Comparison of Similarity Measures for Trajectory Clustering in Outdoor Surveillance Scenes

Zhang Zhang, Kaiqi Huang, Tieniu Tan
National Lab of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, China, 100080
Email: {zzhang, kqhuang, tnt}@nlpr.ia.ac.cn

Abstract

This paper compares different similarity measures used for trajectory clustering in outdoor surveillance scenes. Six similarity measures are presented and the performance is evaluated by Correct Clustering Rate (CCR) and Time Cost (TC). The experimental results demonstrate that in outdoor surveillance scenes, the simpler PCA+Euclidean distance is competent for the clustering task even in case of noise, as more complex similarity measures such as DTW, LCSS are not efficient due to their high computational cost .

1. Introduction

In outdoor surveillance scenes, many applications depend on analyzing motion trajectory, such as path modeling [2] [4], activity recognition [1], and anomaly detection [3] [5]. Recent advances in object tracking also made it possible to do data mining in a large trajectories set. Cluster analysis is a common mining method. For similarity based clustering, a key issue is how to measure the similarity between two trajectories. Many methods ranging from simple Euclidean distance to complex DTW (Dynamic Time Warping) and LCSS (Longest Common Subsequence) distance have been applied to measure trajectory similarity for different applications. We list some recent work on trajectory analysis in Table 1, including the similarity measure and its application. However, it is still unclear which measure is more appropriate for the trajectory clustering task in outdoor surveillance scenes.

In this paper, on the basis of a large trajectory dataset we do clustering with six different similarity measures in case of different noise. The performance is evaluated by Correct Clustering Rate (CCR) that indicates the validity of each measure and Time Cost (TC) that indicates the computational cost. By analyzing the experimental results, we try to answer the above mentioned question.

Table1: Some recent work on trajectory analysis

Work	Measure	Application
Lou [1]	Hausdorff	Surveillance
Junejo [2]	Hausdorff	Surveillance
Porikli [3]	HMM	Surveillance
Fu [5]	Euclidean	Surveillance
Bashir [6]	PCA+Euclidean	Sign language recognition
Keogh [7]	DTW	Sign language recognition
Buzan [4]	LCSS	Surveillance
Vlachos [8]	LCSS	Sign language& Spelling recognition

2. Trajectory Distances

A trajectory is a time series of 2D or 3D coordinates. A 2D trajectory is shown below:

$$((a_1^x, a_1^y), (a_2^x, a_2^y), \ \ldots, \ (a_n^x, a_n^y))$$

Let A and B be two 2D trajectories with size N and M respectively. Six similarity measures are simply presented as follows (Details can be found in the corresponding reference):

2.1. Euclidean distance

The average distance between corresponding points on two trajectories is adopted, as shown in [5].

$$D_1(A, B) = \frac{1}{N}\sum_{n=1}^{N}[(a_n^x - b_n^x)^2 + (a_n^y - b_n^y)^2]^{\frac{1}{2}} \quad (1)$$

Note the length of both trajectories must be the same.

2.2. PCA+Euclidean distance

As proposed in [6], a trajectory is first represented as a 1-D signal by concatenating the x- and the y-projections. Then the signal is converted into the first few PCA (Principle Components Analysis) coefficients. The trajectory similarity is computed as the Euclidean distance between the PCA coefficients [6]:

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

$$D_2(A,B) = [\sum_{k=1}^{K} (a_k^c - b_k^c)^2]^{\frac{1}{2}} \quad (2)$$

where a_k^c and b_k^c is the k-th PCA coefficient of A and B respectively, and $K<<2N$, N is the size of trajectory. In our experiments, K is assigned as 5 using the method proposed in [6]. Besides this, we resample all trajectories to the median size of the dataset before using the Euclidean and PCA+ Euclidean distance.

2.3. Hausdorff distance.

In [1-2], the Hausdorff distance is adopted to express the spatial similarity between two trajectories.

$$D_3(A,B) = \max\{d(A,B), d(B,A)\} \quad (3)$$

Where $d(A,B) = \max_{a \in A} \min_{b \in B} \| a - b \| \quad (4)$

2.4. HMM-based distance.

This method is proposed by Porikli [3]. Each trajectory is fitted by an HMM (Hidden Markov Model). The HMM-based distance is [3]:

$$D_4(A,B) = | d(A;\lambda_A) + d(B;\lambda_B) - d(A;\lambda_B) - d(B;\lambda_A) | \quad (5)$$

where $d(A;\lambda_A)$ and $d(B;\lambda_B)$ indicate the likelihood of the trajectories to their own model. The terms $d(A;\lambda_B)$, $d(B;\lambda_A)$ reveal the likelihood of a trajectory generated by the other trajectory's model.

