Segmenting Trajectories by Movement States

Maike Buchin, Helmut Kruckenberg and Andrea Kölzsch

Abstract Dividing movement trajectories according to different movement states of animals has become a challenge in movement ecology, as well as in algorithm development. In this study, we revisit and extend a framework for trajectory segmentation based on spatio-temporal criteria for this purpose. We adapt and extend the framework to the setting of segmentation according to the individual movement states of an object, in particular an animal. We implement the framework and evaluate it at the example of tracks of migrating geese.

Keywords Trajectory segmentation · Experimental evaluation

1 Introduction

Due to advances in tracking technology, like GPS, a growing amount of trajectory data are being collected. Trajectory data occur in many contexts, such as sports, biology, traffic analysis, and defense. In particular, they occur in movement

M. Buchin (✉)
Department of Mathematics and Computer Science, TU, Eindhoven, The Netherlands
e-mail: m.e.buchin@tue.nl

H. Kruckenberg
European White-Fronted Goose Research Programme, Verden, Germany

A. Kölzsch
Department of Animal Ecology and Project Group Movement Ecology,
Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands

ecology (Nathan et al. 2008): biologists use GPS tracks of animals to aid in the under-
standing of their movement. To analyze large amounts of trajectory data, efficient
and effective methods are needed. *Automatic Movement Analysis* has emerged as a
research field addressing this challenge (Shamoun-Baranes et al. 2012).

One important task is *trajectory segmentation*: cutting the trajectory into smaller
pieces. This may have different purposes, e.g., data handling (efficient storing and
searching) or analysis (learning about the underlying movement). We focus on the
latter, and in particular on segmentation based on individual movement states of the
object (animal). That is, we are given a trajectory and a set of classes of movement
states, and we want to segment the trajectory according to these movement states.
For instance, we want to segment a bird trajectory such that the bird was flying,
foraging, or resting on each piece. More generally, this is a segmentation based on
the *semantics* of the trajectory.

Recently, Buchin et al. (2011) proposed a framework for trajectory segmentation
based on spatio-temporal criteria. We revisit this framework and extend it for the
purpose of trajectory segmentation based on the individual movement states of an
object. As a prototype example, we consider segmenting the trajectories of migrating
geese.

Our framework is semi-automatic in the sense that the parameters for segmen-
tation need to be input manually, and the resulting segmentation is then computed
automatically. Other approaches (e.g. Fauchald and Tveraa 2003; van Moorter et al.
2010) are fully automatic, that is, no user input is required. We see the strength of
our approach in that expert knowledge can be used to effectively describe movement
states. We will discuss this further in Sect. 4.

Results. We extend and implement the framework presented in Buchin et al. (2011)
for segmentation based on movement states of an object. In particular, we

- identify and discuss relevant criteria,
- adapt the framework for describing movement states,
- integrate a time criterion,
- allow outliers in the data,
- present the first implementation and experimental evaluation of the framework.

2 Extending the Framework

2.1 Existing Framework

The existing framework (Buchin et al. 2011) segments a trajectory such that
each segment is homogenous in the sense that a set of spatio-temporal crite-
ria are fulfilled. For this, it considers spatio-temporal *attributes*, such as head-
ing, speed, and location. It uses *criteria* on the attributes, such as bounding the
heading angular range, bounding the speed ratio, or requiring the locations to lie
in a disk of given radius. The criteria are required to be monotone in the sense

that if a segment fulfills a criterion, then so does every subsegment of the segment. Criteria can be combined, by a boolean or a linear combination. We call the result a *set of criteria*. The framework provides algorithms for segmenting a trajectory into the fewest number of pieces such that each piece fulfills the set of criteria.

The algorithms are greedy strategies: *incremental-search* and *double-and-search*. These greedy strategies can be described as follows. The algorithm starts at the beginning of the trajectory and finds the longest segment that fulfills the set of criteria. Then it starts new at the end point of the segment just found, and again finds the longest segment fulfilling the criteria. And so on. The incremental-search finds the longest segment by incrementing the test segment by one in each step. Double-and-search uses first an exponential search on the segment size, until this fails, and then a binary search between the last two points (the last where it succeeded and the first where it failed). The run time of both algorithms depends on the chosen criterion. In most cases, an overall $O(n \log n)$ run time is achieved, where n is the number of edges of the trajectory to be segmented (Buchin et al. 2011).

