Extracting Geometric Representations Of Trajectories Using Topological Data Analysis

Thomas Edwards*[1], Christopher B. Jones[1], and Padraig Corcoran[1]

[1]School of Computer Science & Informatics, Cardiff University

*Email:EdwardsTJ1@cardiff.ac.uk

Abstract

Object tracking finds applications in many research areas. Examples include tracking weather phenomena such as a snow storm and tracking migration data and identifying migration patterns of bird species. However, tracking objects is a challenging task since an object's topological properties can change over time. Previous research proposed a method for tracking objects with dynamic topology, based on using zig-zag persistent homology principles. Our paper builds on that research by using the method for identifying objects within a dataset in order to extract and visualise trajectories of the moving objects. We create the trajectories based on the centroids of the objects in each time step. We also perform trajectory clustering for reducing noise in data and identifying main movement patterns. The approach is demonstrated with respect to tracking rain clouds in radar imagery.

Keywords: topological data analysis, trajectories representations, trajectories clustering.

1 Introduction

Object tracking finds applications in many research problems. For example, when making inferences concerning future weather conditions, it is necessary to track weather phenomena such as a snow storm. Another application is tracking migration data and identifying migration patterns of wildlife. However, a big challenge of tracking objects with dynamic topology is that the object's topological properties can change over time. For instance, splitting an object into multiple objects or the merging of multiple objects into a single object (Corcoran and Jones, 2018). A robust method is therefore required to keep track of the appearance and disappearance of individual objects.

Corcoran and Jones (2017) presented a model that encodes the spatio-temporal characteristics of topological features of objects, such as holes and connected components. The persistence of topological features with respect to time is computed using zig-zag persistent homology. Zig-zag homology gives a set of intervals representing the periods of existence of the topological features in question (Carlsson and De Silva, 2010). In order to facilitate statistical and data mining techniques the set of intervals are converted into a persistence landscape. A persistence landscape is a vector space representation of topological features, which makes it easy to be combined with statistical and machine learning tools (Bubenik, 2015). Bubenik introduces a set of different algorithms for calculating persistence landscapes in (Bubenik and Dłotko, 2017).

*

Corcoran and Jones (2017) identifies objects that persist between successive time slices and records the start and end duration of each object across the times slices. A limitation of the standard method employed for computing these persistence intervals is that it is only capable of inferring the appearance and subsequent disappearance of objects but does not maintain their identities relative to their respective regions (components) in the source image. Corcoran and Jones (2018) extend these persistent homology methods to attach unique identifiers to objects with dynamic topology, from their creation to disappearance, keeping track of the image locations of the objects. When one object is merged with another, one of the objects will lose their original identity, while when an object splits one or more separate identities will be attached to the newly spawned objects (depending on how many there are). It may be noted that earlier work that implemented methods for tracking topological change, notably (Worboys and Duckham, 2006), employed a rule-based approach that was acknowledged as not being complete with regard to all possible change situations.

The widespread use of location-aware devices has led to an increasing availability of trajectory data. As a result, researchers have devoted efforts to developing analysis methods including different data mining methods for trajectories (Yuan et al., 2017), (Mazimpaka and Timpf, 2016). Data mining techniques depend on the type of objects whose trajectory is in focus, e.g. people, animals, weather data, and the application, e.g. hot-spot discovery, extraction of mobility profile, discovery of interaction between animals (Mazimpaka and Timpf, 2016). Main types of analysis are classification, clustering, frequent pattern mining and group pattern mining (Mazimpaka and Timpf, 2016). Clustering is a popular method for analyzing trajectories because it provides useful insight into data without the need for a training set (Mazimpaka and Timpf, 2016). Trajectory clustering aims at finding trajectories that are of the same (or similar) pattern, or distinguishing some undesired behaviours, such as outliers (Yuan et al., 2017).

Our aim is to illustrate how the topological data analysis-based methods developed in Corcoran and Jones (2018) can be used to extract trajectories of identified objects. We create trajectories by calculating centroids of object's regions at each time slice and then connecting the centroids. We also perform trajectory clustering in an attempt to identify similar trajectories and main movement patterns. We apply the methods to weather data, specifically tracking rainfall in radar imagery. The images are obtained from the UK Meteorological (Met) Office. The Met Office provides this data at 15 minute intervals. For a given time, the image data in question categorises the rainfall level at each location in a 500x500 regular grid over Ireland and UK. Given this data, we consider the problem of tracking objects corresponding to spatially close path-connected components of $\mathcal{R}^2$ with a rainfall level greater than a threshold (Corcoran and Jones, 2018).

