IMPROVEMENTS IN MEDIAN TRAJECTORY ANALYSIS

Volodymyr Zaslavskyi, Yurii Pushkarenko
Taras Shevchenko National University of Kyiv
Ukraine, Kyiv
zas@unicyb.kiev.ua
yurii.pushkarenko@gmail.com

Annotation: This research work is focused on the concept of finding a median trajectory among a bundle of trajectories with specific source and destination points. The main idea is to use the timing information associated with trajectories to improve existing methods for trajectory analysis.
Key words: Map-Matching, Geo-Spatial Analysis, Computational Geometry, Cluster Analysis.

1. Introduction

Trajectory analysis and related algorithms have recently attracted substantial attention, thanks to technological advances in navigation and mapping systems. Nevertheless, some fundamental concepts are still lacking a thorough study. The identification of a *middle* (representative) trajectory in a bundle of trajectories is one of them. Without conscious reasoning, a middle trajectory is a trajectory that lies in the middle of a collection of trajectories. However, this definition is far from being comprehensive.

1.1. Trajectory Definition

A formal definition of trajectory is specified in [1] as a time-stamped path taken by a moving entity, represented by a sequence of n tuples of

points and time stamps $(p_0, t_0),\ (p_1, t_1), \ldots, (p_{n-1}, t_{n-1})$. Points have

spatial and temporal components. The spatial component typically

represents a two or three dimensional space. Here, we assume that the space is two dimensional. A bundle of m distinct trajectories $T_0, \ldots, T_{m-1}$ with the same start and ending points, therefore results in an input size of $\theta(nm)$.

In an ideal situation, the time stamps of all trajectories in the bundle are exactly the same, but this is usually not the case. Generally, trajectories are collected with different or irregular sampling rates, at different times, and data can be missing as well. In between time stamps, we have no information about the actual movement path of the entity. The standard assumption can be that the entity moves with constant velocity from a time-stamped point to the next time-stamped point in a straight line. This assumption leads to an approximation of the actual data, which becomes more inaccurate when sampled at longer intervals. As a result, the path of a trajectory is considered as a polygonal curve with n edges that can self-intersect, and can have repeated vertices at the same location if the entity stands still. The number of points defining a trajectory is usually much larger than the number of trajectories in a bundle $n \gg m$.

1.2. Motivation

Various methods for trajectory analysis have been developed in different fields of science including computational geometry and data mining. Trajectory data sets can be analyzed in a variety of ways. They are usually clustered into a collection of subsets that have a high similarity regarding (a) certain property(-ies), such as location. Nevertheless, processing large amounts of trajectory data is a challenge. Trajectory data compression can be regarded as a solution to address this problem, particularly for the trajectories with huge sampled data points to improve the efficiency of computations. This compression can be carried out for individual trajectories, but in most applications, the data contains similar bundle of trajectories in terms of space and/or time, and as such, an alternative method would be to compress a trajectory bundle to represent similar trajectories with a single representative trajectory. We call this

representative trajectory a middle trajectory without loss of generality. Furthermore, using a single representative trajectory enables processing data in a simple, robust and more predictive way. As an example, a representative trajectory for an specific route can be used to predict where a vehicle will be at a certain time. Alternative applications of middle trajectories include clustering and visualizations. As an example, a middle trajectory can act as the medoid in $k - medoid$ clustering. In the context

of visualization, instead of showing a large collection of trajectories, one may identify subsets of similar ones, and replace them by a middle trajectory whose width is determined by the size of the subset.

