

Identifying Characteristics of Collective Motion from GPS Running Data

Zena Wood¹, Antony Galton¹

¹College of Engineering, Mathematics and Physical Sciences, University of Exeter, EX4 4QF, UK
Email: {zmwood/apgalton}@ex.ac.uk

1. Introduction

Much of the research that has been carried out into movement patterns focuses on one level of granularity (Andrienko and Andrienko 2007, Dodge et al. 2008). However, our research into collectives has shown that this is not adequate when representing collective motion (i.e., the motion exhibited by a collective). We have developed a framework (the Three Level Analysis (TLA) framework) that analyses the motion of a collective on three levels of spatial granularity.

The ultimate aim of the research reported in this paper is to develop a system that, using a combination of visual and automatic analysis, identifies whether or not a collective is present within a data set and if so, the type of collective it is. The TLA framework could be used as the basis of such a system but to do so, within the dataset, it must be possible to identify and extract the movement patterns occurring at each of the three levels as set out in the framework; and, to identify and extract the set of episodes occurring at each of these three levels. To test whether this is possible GPS data has been collected from the runners of various races around Exeter; it is this data set that is presented within this paper.

A brief overview is given of the TLA framework (section 2) followed by why we believe this framework to be a possible bench-mark data set (section 3). Section 4 outlines how we are using the data set to address the problem of validating our framework. The paper concludes with a statement of how we believe an analysis of our current results could allow the identification of collective motion within a data set (section 5).

2. The Three Level Analysis (TLA) Framework

Although a collective comprises a group of individuals, when analysing the movement pattern of a collective, it is not sufficient to simply consider the aggregated motions of the individual members; the motions of the individuals may be qualitatively different from that of the collective (Wood and Galton 2009). Consider a crowd which slowly drifts east. The collective, when considered as a single unit, can be observed as moving in an easterly direction but the individuals as moving around randomly.

The TLA framework examines three levels of a collective's motion: the movement of a collective when considered as a point, the evolution of the region occupied by the collective (referred to as the footprint) and the movements of the individuals. These three levels can be thought of as three distinct levels of granularity.

The way in which the motion is described is dependent on the granularity at which it is observed. Within the TLA framework this is accounted for by defining a suitable set of *episodes* for each of the three levels of granularity; an episode is a maximal chunk of homogeneous process at a given level of granularity. This approach could be seen as similar to the use of primitives by Dodge et al. (2008) and Andrienko and Andrienko (2007). However, unlike these approaches, the TLA framework allows the movement pattern to be observed at multiple levels of granularity where at each level

different episodes may become apparent. A more detailed account of the TLA framework can be found in (Wood and Galton 2010).

3. The Proposed Data Set

The runners of a race form a collective where their movement patterns are crucial— you cannot have a race where the individual runners are not moving! Since many products are now available to runners that allow them to record GPS data from their races or training sessions, it is possible to collect a large data set. We focussed on the most popular brand of GPS running watches (Garmin) and asked runners to volunteer a copy of their data after the race for research purposes. All data is anonymous and, therefore, we found that many people were happy to volunteer their data.

Each of the Garmin GPS watches records the data in one of two formats: GPX or TCX. Both are XML formats but the former is a lightweight version that only records the essential information at each time step: longitude, latitude and elevation. TCX files allow additional information to be recorded such as calories and heart-rate. The relevant XML schemas for both formats have been published by the company and are in the public domain. As well as easily converting between the two formats, TCX and GPX files can be converted into KML format; this allows the data to be overlaid onto a map and visually analysed. Software has been written to perform this conversion but also to allow the data from multiple individuals to be overlaid onto the same map.