2.2 Analysis of Criteria

Attributes and Criteria. Although the framework presented in Buchin et al. (2011) allows for many criteria, we found the basic and intuitive attributes, speed, location, heading, and time, are well suited for describing movement states. For each, we give one (or two) intuitive criteria. Except for the time criterion, these criteria were suggested in Buchin et al. (2011). We will use these criteria in the next section for segmenting migrating geese data. Figure 1 illustrates the criteria location in disk and heading angular range.

Location disk criterion: all points of the segment must lie in a disk of a given radius.
Heading angular range criterion: the angular range of the heading vectors is bounded by a given angle.
Speed minimum/maximum value: all speed values of the segment must be larger/smaller than the given speed value.
Time minimum/maximum duration: the duration of a segment is larger/smaller than the given duration.

Properties. The criteria have (some of) the following properties.

- *monotone*: if a segment fulfills the criterion, then so does each subsegment.
- *linear*: all sample points of a segment need to be checked for the criterion.
- *constant-update*: checking a segment of size one larger can be done in constant time.
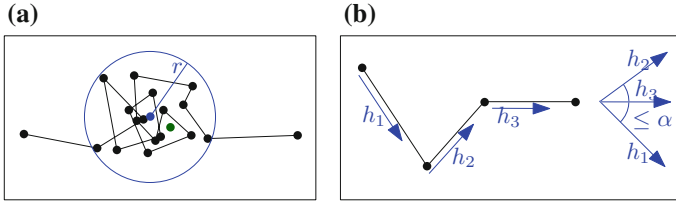- *relative/absolute*: a criterion that depends on relative/absolute values of attributes

Fig. 1 Illustration of criteria. **a** Location in disk (of radius r). **b** Heading angular range ($\leq \alpha$)

Table 1 Attributes and criteria and the properties they fulfill

Attribute	Criterion	Mono.	Line.	Cons.	Rela.
Location	Disk	Yes	Yes	No	Yes
Heading	Angular Range	Yes	Yes	Yes	Yes
Speed	Minimal value	Yes	Yes	Yes	No
	Maximal value	Yes	Yes	Yes	No
Time	Minimal duration	No	No	Yes	Yes
	Maximal duration	Yes	No	Yes	Yes

Mono. stands for monotone, line. for linear, cons. for constant-update, and rela. for relative

Monotonicity was already discussed in Buchin et al. (2011) and is required for the greedy strategies to be optimal (i.e., result in minimal segmentations). It seems to be a natural requirement, however, there are non-monotone criteria, such as the angle and tube criterion suggested by Turchin (1998) (see below), and criteria on average values.

The properties linear and constant-update affect the runtime of the algorithms. For a linear, constant-update criterion, incremental-search is asymptotically faster than double-and-search, (linear instead of $O(n \log n)$ run time). For a linear criterion with $O(\log n)$ update cost, incremental and double-and-search are asymptotically equally fast.

A relative criterion completely partitions the trajectory, whereas an absolute criterion may give unclassified segments, i.e., pieces not fulfilling the criterion. In a segmentation according to movement states, unclassified segments correspond to segments where the movement states could not be classified.

Table 1 shows which of the criteria listed above are monotone, linear, constant-update, and relative. These properties are all straightforward except for updating heading angular range in constant time, which we describe next.

Heading angular range can be updated in constant time for angles up to 180° using the fact that only one angle between two heading directions needs to be taken into account. That is, one can simply maintain the cone, smaller than 180°, containing all directions, as long as this exists. For larger angles, say $360 - \gamma$, we subdivide the space of angles into $360/\gamma$ ranges of size γ. In each range we maintain the smallest and largest heading direction in constant time and deduce the maximum range from this. The resulting update time remains constant, but the constant increases with decreasing γ.

Further Criteria. There are many further criteria. In Buchin et al. (2011) the authors propose also location diameter, curvature, curviness, sinuosity, and a shape fitting criteria. Turchin (1998) introduced an angle and a tube criterion. Both of the latter criteria use the line from start to end point of a segment as reference line and bound the angle or the location of the points in between with respect to this line. Because they depend on the line from start to end of a segment, these criteria are not monotone.

2.3 Combining Criteria

Describing Movement States. In Buchin et al. (2011) as combination of criteria, a boolean formula in conjunctive normal form or a linear combination was suggested. We propose instead using a boolean formula in disjunctive normal form for describing movement states.