2.5. LCSS distance.

The LCSS (Longest Common Subsequence) distance finds the alignment between two sequences that maximize the length of common subsequence. Let *Head(A)* be the first *N-1* points in *A*, *Head(B)* be the first *M-1* points in *B*. Given δ and ε, the *LCSS(A, B)* can be defined as follows [8]:

$$\begin{cases} 0 & if \ A \, or \, B \ is \ empty \\ 1 + LCSS_{\delta,\varepsilon}(Head(A), Head(B)), & \\ & if \ \| a_N - b_M \| < \varepsilon \ and \, | N - M | \le \delta \\ \max(LCSS_{\delta,\varepsilon}(Head(A), B), LCSS_{\delta,\varepsilon}(A, Head(B))) & \end{cases} \quad (6)$$

In this paper, the distance is defined as:

$$D_5(\delta, \varepsilon, A, B) = 1 - \frac{LCSS_{\delta,\varepsilon}(A,B)}{\max(N,M)} \quad (7)$$

2.6. DTW distance.

It is also an alignment based distance. The basic idea behind DTW (Dynamic Time Warping) is to find out the warping path W between two trajectories that minimizes the warping cost.

The DTW distance can be represented as follows [7]:

$$D_6(A,B) = \min\{\frac{1}{K}[\sum_{k=1}^{K} w_k]^{\frac{1}{2}}\} \quad (8)$$

where the w_k is the k-th element of a warping path. In our experiment, we use the faster version in [7].

3. Dataset

In outdoor surveillance scenes, trajectory clusters are often represented by routes or paths most commonly taken by motion objects. So for a convincing clustering comparison, all kinds of paths with different position, orientation and speed should be covered in the dataset. According to the principle, we chose the scene shown in the Fig.1. The tracker presented in [9] is used to obtain motion objects' trajectories. 130 trajectories acquired from a 3-hour video are labeled manually with 13 clusters as ground truth. The average length is around 127 points. The shortest one is 38 and the longest one 447. The right part of Fig.1 illustrates the 13 trajectory clusters, where each line denotes a cluster; the arrows indicate the orientation. The blue clusters are formed by bicycle and the red are formed by pedestrian, which denote the speed variation between different clusters.

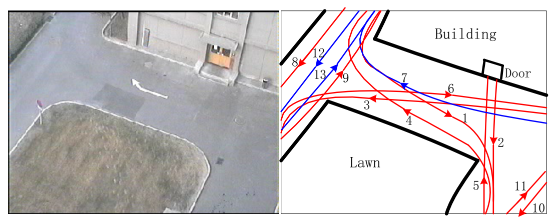

Figure1:Illustration of scene and trajectory clusters

4. Experimental Results

Let G be the ground truth set with M clusters, C be the resulting clusters set. Given a cluster $c_i \in C$, we can find out the corresponding cluster $g_m \in G$ that maximize $| c_i \cap g_j |$ Then CCR(Correct Clustering Rate) is defined to evaluate the validity of each measure.

$$CCR = \frac{1}{N} \sum_{i}^{M} p_i \quad (9)$$

where N is the total number of trajectories, p_i denote the number of the correct clustered trajectories in i-th resulting cluster. It is computed as follows.

$$p_i = \begin{cases} 0 & if \ \exists c_k \in C, | c_k \cap g_m | > | c_i \cap g_m | \\ | c_i \cap g_m | & otherwise \end{cases} \quad (10)$$

Then the following experiments are run on a PC with P4 3.0G and 512 RAM.

4.1. Clustering with original data

We first construct the similarity matrixes using each similarity measure. Then we perform the spectral clustering algorithm proposed in [10] to partition the original trajectory dataset into 13 clusters. Spectral clustering has been used in recent work [3] [5], and the advantage has been discussed in [5].

The average CCR and Time Cost (TC) during five runs are shown in Table2. The PCA+Euclidean, Euclidean, DTW and LCSS distance obtain similar CCR. By examining the clustering results, we found only the LCSS distance successfully classify the two clusters occurring in the same path and same orientation but different speed, such as the cluster9 and cluster13 shown in Fig.1. That is because with the Euclidean and PCA+Euclidean distance, the speed information is erased due to the resampling. Compared with DTW, where all points of the trajectories are matched in the warp path, only the LCSS distance find the correct alignment between the two trajectories with different speed. On the other hand, the DTW and LCSS distance take more time than the Euclidean and PCA+Euclidean distance. Furthermore, the LCSS distance requires the setting of two appropriate parameters, which is an exhausting work. In this work, we set δ as 20% of the shorter one's length of two trajectories, ε as 10 trail and error.