2 Methodology

The methodology uses the objects' locations produced by the application of Corcoran and Jones (2018) as an input. It consists of 2D array representations of cloud images, where the two indices represent the x and y pixel coordinates while the value of each array element is the unique identifier of the respective pixel at that location. All pixels that belong to the same object (i.e. cloud region) will have the same unique identifier, where that identifier serves simply to distinguish the objects from each other. Each object persists across a consecutive sequence of time slices starting at time slice T_1 and finishing at time slice T_n. These persistence intervals are provided as output from the zig-zag persistent homology procedure.

An overview of the methodology is given in Figure 1. The methodology consists of 5 main steps.

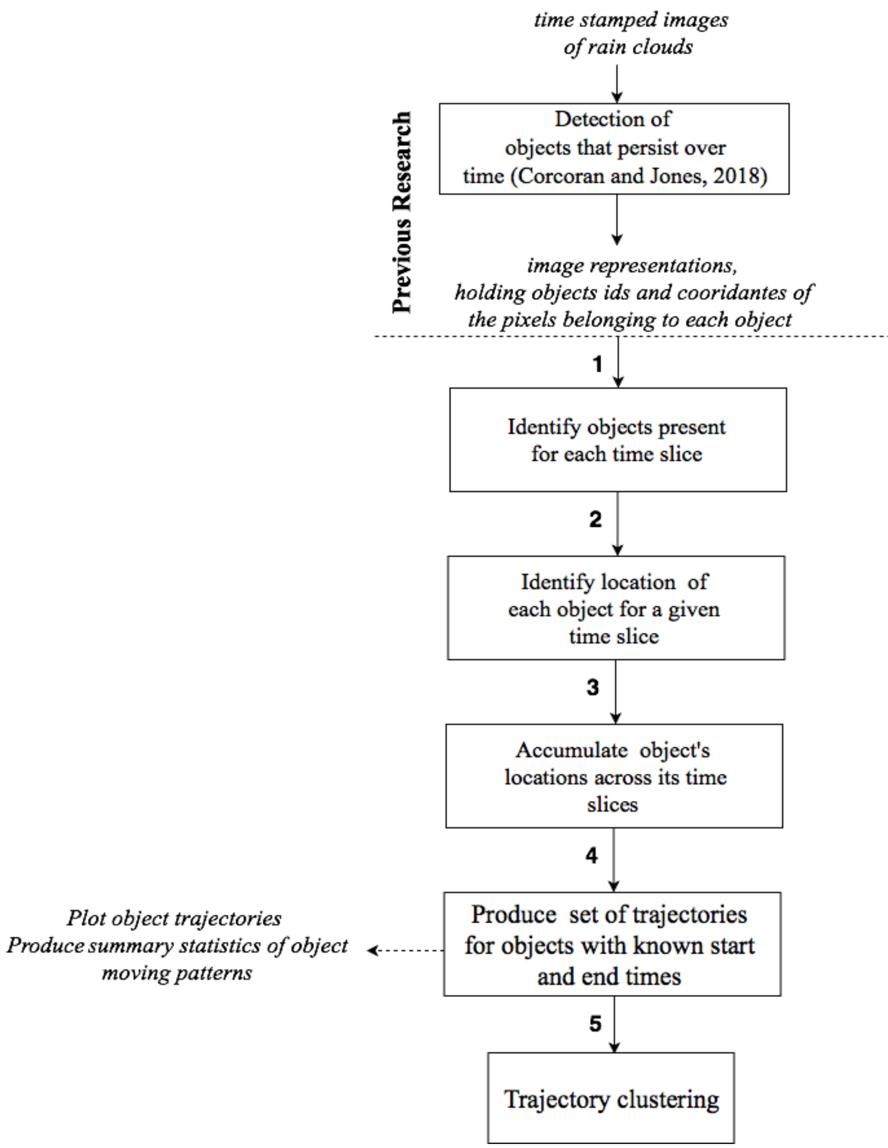

Figure 1: Methodology overview

In *Step 1* we identify objects present for each time slice. We do this by finding all the pixels that belong to the same object for a given time slice. We also remove pixels with no objects in them (i.e., they have value equals '0'). At the end of this step, each object id is associated with a list of all the x and y coordinates of the pixels that belong to this object for a given time slice.