A boolean formula in disjunctive normal form (dnf) describing k different movement states has the form

clause-1 OR clause-2 OR ... OR clause-k

where each clause has the form

criterion-1 AND criterion-2 AND ... AND criterion-l

where the number l of criteria depends on the described movement states. For instance, we may use a dnf formula for describing *flying* or *resting*. If we describe flying as minimum speed and minimum time duration and resting as maximum speed (for a simple example, without parameters), we get in total the formula

(minimum speed AND minimum time duration) OR maximum speed.

Choosing criteria. The number of clauses corresponds to the number of movement states to be used for segmentation (and are thus fixed). For describing a movement state, the more criteria are used, the harder the corresponding clause is to fulfill. Note that a criterion should only be used, if the data contains the corresponding information. For instance, currently GPS data does not contain accurate altitude data, thus, this criterion should not be used in segmentation.

2.4 Algorithms

Basic algorithms. In Buchin et al. (2011) two basic algorithms are suggested: *incremental-search* and *double-and-search* (see Sect. 2.1). Incremental is asymptotically faster for linear, constant-update criteria. For linear, $O(\log n)$-update criteria,

the two are asymptotically equally fast. Double-and-search is asymptotically faster for criteria with higher update cost.

Which of the two algorithms is faster for a specific application depends on the trajectory and the set of criteria. If some criteria are linear and constant-update, and some are not, it cannot be said in advance. A naive strategy is to use the algorithm that is better for a majority of the criteria.

Non-monotone criteria. For non-monotone criteria no efficient algorithm is known. For linear, non-monotone criteria, a linear search from the back finds a single longest segment. A linear search from the front (i.e., incremental-search) may result in a non-optimal segmentation in terms of size, i.e., more and shorter segments than necessary may be found.

Continuous Segmentation. For each algorithm, we have the choice between discrete and continuous segmentation, i.e., whether we segment only at sample points or also in between. Segmenting in between sample points in Buchin et al. (2011) is based on the linear motion model, i.e., interpolating linearly in between sampling points. Whether discrete or continuous segmentation is more suitable depends on the application, and in particular on the sampling of the trajectory. If a trajectory is densely sampled, then a discrete segmentation will not differ much from a continuous segmentation. If a trajectory is sparsely sampled, then a discrete and a continuous segmentation may differ considerably. In this case, segmenting in between sample points may be superior. However, for sparse sampling, the linear motion model is not realistic, which applies not only to the location but also to movement parameters (e.g., speed) in between sampling points. Therefore, in our study we chose discrete segmentation.

Extensions. For the two basic algorithms, we developed the following two extensions.

- allowing a min time criterion
- allowing a constant number of outlier

Time Criterion. Minimum time duration is an important, but non-monotone criterion, which is not handled by the original framework. With this criterion it is possible to filter out segments of short duration. Therefore we extend the framework to include it.

Two simple approaches for incorporating it in our framework are:

- Ignore a time criterion until all other criteria of a clause fail. Then test minimum time.
- First find the time threshold and test there.

The first method is simple, but non-optimal. We may discard segments after having tested criteria on them. The second method is faster, if we can find the time threshold, i.e., a segment of the trajectory of the given duration, efficiently (e.g., in regularly sampled data). However, in irregularly sampled data finding time threshold may be time-consuming as well.

Outlier. Outliers, or noise in the data, occur often in todays GPS data. They may influence the analysis, in our case they may cause unwanted cuts in the segmentation, by causing criteria to fail. There are several strategies to deal with these, see also the

discussion in Buchin et al. (2011). We propose a simple strategy of ignoring a (small) constant number of points (outliers) per segment in the segmentation. We change our testing strategy, for each criterion separately, to ignore up to a certain amount of points per segment in total, with up to a certain amount of these consecutive. This is useful not only for noise due to GPS error, but also for "relaxing" the criteria describing movement states. Allowing a constant number of outlier points does not affect the monotonicity of criteria, and thus our greedy strategies are still optimal (in the size of the resulting segmentation). A different approach would be to require that only at least a given percentage of a segment fulfills the given criteria. In this case, segmentation is not anymore monotone (extending to the front may allow extending further to the end). Thus, for this we would need other computation approaches, which we leave to future work.

