purpose for one or more individuals. Places were identified by performing automatic proximity analysis of all vehicles using hash grids for efficient detection [4]. Locations where any vehicle remained stationary (horizontal gaps between red bars in Figure 1) for longer than a user-defined threshold were used as candidate places. Multiple vehicles stopped within a user-defined distance to the same location increased the probability of the location representing a significant place. A place's function was identified by a combination of the temporal pattern of visits (e.g. home locations outside office hours) and credit card transactions (e.g. cafe purchases).

Once place types and locations had been identified, proximity detection could be used to automatically build the *event view* (top-left of Figure 1) using a calendar-type layout coloured by place type (`sLayout(/,SF); sOrder(/,[$timeOfDay,$dayOfYear]; sColor(/,$place);`). This allowed common routines of individuals to be easily identified (e.g. weekday home–coffee shop–work behaviours of many employees).

One further facet of the data was emphasised in the *co-location view* – using a similar temporal layout to

• *Jo Wood is at the giCentre, City University London, E-mail: j.d.wood@city.ac.uk.*

the movement view but colouring according to stationary location at known places rather than by movement (`sLayout(/,CA); sOrder(/,[$gpsTime,$vehicle]); sColor(/,$place);`). This provided a single view showing if, where and when multiple vehicles met at the same location and time (see Figure 2). It allowed social behaviours to be more easily identified (e.g., a Friday night party; a couple visiting tourist attractions together; post-work group visits to restaurants etc.).

Together, these coordinated facet views were used to build a picture of routine behaviour (e.g., weekday work routines); less common but benign activities (e.g., weekend leisure activities and social gatherings) as well as flag suspicious behaviour warranting further investigation (e.g., apparent surveillance of executive employees by security staff and visits to undeclared safe-houses).

3 Privacy and the Ethics of Behaviour Detection

There may be noble reasons for using tracking data to infer certain behaviours, but the development of systems to aid behaviour detection should be carried out with a sensitivity to ethical concerns. It is notable that in the scenario described in the VAST challenge, people were unaware they were being tracked via their vehicles. Indeed the value of the data is enhanced by that very ignorance. When combined with any form of coercion, such as might be applied by the fictitious 'GASTech' company in the VAST Challenge scenario, it is arguably a form of unregulated *geoslavery* [3]. Even where subjects voluntarily contribute location-based data about themselves, there may be ignorance of how those data may be combined with other data to construct a more complete form of surveillance (e.g., bicycle tracking data combined with Flickr and Foursquare checkins [5]).

The primary aim of the challenge was to detect possible criminal behaviour, yet in so doing it was necessary to identify other legal behaviour of individuals without their consent. Does the end justify the means? Is there a clear boundary between behaviour that justifies tracking and that which does not? Two activities were detected that sit somewhere closer to that boundary than either strictly illegal or routine benign behaviour. Figure 2 shows how repeated midday visits to the Chostus Hotel by two GASTech employees were detected using the co-location viewer. The tracking evidence might suggest a covert affair between the two of them, but should they have the right to do so, even if on 'company time'? What of alternative explanations such as making arrangements for hospitality, or the hotel car park being used to park the vehicles while conducting other business?

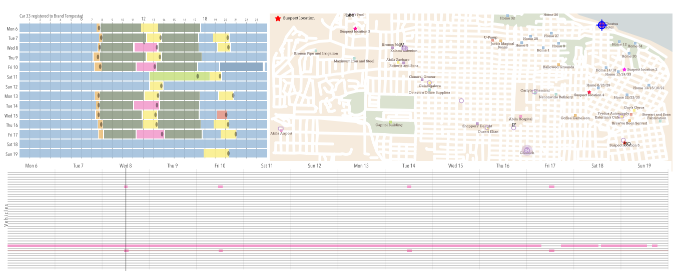

Fig. 2. Co-location at the Chostus Hotel. The short pink bars in the co-location view (lower part) show four visits to the hotel by Isande Borrasca and Brand Tempestad each around midday. The long pink bar represents a vehicle that remained at the hotel for most of the duration.

