

SYSTEMATIZATION OF THE METHODS OF EVALUATING DISTANCES BETWEEN TRAJECTORIES BY DESCRIBING THE FEATURES OF THEM.

Tomoyuki SUZUKI, Kyoichiro KATABIRA,
Huijing ZHAO Visiting Associate Professor, Ryosuke SHBASAKI Professor
Center for Spatial Information Science, University of Tokyo
5-1-5 Kashiwanoha Kashiwa-shi, Chiba 277-8568, JAPAN
TEL: +81-4-04-7136-4307 FAX: +81-4-04-7136-4292
E-mail: {t-suzuki, katabira, chou, shiba}@csis.u-tokyo.ac.jp

KEY WORDS: Trajectory Distance, Metric, Systematization

ABSTRACT

Recently, as the progress of positioning techniques with GPS, and tracking techniques of moving objects with camera, laser range scanner etc, many trajectories have been obtained and stored. But the methods of analyzing these trajectories are not enough proposed, and not systematized. So it's difficult to say that we make the best use of many trajectories. In this paper, we straighten up and systematize the trajectory distance metrics according to the applicable cases of the metrics. By that means, it becomes easier that we can extract the similar trajectory under specific condition, or that we can classify the trajectories.

1. INTRODUCTION

Recently, as the progress of positioning techniques with GPS, and tracking techniques of moving objects with camera, laser range scanner etc, many trajectories have been obtained and stored. The methods of evaluating distances between trajectories and searching for similar trajectories are very beneficial to attain many purposes. For example, prediction of prospective human behaviors at a certain space, detection of the suspicious action, acquisition of the hints on the new service, optimization of spatial structure and so on.

But the proposed metrics are fundamental and applicable only to particular conditions. So, it's difficult to say that we can best use of heavily obtained and stored trajectories.

In addition, according to the tracked objects, each trajectory has different property. So the considered conditions when evaluating distances are different according to the contexts. In these cases, we need to apply the proposed fundamental metrics to the context in order to evaluating distances. However, we cannot instantly recognize which metric to apply, because of not accomplishing the systematization of the existing metrics. From above, when evaluating distances, there is no foundation that is able to meet the various demands because there is no existing study on the systematization of the existing metrics. In terms of this circumstance, it's difficult to say that we can best use of trajectories, too.

Therefore, we will straighten up the considered conditions when evaluating distances, and metrics proposed not only in the study on the trajectory distance but also in the study on shape analysis. And we systematize them. Besides, through case study in section 4, we will show how to actually evaluate distances under the specific condition with the systematized table and how to apply the existing trajectory distance metric to the specific condition.

2. EXISTING TRAJECTORY DISTANCE METRICS

In this section, we will detail not only the metrics proposed as the trajectory distances but also the metrics proposed in the study on shape analysis. We adopt the following notation to represent the trajectories.

$$trajA : \{(x_{1,A}, y_{1,A}, t_{1,A}), (x_{2,A}, y_{2,A}, t_{2,A}), \dots, (x_{n,A}, y_{n,A}, t_{n,A})\}$$

In [5], trajectory distance is defined as the mean of Euclidean distances between corresponding points of trajectories. Trajectory distance d_1 is given as

$$d_1(A, B) = \frac{1}{n} \sum_{i=1}^n \sqrt{(x_{i,A} - x_{i,B})^2 + (y_{i,A} - y_{i,B})^2} \quad (1)$$

In [4], trajectory distance is defined as Time-Warping distance permitting a point of one trajectory to correspond to several continuous points of the other trajectory. Time-Warping distance is also the mean of Euclidean distances between corresponding points of trajectories and given as

$$d_2(A, B) = \frac{1}{\max(m, n)} TW(m, n) \quad (2)$$

where m, n is the number of points of $trajA, trajB$, TW is $m \times n$ matrix and each cell of TW is given as following equation using dynamic programming

$$TW(i, j) = \sqrt{(x_{i,A} - x_{i,B})^2 + (y_{i,A} - y_{i,B})^2} + \min\{TW(i-1, j), TW(i, j-1), TW(i-1, j-1)\} \quad (3)$$

In [4], trajectory distance is defined as the mean of the Euclidean distances between a certain point of one trajectory and the point of the other trajectory which is the nearest to the point.

Trajectory distance d_3 is given as

$$d_3(A, B) = \max\left(\frac{1}{n} \sum_{i=1}^n d_{i,A,B}, \frac{1}{m} \sum_{i=1}^m d_{i,B,A}\right) \quad (4)$$

where $d_{i,A,B}$ is the Euclidean distance between the point i of $trajA$ and the point of $trajB$ which is the nearest to the point i .