In *Step 2* we identify the location of each object for a given time slice. We approximate the location of the objects by finding their centroids. Each object is a finite set R of n elements. Thus, the centroid is the mean of the elements in the set R (see Equation 1). At the end of this step, we have each object associated with the x and y coordinates of the centroids of this object for each time slice. In *Step 3* we accumulate object's locations across its time slices. Specifically, we accumulate the sequence of centroid coordinates for each unique object across its time slices. In *Step 4* we produce a set of trajectories and visualize them.

In *Step 5* we perform clustering of the object trajectories. Our goal is to create clusters containing similar trajectories in order to identify movement patterns. We perform trajectory clustering by adapting the QuickBundle (QB) algorithm, originally created for use in magnetic resonance imaging to cluster white matter fibres. Each QB cluster can be represented by a single centroid streamline, which is a sequence of points. We selected this algorithm for the following reasons:

1. It provides results in timely-manner - the algorithm has been created to deal with large datasets and return results fast.
2. It has been specifically created for data from 3D imagery of white tissues, which resembles the structure of trajectories - the algorithm was created for simplifying tractography data of white tissues. Tractography data is a dataset composed of streamlines.
3. QuickBundle uses a symmetric distance function called minimum average direct-flip (MDF) distance which takes into account the sequential nature of streamlines - other distance measures treat streamlines as a bag of points where every point on the first streamline is to be compared with every point on the second streamline, and vice versa.

The number of clusters produced depends on adjusting a threshold value. High threshold values will produce less clusters with more trajectories in them while a small threshold value will produce smaller clusters. Before clustering, we defined a distance function between the trajectories using Euclidean distance. We chose Euclidean distance because of its simplicity and lack of a threshold that needs adjusting. The MDF measure requires trajectories to have the same length. Therefore, we re-sample the trajectories to have the same number of points. This is achieved using linear interpolation.

$$Object_[centroid] = \frac{\sum_{i=1}^{n} R_i}{n} \qquad \text{Equation 1}$$

3 Results

The methods were applied to meteorological data collected over a 12 hour period. Cloud images were recorded at 15 minute intervals, thus we have 48 time slices. After extracting the objects from these images with respect to time, we obtained six objects. We identify objects, as indicated above, using the unique identifier of the pixels that belong to the same objects from processed imagery data (Corcoran and Jones, 2018). In Table 1 we give summary statistics of the existence of the objects over time.

Table 1: Summary statistics of objects existence over time

Object ID	Start time	End time
Object 1	0	47
Object 2	0	47
Object 3	0	47
Object 4	11	37
Object 5	11	46
Object 6	22	22

Object 6 appears in only one time slice (see Table 1). Thus, it is not visible in Figure 2 and Figure 3 except as a dot.

3.1 Trajectories representation

Figure 2 and Figure 3 represent the trajectories of the moving objects, where Figure 2 provides linear representation and Figure 3 shows trajectories changes over time.