Using the movement view in combination with the geospatial and event views it was possible to detect unusual evening travel behaviour. In one case, an individual – Bertrand Ovan (Figure 3) – was prominent in driving between several bars within a two-hour period. Is this evidence of drink-driving? Of an alcohol problem? Perhaps not, but even if it were, should anyone have the right to look for such behaviour, and if found, do they have a responsibility to act upon the information?

Fictitious scenarios like that of the VAST Challenge allow us to explore and expose these ethical issues without direct harm, but we must be aware the danger of 'normalising' surveillance, especially where

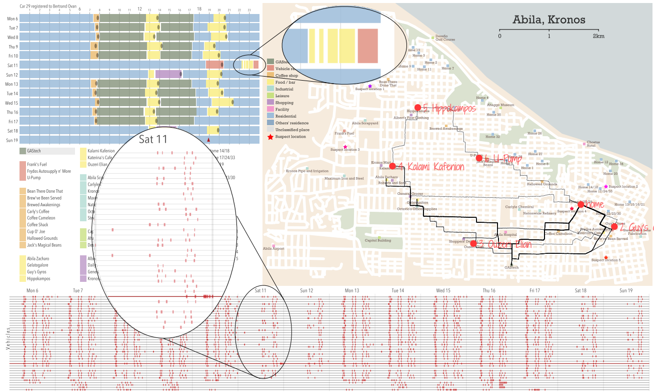

Fig. 3. Saturday night bar-cruise of Bertrand Ovan detected via the coordinated movement, event and geospatial views.

there can be uncertainty in behaviour inference. It suggests that capturing and symbolising that uncertainty may not only have a practical benefit, but that we are under a moral obligation to do so if the consequences of the decisions made have an impact on individuals. More positively, publishing advances in visual analytic techniques using such scenarios do provide a platform for informing the wider debate on privacy, surveillance and data ownership.

4 Conclusions

With data as simple as GPS track records and time-stamped credit card transactions, it is possible to build up quite detailed pictures of individuals' behaviours. This was achieved by projecting different facets of the data in a set of coordinated views, each emphasising a different quality of the data such as space, time or co-location. A formal mapping of the characteristics of those faceted views (via HiVE in this instance) allows the design space to be explored systematically in order to optimise the design to answer the tasks in question.

For addressing geospatial challenges such as this one, the transformation from location to place and then from place to behaviour has proven a successful strategy. Yet it raises concerns about geoprivacy, both influencing how we might design such visual analytic systems to protect individuals from incorrect inferences made about their behaviour, as well as forcing us to confront the ethics of surveillance and behaviour detection.

References

[1] R. Beecham and J. Wood. Exploring gendered cycling behaviours within a large-scale behavioural data-set. *Transportation Planning and Technology*, 37(1):83–97, 2014.

[2] K. Cook, G. Grinstein, and M. Whiting. Vast challenge 2014: The kronos incident. `http://vacommunity.org/VAST+Challenge+2014`, 2014.

[3] J. E. Dobson and P. F. Fisher. Geoslavery. *IEEE Technology and Society Magazine*, 22(1):47–52, 2003.

[4] A. Gregory, M. C. Lin, S. Gottschalk, and R. Taylor. A framework for fast and accurate collision detection for haptic interaction. In *ACM SIGGRAPH 2005 Courses*, page 34. ACM, 2005.

[5] J. Siddle. I know where you were last summer. `http://vartree.blogspot.co.uk/2014/04/i-know-where-you-were-last-summer.html`, 2014.

[6] A. Slingsby, R. Beecham, and J. Wood. Visual analysis of social networks in space and time using smartphone logs. *Pervasive and Mobile Computing*, 9(6):848–864, 2013.

[7] A. Slingsby, J. Dykes, and J. Wood. Configuring hierarchical layouts to address research questions. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):977–984, 2009.

[8] R. Walker, A. Slingsby, J. Dykes, K. Xu, J. Wood, P. H. Nguyen, D. Stephens, B. Wong, and Y. Zheng. An extensible framework for provenance in human terrain visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2139–2148, 2013.