In [2], trajectory distance doesn't include the coordinates of the points of trajectories. Trajectory distance d_4 is defined with the parameters of Hidden Markov Model.

In [6], Procrustean distance, which describes the features of the shape as the unit vector and defines the difference of the shapes as the distance on the sphere, is proposed. Procrustean distance is applicable to the trajectory distance when permitting shifting, rotating, normalizing the trajectories. Procrustean distance d_5 is given as

$$d_5(A, B) = \text{Arc cos} \left\{ \left| \sum_{j=1}^n (x'_{j,A} + iy'_{j,A})(x'_{j,B} + iy'_{j,B}) \right| \right\} \quad (5)$$

where x'_i, y'_i are the normalized coordinates

In [1], with deciding the upper and lower bounds of shifting and scaling parameters and a tolerance ε in advance, and if $trajA$ can be shifted and scaled within the specified bounds to produce a modified $trajA'$ whose distance from $trajB$ is within ε , trajectory distance d_6 which means $trajA$ is boundedly similar to $trajB$ is defined.

In [7], Levenshtein distance, which is the distance of strings describing the shapes and made by replacing the features of the shapes with the string, is proposed. With making strings according to the features of the shapes of trajectories, Levenshtein distance d_7 can be defined as the trajectory distance metric.

3. SYSTEMATIZATION OF THE EXISTING TRAJECTORY DISTANCE METRICS

In this section, we will make the table that becomes the basic frame to systematize the trajectory distance metrics. The main axes of the table are “time” and “space” that are very important factors of the trajectories.

First, about the time axis, we will divide it into four states whether to consider absolute time, relative time and time-scale. We show the names of four states and the method of dealing with trajectory at each state in Table 1.

Table 1 Four states of time axis and the direction of each state

Name of State	Method of Dealing with Trajectory
Abs Time	Permitting only the correspondences between the features tracked at the same absolute time.
Rel Time, Time-Scale Invariable	Permitting only the correspondences between the features tracked at the same time from when the first point was tracked.
Rel Time, Time-Scale Variable	After permitting “the shifting of time”, let a feature correspond to another feature, where “the shifting of time” means the correspondence between features tracked at the different time from when the first point was tracked.
OFF	Not considering absolute time and relative time completely.

Second, about the space axis, we will divide it into three states whether to consider coordinates, shifting, rotating, scaling to normalize, and other special or complex transformation. We show the names of three states and the method of dealing with trajectory at each state in Table 2.

Table 2 Three states of space axis and the direction of each state

Name of State	Method of Dealing with Trajectory
Coordinates	Using tracked coordinates to evaluate distances.
Shift, Rotate, Normalize	Permitting to shift, rotate and scale to normalize.
Special, Complex	Using special or complex transformation except above two.

By combining Table.1 and Table.2, basic frame to systematize the trajectory distance metrics is attained. In Table. 3, we show the basic frame, where the existing trajectory distance metrics given in section 2 and the arrow illustrating degree of free of each axis.

Table. 3 Systematization of trajectory distance metrics

				[Space Axis]		
				Degree of Free		
				Coordinates	Shift, Rotate, Normalize	Special, Complex
Degree of Free [Time Axis]	Abs Time	Time-scale	Invariable	Trajectory Distance d_1	Procrustean Distance d_5	
			Variable	Time-Warping Distance d_2	Trajectory Distance d_6	Trajectory Distance d_4 Levenshtein Distance d_7
	Rel Time	OFF		Trajectory Distance d_3		
	OFF					

Here, the cell whose degree of free is big means that the cell is subdivided into several cells. When you want to evaluate distances under the specific condition illustrated as the subdivided cells, you have to apply the existing metric in Table.3 in order to correspond to the specific condition. Also, the existing metrics is applicable to the cell there is no metrics in Table.3. In addition, the cell whose degree of free is big means that the computation time of the cell is generally long while computation time differs from the adopted algorithm.

4. CASE STUDY: CLASSIFICATION OF TRAJECTORIES

In this section, we will show through case study that the existing metrics are applicable to the specific conditions illustrated as the subdivided cells from the cell in Table.3. In case study, we target on the 319 trajectories that were tracked at the concourse of the railway station in Tokyo for ten minutes and the number of including points is between 40 and 80, and classify them by cluster analysis. Classification of trajectories is very beneficial to predict the human behavior and investigate traffic flow.