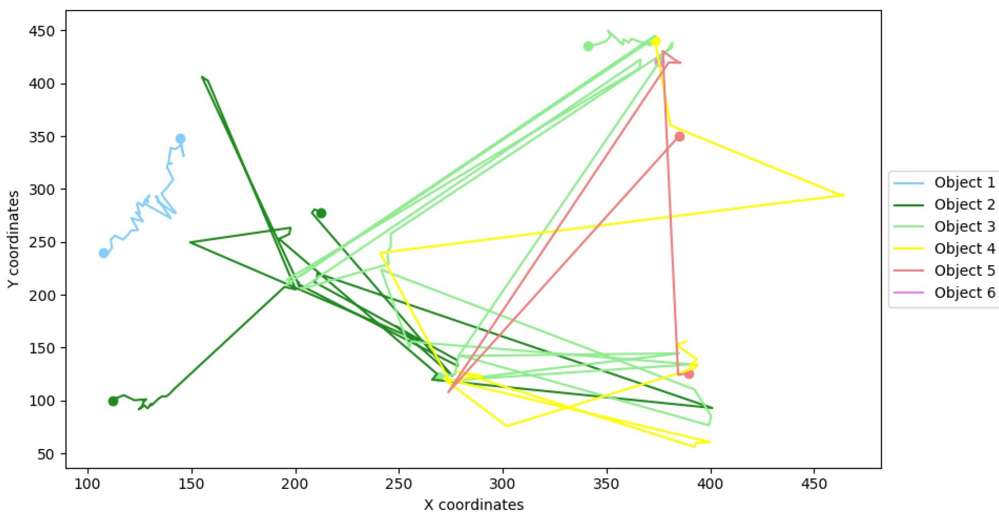

Figure 2: Cloud movement trajectories representation: line representation

The trajectories of the objects (see Figure 3) appear to be subject to some very sudden changes of location which is unusual for weather data. This occurs here as a consequence of objects merging and splitting between time slices. This is demonstrated in Figure 4 where we display object movements over three time slices. When large objects become connected it results in a dramatic change in the centroid of the merged object. Figure 4 demonstrates this with the yellow object merging with the blue object merging in "time 2", then in "time 3" the blue object splits creating the red object. Once an object is merged into another object or it disappears in a time slice, it is destroyed. An object cannot re-appear once it has been destroyed, according to the persistence homology approach used for identifying objects over time.

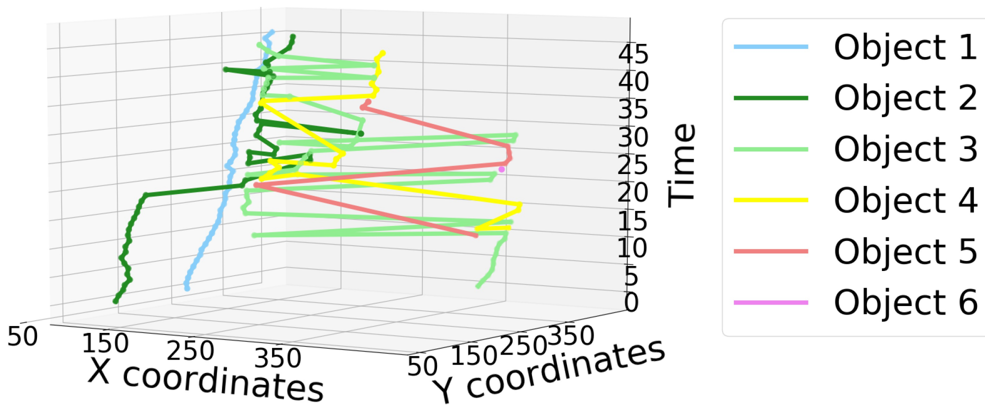

Figure 3: Cloud movement trajectories representation:space time cube representation