3 Use Case: Migrating Geese

3.1 Geese Data

We implemented our methods (in Java) and tested these on GPS tracks from greater white-fronted geese (Anser albifrons albifrons). The given data of those geese spanned the time of their spring migration (March to June), during which they migrated from the Netherlands to Siberia. The geese where equipped with combined argos/gps microwave tags, and sampled at most every 2 h. See Kruckenberg et al. (2008), van Wijk et al. (2011), and www.blessgans.de for a more detailed description of the GPS devices, and www.blessgans.de for more information on the data collection.

The segmentation goal for this data set was to segment migration flight and stopovers (including wintering, breeding, and moult). These movement states can be described by spatio-temporal attributes as follows.

flight: little variation in heading, speed at least 20 km/h for at least 5 h
stopover: stay in 30 km radius for at least 48 h

3.2 Geese Criteria

The description of movement states directly translates to the following criteria:

flight: bounded heading angular range AND minimum speed AND min time duration
stopover: location in disk AND minimum time duration

However, because speed values were noisy, we did not use the minimum speed for flight (also not with outliers, see the discussion below). The bounded heading for flight sometimes (artificially) cut pieces of flight into smaller pieces. Due to this, we

Fig. 2 **a** Manual and **b** computed segmentation of two migrating geese. *Grey* indicates flight, *red* stopover

also omitted the minimum time duration for flight. Otherwise pieces of flight would have been unclassified.

We varied the parameter for heading angular range, i.e., the angle bounding the angular range. We found that 120° gave best results, which coincides with the domain knowledge of this parameter, see the discussion below.

Thus, finally we used the following reduced criteria (with these parameters). Note that these are all relative criteria, thus resulting in a complete segmentation.

flight: bounded heading angular range (120°)
stopover: location in disk (30 km) AND min time duration (48 h)

3.3 Results

From the resulting segmentations (Figs. 2, 3) in comparison to expert classification it becomes clear that the algorithm catches the general pattern of migration flight and stopover in space as well as in time.

Evaluation. The automatic and manual segmentation are very similar on a global scale, that is, the same stopovers are detected. However, there are differences locally. Sometimes short stops are not picked up by the algorithm, but the main stopovers are well detected. Also, the automatic segmentation cuts flight more often, because
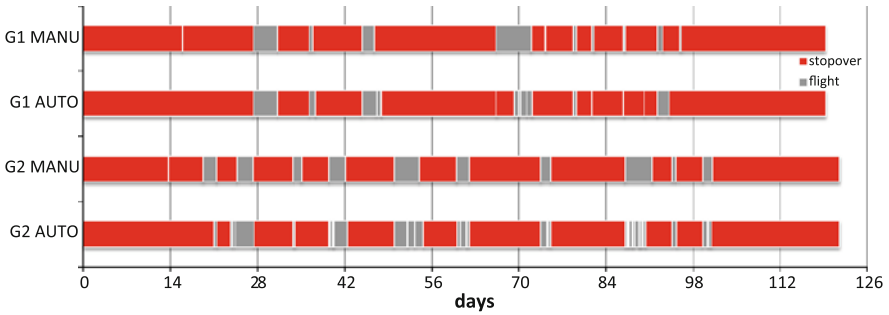
Fig. 3 Comparison of manual and computed segmentation of the tracks of two migrating geese. *Grey* indicates flight, *red* stopover. Day 1 refers to 1 March
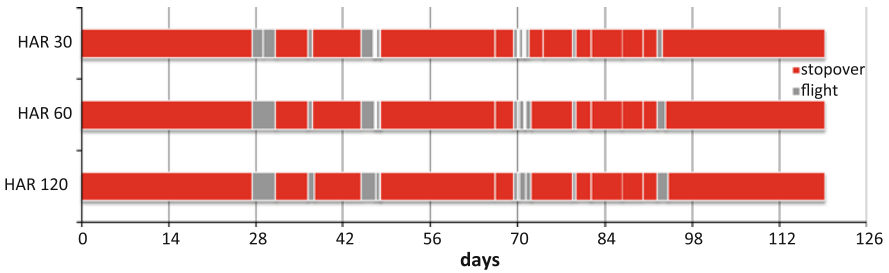

Fig. 4 Varying the parameter for heading angular range (HAR) in the computed segmentation by 120, 60 %, and 30 %. *Grey* indicates flight, *red* stopover. Day 1 refers to 1 March

of larger variation in heading. The manual segmentation has some longer stopovers, due to taking into account geographical information (e.g. lakes).