We will evaluate distances in consideration of relative time, variable time-scale and coordinates of trajectories. Then, we can instantly recognize Time-Warping distance as the basic metric of this condition with Table.3. This is one of the advantages due to the systematization of the metrics. Besides, we add the specific condition. The specific condition is that we extract the similar part of the trajectories and evaluate distances between them by varying the first point and the last point of compared part of trajectories.

While Time-Warping distance is the metric permitting a point of one trajectory to correspond to several continuous points of the other trajectory, and the metric inevitably corresponding the first point of a trajectory to that of the other trajectory, the last point of a trajectory to that of the other trajectory, we modify the algorithm to apply to the additional condition. If the first / last point of a trajectory corresponds to several points of the other trajectory, the modified algorithm permits only the correspondence to the point tracked at the latest / earliest time of them. This modification of algorithm is applicable to the additional condition. In Figure.1, we show the simple image of the modified algorithm of Time-Warping distance, where the broken lines between two points illustrate the correspondences of them.

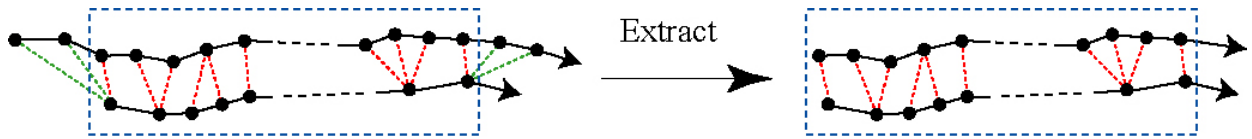


Figure. 1 Simple image of the modified algorithm of Time-Warping distance

In Figure. 2, we show all trajectories and the map of the concourse, and part of the extracted main stream by cluster analysis when we classified all trajectories into 12 clusters.