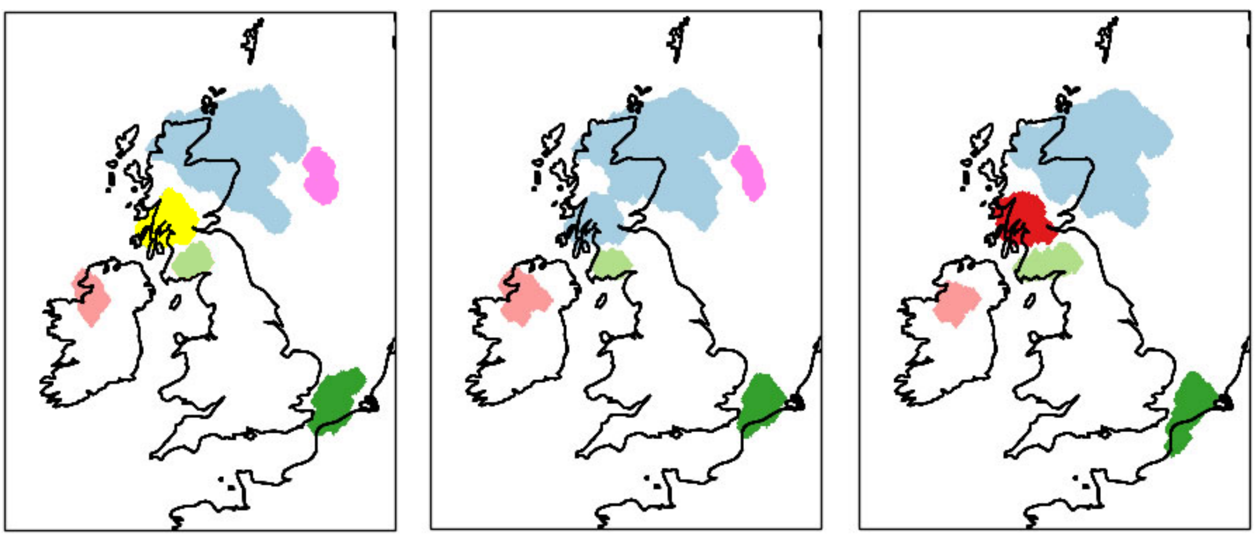

Figure 4: Original photo images: Object at time 1 (left), Object at time 2 (middle), Object at time 3 (right)

3.2 Trajectories clusters

In order to limit the noise in the data and identify similar movement patterns we performed clustering of the object's trajectories (see Figure 5). A summary of the clusters is given in Table 2.

Table 2: Summary statistics of objects existence over time

Cluster ID ID	Trajectory ID
1	1
2	2
3	3,4,5,6

In Figure 5 all object trajectories that belong to the same cluster have the same colour.

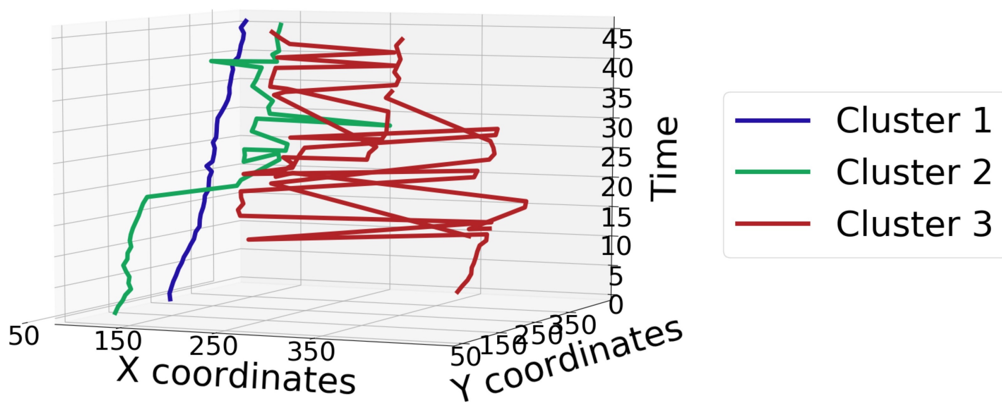

Figure 5: Cloud movement trajectories representated with a space time cube

4 Conclusions and Future work

The paper presented a way of using existing topological data analysis methods for tracking spatio-temporal phenomena in order to extract the geometric representations of trajectories. The results show that topological data analysis methods for identifying objects with dynamic topology can be used for extracting object's trajectories and illustrating object movement. Clustering the trajectories helped reduce noise from data, identify similar trajectories and find patterns of movement.

With regard to future work, the proposed model can be applied to studying movement patterns in other types of data such as bird migration. We will also look at developing improved clustering methods for summarizing trajectories in noisy data.

5 References

Bubenik, P.
2015. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102.

Bubenik, P. and P. Dłotko
2017. A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation*, 78:91–114.

Carlsson, G. and V. De Silva
2010. Zigzag persistence. *Foundations of computational mathematics*, 10(4):367–405.

Corcoran, P. and C. B. Jones
2017. Modelling topological features of swarm behaviour in space and time with persistence landscapes. *IEEE Access*, 5:18534–18544.

Corcoran, P. and C. B. Jones
2018. Robust tracking of objects with dynamic topology. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Pp. 428–431. ACM.

Mazimpaka, J. D. and S. Timpf
2016. Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science*, 2016(13):61–99.

Worboys, M. and M. Duckham
2006. Monitoring qualitative spatiotemporal change for geosensor networks. *International Journal of Geographical Information Science*, 20(10):1087–1108.

Yuan, G., P. Sun, J. Zhao, D. Li, and C. Wang
2017. A review of moving object trajectory clustering algorithms. *Artificial Intelligence Review*, 47(1):123–144.