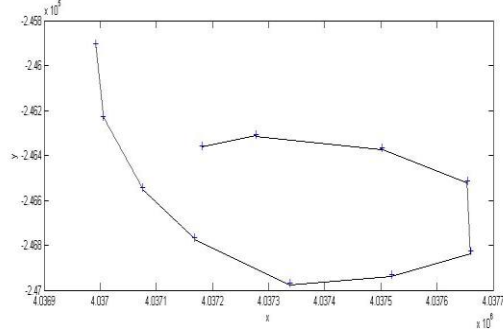
Along with the advantages of a specified structure and the possibility of collecting a large data set, another benefit of our data is that the sampling rates are frequent. Each user has the option of automatic or manual time recording. The former results in a new sample being recorded each time there is a significant change in speed or direction. The latter allows the user to record data at one, two, three or four second intervals. An examination of the data that we have collected indicates little deviation in the sampling rates between those that have chosen automatic recording compared to manual.

4. Searching for Collective Motion within the Data Set

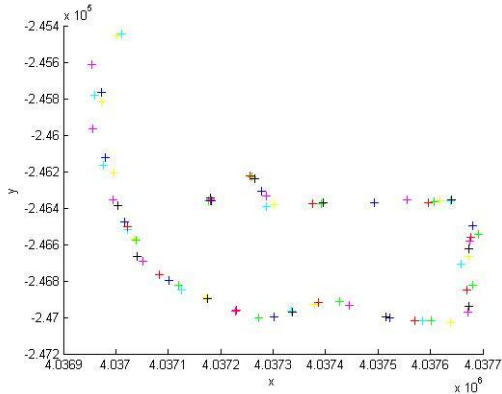
We believe that by examining a data set using the TLA framework it may be possible to identify the presence of a collective by searching for characteristics of collective motion. However, a pre-requisite for this is the extraction of the three movement patterns and the identification of the episodes that they each comprise. One of us (Wood) has written a computer program that will take in all of the data gathered from a particular race and apply the TLA framework. This section details the current results from this program.

4.1 Extracting the Movement Patterns

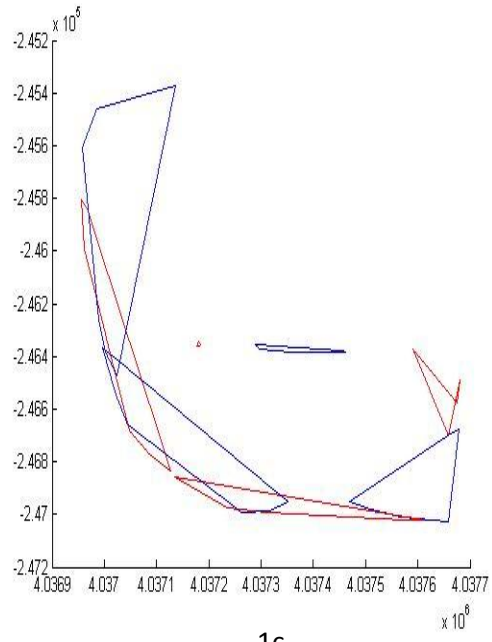
Automatic processing has been used to extract the necessary movement patterns from the data set. We have GPS data for each individual. To observe the movement of the collective when considered as a single entity the group's centroid, computed as the average position over all the members of the group, has been taken as a representative point. Methods that establish the footprint at each time step have been proposed by Dupenois and Galton (2009, 2010). Currently, the program calculates the convex hull of the group of individuals and uses this as its footprint. More sophisticated footprint algorithms will be used in future research. Figures 1a, 1b and 1c show the movement patterns that have been extracted for the individual, centroid and footprint respectively.



1a



1b



1c

Figure 1. The three extracted movement patterns over the first 1000 time steps.

4.2 Identifying episodes

Since the relevant episode-types are pre-defined within the TLA framework, they can be found within the dataset through computation. Simple examples include: when the position of the centroid is the same at t_1 and t_2 , an episode of uniform motion has occurred during this interval; and when the size of the footprint increases during an interval, an episode of expansion has occurred. Figure 2 is an example of one of the graphs that has been output by the program. This graph analyses the speed of the centroid's motion at different levels of granularity (vertical axis) over time (horizontal axis). The episode-types are identified by colour-coding. However, for this example, numbers two, four, and five have been used to aid understanding with these numbers representing the episode types: acceleration from start, acceleration and decelerated motion respectively.