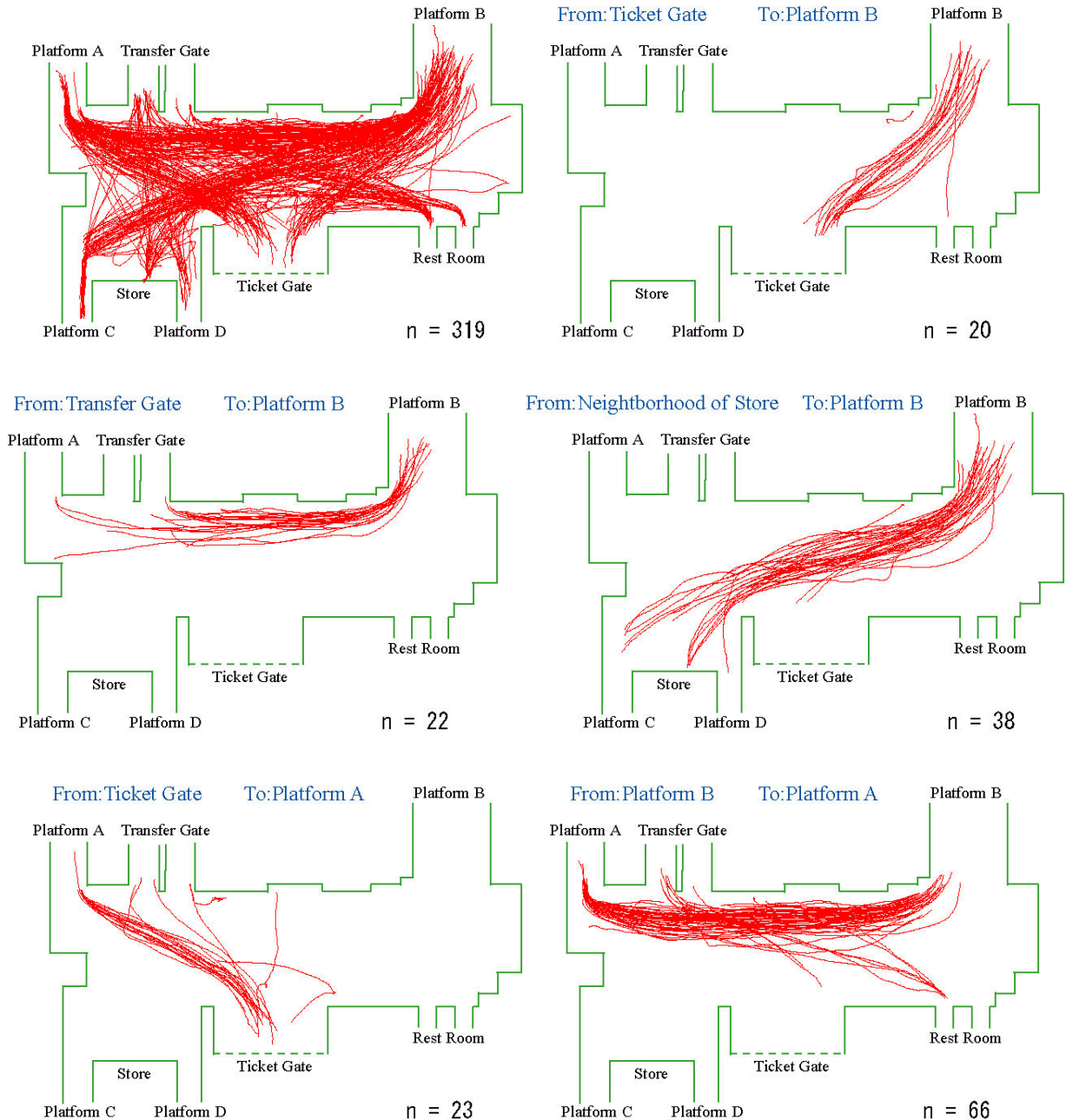


Figure.2 All trajectories and the map of the concourse, and the extracted main stream by cluster analysis, where n means the number of trajectories in each figure

Here, in Figure. 3, we show one of 12 clusters made by cluster analysis using normal algorithm of Time-Warping distance.

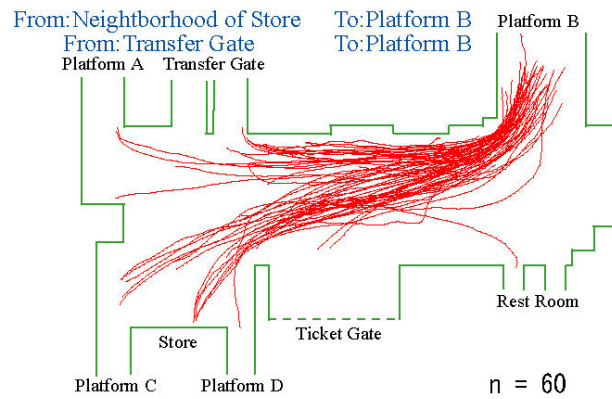


Figure. 3 Part of result of cluster analysis using normal algorithm of Time-Warping distance

The streams extracted in Figure.2 aren't clearly classified in Figure.3. This is because the extraction of the similar part of the trajectories means emphasizing the position of the first and last point of the trajectory. This result proves that the modified algorithm of Time-Warping distance is enough useful to classify the trajectories.

On the other hand, the modified algorithm has one problem. If the extracted points are few, evaluation of trajectory distances is not properly made. So there are some clusters including unsuitable trajectories. One solution of this problem is to adopt the number of the extracted points as the index like reliability of evaluation when evaluating trajectory distances.

5. CONCLUSION AND FUTURE WORKS

Case study in section 4 proves that systematization of trajectory distance metrics is beneficial and that existing metrics is applicable to the specific conditions.

For future works, accomplishment of more exhaustive systematization of trajectory distance metrics including the specific conditions and examination of optimum modification of existing metrics or proposal for new metrics to be applicable to the specific conditions, which are very important.

References

- [1] D, Q, Goldin., T, D, Millstein., and A, Kutlu., 2004. Bounded Similarity querying for time-series data. *Information and Computation*, 194 (2), pp. 203-241.
- [2] F, Porikli., 2004. Trajectory Distance Metric Using Hidden Markov Model based Representation. *European Conference on Computer Vision 2004*.
- [3] P, Laurinen., P, Siirtola., and J. Roning., 2006. Efficient Algorithm for Calculating Similarity between Trajectories Containing an Increasing Dimension. *Artificial Intelligence and Applications 2006*.
- [4] S, Oomomo., H, Chen., K, Furuse., and N, Ohbo., 2005. Adaptive Time Warping. *DBSJ Letters*, Vol. 2, No. 2, pp. 57-60.
- [5] Y, Yanagisawa., J, Akahani., and T, Satoh., 2003. Shape-Based Similarity Query For Trajectory of Mobile Objects. In *Proc. of the 4th Intl. Conf. on Mobile Data Management*, pp. 63-77.
- [6] I, L, Dryden., and K, V, Mardia., 1998. *Statistical Shape Analysis*. pp. 61-65.
- [7] S, Nakagawa., 1999. *Pattern Information Processing*. pp. 153-169.