1.3. Problem Definition

Given a set of m trajectories $T = \{T_0, \dots, T_{m-1}\}$ with the same source

and destination points s and d, we want to find a representative

trajectory T_r that minimizes a defined distance function to all trajectories

$T_i, 0 \leqslant i < m$ in the set and T_r contains points from input trajectories. In

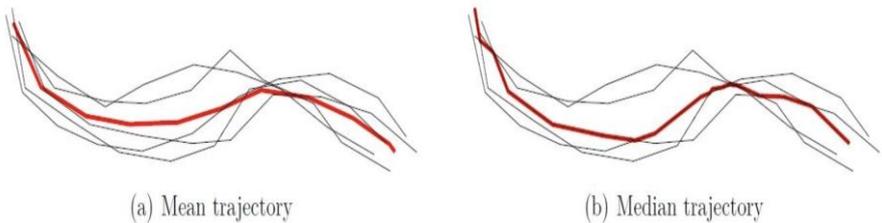case not specified otherwise, the distance function is defined as the Euclidean distance between trajectories.

(a) Mean trajectory   (b) Median trajectory

Fig. 1: Examples of (a) Mean trajectory (b) Median trajectory

A representative trajectory has to be in the middle of all trajectories in the bundle. However, we first have to define what is considered middle. There are two main types of interpretations of what a middle trajectory is for a set of input trajectories: *median* and *mean* trajectory [2]. They can produce a trajectory that is in the middle, regarding space or time, or both. There could be different interpretations for the definitions of mean and median here compared to the standard arithmetic definitions of these parameters, however the general intuition remains the same. The main difference between a mean and a median trajectory is that the latter uses only points of the input trajectories. An alternative definition for a median trajectory is to use only the edges of the trajectories present in the input set, or parts of the edges, and switch at intersections [1].
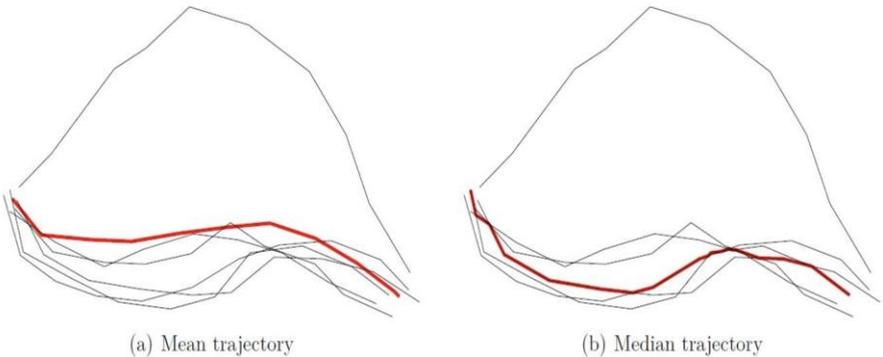

(a) Mean trajectory   (b) Median trajectory

Fig. 2: Effects of outliers on (a) Mean trajectory (b) Median trajectory

Both mean and median trajectories have their own applications. Mean trajectory which is based on the spatial component of the input trajectories looks more accurate compared to a trajectory that is restricted to one of the inputs. Nevertheless, there can be a strong reason why one prefers a median trajectory over the mean one. A typical example would be for moving objects that have to avoid obstacles. The same property should also hold for the middle trajectory. If no information about the obstacles is known in advance, then the only way to ensure that the representative trajectory crosses no obstacle is to use parts of the original input trajectories. On the other hand, if such information is known, then it may

also be possible to use this information and make the mean path go around the obstacles. Fig. 1 shows examples of mean and median trajectories. The median one follows the existing trajectories and thus avoids potential obstacles if any is defined. The mean trajectory does not necessarily follow the input trajectories and may pass through obstacles. One major disadvantage for mean trajectory is its sensitivity to outliers. Data sets of input trajectories containing substantially diverging measurements will influence a calculated mean trajectory. Fig. 2 shows an example of this phenomenon.

References.

[1] Buchin, K., Buchin, M., van Kreveld, M., Löffler, M., Silveira, R., Wenk, C., Wiratma, L., Median Trajectories, *Algorithmica* 66(3), pp. 595–614, July 2013.

[2] Vermeulen, T., Algorithms for finding a middle trajectory, Master's Thesis, TU Eindhoven, October 2013.