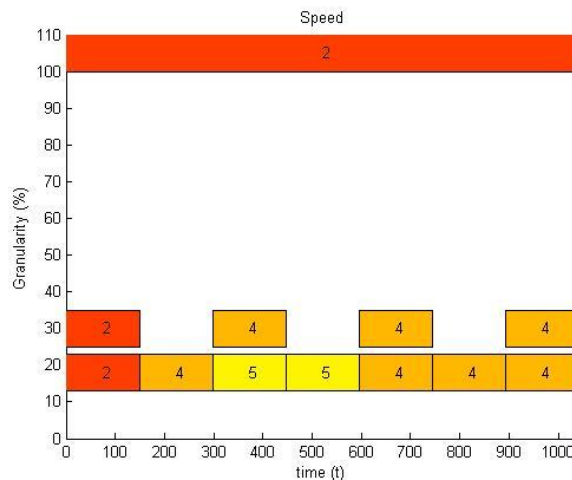


Figure 2. The analysis of the centroid's motion when considering the speed of the motion.

5. Further Work

The program that has been written can extract the three movement patterns from the data set, as given in the TLA framework, and analyse each pattern according to the set of pre-defined episode-types. However, to determine whether these movement patterns indicate the presence of a collective, further analysis is needed to establish the characteristics of collective motion. Such characteristics could be found by identifying any relationships that may exist between the three extracted movement patterns. For example, if the motions of the individuals and the centroid are qualitatively similar it could be considered as evidence of coherence and, therefore, that the individuals are part of a collective. In comparison, when the motions are qualitatively distinct evidence exists of minimal coherence. However, consider the dancers around a maypole. They all move around the pole but the centroid of the group would appear stationary. These two motions are qualitatively distinct but there is a relationship between the movement of the individuals and the centroid of the group. This is emphasised by an examination of the evolution of the footprint which would be seen as expanding and contracting as the dancers moved away from and towards the pole.

If it is to be established what type of collective is present, the different characteristics that each type of collective exhibits must also be established. However, the analysis of one data set is not sufficient to establish this information. More data must be gathered where other types of collective may be present, the TLA framework applied and the results analysed. For example, a transition matrix could be produced for each type of collective that shows the probabilities for each episode type being followed by one of the other pre-defined episode-types; each type of collective could have an identifiable transition matrix.

6. Conclusion

Collective motion must be considered on more than one level of granularity if it is to be sufficiently analysed. A data set has been collected and the TLA framework applied. Three movement patterns have been extracted from this data and analysed according to the pre-defined episode-types within the framework. However, if a program is to be produced that allows the use of the TLA framework to identify the presence of a collective and its type, the characteristics of collective motion must be identified by examining the relationships that exist between the extracted movement patterns. The TLA framework must also be applied to additional data sets that may contain different types of collective (i.e., not runners).

References

- N. Andrienko and G. Andrienko, (2007). Designing visual analytics methods for massive collections of movement data. *Cartographica*, 42(2):117–138.
- S. Dodge, R. Weibel, and A.K Lautenschütz, (2008). Towards a taxonomy of movement patterns. *Information Visualization*, 7:240–252.
- Z. Wood and A. Galton, (2009). Classifying Collective Motion. In Björn Gottfried and Hamid Aghajan (Eds.), *Behaviour Monitoring and Interpretation - BMI: Smart Environments*, IOS Press, Amsterdam, 129-155.
- Z. Wood and A. Galton, (2010). Zooming in on Collective Motion. (Submitted to: *Workshop on Spatio-Temporal Dynamics* (ECAI 2010)).
- M. Dupenois and A.P. Galton, (2009). Assigning Footprints to Dot Sets: An Analytical Survey, COSIT'09, Aber Wrac'h, Brittany, France, 21st - 25th Sep 2009. *Spatial Information Theory*, vol. LNCS 5756, 227-244.
- M. Dupenois and A. Galton, (2010). The Use of Change Identifiers to Update Footprints of Dot Patterns in Real Time (Submitted to: *Workshop on Spatio-Temporal Dynamics* (ECAI 2010)).