Visual Analytic Design for Contextualising Sensor Data
VAST Challenge 2016 Award: Outstanding Presentation of Patterns in Context

Jo Wood*
giCentre (http://gicentre.org), City University London

ABSTRACT

The design choices behind a visual analytic system for monitoring a geosensor network are described and justified. Sparse spatial but fine-grained temporal data drive the layout choices privileging temporal pattern and classified location symbolised by colour. Additional context is provided by clustering space-time signatures and ordering data by cluster size. Uncertainty in classification is reflected in interpolation of categorical colours and interaction allowing multiple mappings of high dimensional clusters onto a one-dimensional ordering.

1 INTRODUCTION

This paper explores some of the processes involved in using visual analytics to contextualise a rich multi-dimensional set of sensor data. It is illustrated using the data and scenario described in the VAST Challenge 2016 Mini Challenge 2 (`vacommunity.org/VAST+Challenge+2016`). The challenge involved identifying routine and anomalous behaviour based on two groups of sensor data in a three storey building. Movement data of 115 employees were provided via ‘prox card’ sensors at fixed and mobile locations that registered when prox cards carried by staff were detected within a short distance of the sensor. Environmental conditions within the building were measured by 419 Heating, Ventilation and Air Conditioning (HVAC) sensors. Static prox card and HVAC sensors had a spatial precision that was generally coarser than individual room locations, being divided into ‘zones’ that comprised aggregations of between 1 and 30 rooms. Figure 1 shows an example of the prox and HVAC zones on the second floor. Data from all sensors were provided for a two week period with a temporal resolution of 5 minutes for fixed prox and HVAC sensors and 1 minute for the mobile prox sensor.

2 CONTEXTUALISING GEOSENSOR NETWORKS

Analysis of the spatio-temporal patterns of staff movement and environmental conditions from the sensor data falls into a category of problem common in handling distributed geospatial networks [2]. That is, the desired spatial and temporal resolution for analysis is finer than that provided by the sensor network. The sensor network itself is largely at fixed spatial locations (generating so-called ‘checkpoint’ data) and so movement can only be inferred as sensor readings are combined to identify transitions from one sensor zone to another. This not only requires an additional data processing stage but also introduces a degree of spatial and temporal uncertainty in inferred patterns as potentially important behaviour within zones or for short periods of time may remain undetected. While there are some analytic approaches to handling this problem (e.g. modelled interpolation [6, 3]), visual analytics provides the opportunity to contextualise the known data points and so allow the analyst to perform the intelligent interpolation required to infer behaviour. Three elements of visual analytic design were applied to the building sensor data, namely use of a common temporal layout; classification of location by function; and clustering of data by common behaviour.

*e-mail: j.d.wood@city.ac.uk

Figure 1: Prox zones (top), HVAC zones (bottom) and room layout of floor two of the monitored building.

2.1 Temporal Layout

Routine and anomalous patterns in movement and environment will be characterised by both spatial and temporal footprints. Given that the temporal precision (at least to the nearest 5 minute interval) is much finer than the spatial precision, the visual layout of the design privileges time over space, mapping time to the horizontal axis and employee to each row in the vertical axis (see Figure 2). The same layout supports display of HVAC sensor readings that can be overlaid onto employee location bars to support comparison. Interaction allows any individual employee to be selected and their classified location to be arranged in a calendar-type view [5] aligning days of the week vertically and hours of the day horizontally (see Figure 3). This provides the opportunity to make day-by-day comparisons that often reflect common weekday patterns of movement while highlighting anomalous patterns that deviate from that routine.

2.2 Functional Classification

Colour hue is used to symbolise the period spent in any given location with bars of higher saturation symbolising prox card entry/exit

Figure 2: Employee view showing employees' location in their 'home' office zone (pink) or outside their home zone (blue) organised vertically by cluster and horizontally by time.

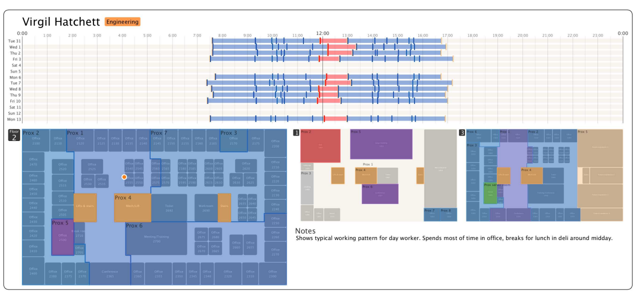

Figure 3: Single employee view showing calendar-type layout and classified locations.

points and lower saturation for interpolated location between prox card readings. While the prox card zone (Figure 1) can be mapped to hue, this provides little direct indication of the likely activity of each employee. So zones were additionally classified by their dominant function (e.g. office space, conference facility, dining etc.) and each given a unique hue evenly spaced in perceptual colour space. Where zones comprised more than one function (e.g. combined office and conference space), colours were linearly interpolated between classes to give some visual indication of classification uncertainty. Finally a zones were classified as being either a 'home' location in which an employee's regular office was located, or an 'away' location in a different prox zone (see Figure 2).

2.3 Cluster-driven Layout

To support both characterisation of routine behaviour and anomalous departure from routine, every employee's two-week pattern of zone location, home-away location and location function was represented as a multidimensional vector and clustered using multiple runs of k-means++ [1]. Despite the relative stability of this form of clustering, large vectors can produce different clusters when run on the same dataset so interaction permitted both re-runs of the clustering and dynamic changes in the number of clusters in order to visually assess their stability. By ordering employees by cluster size, common and anomalous patterns were separated visually (common patterns towards the bottom, unique patterns towards the top).

3 Conclusion

Analysis of the imprecise geosensor 'checkpoint' network data and comparison between staff movement and building environment was facilitated by a set of simple design choices that (i) reflected the importance of change over time in the inference of critical behaviour; (ii) used clustering and layout to automatically separate routine from anomalous movement patterns; (iii) recognised the uncertainty in classification through both symbolisation and the role of interaction. An interactive visual analytic system built following these design elements provided a means of exploring patterns; hypothesising causal and associative relationships within the data set; recording the provenance of observed patterns and inferences [4]; and communicating findings by the analysts.

References

[1] D. Arthur and S. Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[2] M. Duckham. *Decentralized spatial computing: foundations of geosensor networks*. Springer Science & Business Media, 2012.

[3] M. Duckham, M. van Kreveld, R. Purves, B. Speckmann, Y. Tao, K. Verbeek, and J. Wood. Modeling checkpoint-based movement with the earth mover's distance. *Lecture Notes in Computer Science*, 9927, 2016.

[4] R. Walker, A. Slingsby, J. Dykes, K. Xu, J. Wood, P. H. Nguyen, D. Stephens, B. W. 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.

[5] J. J. V. Wijk. Cluster and calendar based visualization of time series data. In *In INFOVIS '99: Proceedings of the 1999 IEEE Symposium on Information Visualization*, pages 4–9. IEEE Computer Society, 1999.

[6] J. Wood. Visualizing personal progress in participatory sports cycling events. *IEEE Computer Graphics and Applications*, 35(4):73–81, 2015.