Varying parameter. We varied the angle parameter for heading angular range, as demonstrated in Fig. 4. A smaller angle results in a higher segmentation of the pieces of flight, whereas a too large angle results in wrongly classifying pieces of stopover as flight. We choose an angle of 120°, which adds only few (artificial) cutting points, and coincides with the expert knowledge of this parameter, since the geese are known to sometimes change their heading.

We compared the segmentation with and without speed and time criterion for flight, and with and without allowing outliers. However, because of a large amount of noise in the speed values, we decided not to use this criterion, even even with allowing outliers. Since the flight pieces contain small loops, the heading criterion will cut some of the flight pieces. In combination with a minimum time criterion, this results in unclassified pieces (instead of flight). Therefore, we decided also not to use a minimum time, but only heading as criterion for flight.

Coordinate projection. In our current implementation, the location in disk criterion is computed on planar coordinates. The original data, however, is given by latitude and longitude values, which we project (using a Universal Transverse Mercator) to (x, y) coordinates. However, a more precise solution would be to compute the criteria

directly based on latitude and longitude values, and instead of Euclidean distance use great-circle or rhumb-line distance. This would be useful, as migration studies use mostly distances based on rhumb lines.

4 Conclusion

The extended segmentation framework for trajectory segmentation based on movement states has proven successful in practice. This is shown by the use case of migrating geese in the previous section. To our knowledge, this is the first automatic approach to segmenting trajectories by spatio-temporal attributes describing movement states.

The success of our approach, however, depends on the possibility to describe movement states with spatio-temporal criteria, and it requires expert knowledge and manual input from the user. This raises the following two questions:

1. How to handle movement states that cannot be described as clearly with the given criteria?
2. How to segment fully automatically, i.e., without input of expert knowledge?

In the first question, the difficulty may come from movement states difficult to describe with spatio-temporal criteria, or because appropriate attributes or criteria are missing. For example, altitude would be useful for flight detection, however, in current GPS altitude data is very imprecise. However, with increasing accuracy of altitude measurements, it would be interesting to include this. Concerning the second question, such a segmentation is not always desirable, however useful for exploratory purposes. We see the strength of our approach in the possibility to manually describe movement states. Other approaches have been developed for (fully) automatic segmentation, e.g., first passage time (Fauchald and Tveraa 2003) and using k-means clustering (van Moorter et al. 2010). In this study, we concentrated on segmentation by spatio-temporal attributes describing movement states of an animal, given the knowledge of a domain expert. If expert knowledge is not available, for instance, methods from machine learning can be employed to complement our approach. In future work, we would like to explore other types of segmentation in movement ecology.

Besides the questions raised above, we plan to include (dynamic) geographic context as a criterion for segmentation. For example, in the case of migrating geese, we would like to use landcover as further criterion for stopovers (grassfields and lakes are important criteria for the geese to stop). Some important context variables are time-dependent, such as grass bloom and snow melt for the migration of geese in this study.

References

Buchin M, Driemel A, van Kreveld M, Sacristán V (2011) Segmenting trajectories: a framework and algorithms using spatiotemporal criteria. J Spatial Inf Sci (3):33–63

Fauchald P, Tveraa T (2003) Using first-passage time in the analysis of area-restricted search and habitat selection. Ecology 84(2):282–288

Kruckenberg H, Müskens G, Ebbinge B (2008) Satellite tracking of greater white-fronted geese anser albifrons during spring migration (2006)—preliminary results, Vogelwelt, pp 338–342

Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008) A movement ecology paradigm for unifying organismal movement research. Proc. National Academy of Sciences 105(49):19052–19059. doi:10.1073/pnas.0800375105

Shamoun-Baranes J, van Loon E, Purves R, Speckmann B, Weiskopf D, Camphuysen C (2012) Analysis and visualization of animal movement. Biol Lett 8(1):6–9

Turchin P (1998) Quantitative Analysis of Movement: measuring and modeling population redistribution in plants and animals. Sinauer Associates, Sunderland, MA

van Moorter B, Visscher DR, Jerde CL, Frair JL, Merrill EH (2010) Identifying movement states from location data using cluster analysis. J Wildl Manag 74(3), 588–594 . http://dx.doi.org/10.2193/2009-155

van Wijk RE, Kölzsch A, Kruckenberg H, Ebbinge BS, Müskens GJDM, Nolet BA (2011) Individually tracked geese follow peaks of temperature acceleration during spring migration. Oikos, http://dx.doi.org/10.1111/j.1600-0706.2011.20083.x