The results of the other two distances are not satisfactory. The Hausdorff distance only use one pairwise points to measure similarity, which cannot distinguish the trajectories in the same path but opposite direction. Furthermore, it takes the highest time cost. The HMM distance trains a statistical model for each trajectory. The result shows it very likely suffers from over-fitting due to the small training data.

The next two experiments are conducted to test the robustness of each measure in case of noise. The Hausdorff and HMM distance do not attend the next comparison due to the poor CCR and high time cost.

Table 2: Clustering Results with original trajectory

	D1	D2	D3	D4	D5	D6
CRR(%)	86.9	89.2	54.6	74.6	88.6	89.2
TC(s)	165	4	21462	672	1313	1363

4.2. Clustering with partial loss

Partial loss means the first or the last few points in a trajectory are lost due to occlusion or loss of tracking. This noise can be seen as a distortion in the time axis. According to [6-8], the Euclidean and PCA+Euclidean distance are very sensitive to such noise. In our experiment, we add this noise to two of ten trajectories in each cluster by cutting off the initial or last N points. N is selected as 10%, 20%, 30%, and 40% of the size of the trajectory. As shown in Table 3, compared with the LCSS and the DTW distance, the CCR's decrease of the Euclidean and the PCA+Euclidean distance are not significant.

Table 3: Clustering Results with partial loss

noise	D1	D2	D5	D6
10%	86.4	86.9	86.3	88.2
20%	84.6	84.6	85.4	87.7
30%	81.5	83.1	83.1	83.1
40%	79.2	81.5	79.2	80.7

4.3. Clustering with Gaussian noise

This noise was added to every trajectory of the dataset, as adopted in [8]:

$$< x_{noise}, y_{noise} > = < x, y > + randn * rValue \quad (11)$$

where *randn* is a Gaussian variable of zero mean and unit variance, and *rValue* is chosen as 5%, 10% and 15% of the range of value on X coordinates.

Table 4: Clustering Results with Gaussian noise

noise	D1	D2	D5	D6
5%	85.4	87.7	85.5	86.2
10%	83.8	86.2	83.8	85.4
15%	53.1	84.6	76.9	83.9

In Table 4, as noise increased to 10%, the CCR of all the distances have only slight decrease. As noise is increased to 15%, the Euclidean distance fails, while the PCA+Euclidean distance is still satisfactory. Furthermore the PCA+Euclidean distance is even slightly better than the DTW and LCSS distance.

4.4. Analysis of Results

Compared with the Euclidean and PCA+Euclidean distance, the LCSS and DTW distance do not clearly exhibit the robustness with noise. The finding can be understood when the application background is considered. The LCSS first proposed for trajectory analysis[8] is used to recognize sign language, where the concern is not position similarity, but shape similarity. To illustrate this, we select 13 trajectories in our dataset and 10 records in the Australian Sign Language dataset [11] to do the following analysis (shown in Fig.2).

First, we compare the shape complexity of different dataset. For each trajectory, we compute a simple index to measure the shape complexity:

$$\delta = \frac{disp}{length} \quad (12)$$

where *disp* denotes the displacement between the initi-

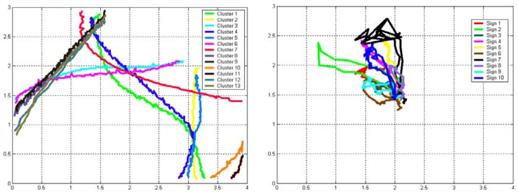

Figure 2: Trajectories in different datasets that are all normalized to a 4×3 coordinate. Left part is our dataset, the right is the ASL dataset, each trajectory denotes a cluster or a sign class.

al point and the final point, *length* denotes the length of whole trajectory. Clearly, the smaller the δ is, the more complex the shape is. The average indices are reported in the first row of Table 5.

Second, we compare the divergence of different dataset. A trajectory's position is depicted by a vector $< \alpha, \beta, \gamma >$, where α is the initial point of the trajectory, β is the middle point (in time axis) and γ is the final point. Then the covariance matrix of the 6-D (2*3) vector is computed. We sum up the diagonal elements of the matrix as the divergence of the dataset (presented in the second row of Table 5).

As shown in Table 5, the shape information in ASL dataset is much richer than that in our dataset. The spread of trajectory distribution in our dataset is much sparser than that in ASL dataset. So generally speaking, in sign dataset, the difference between two clusters mainly arises from the shape difference, while in outdoor surveillance scenes that arises from the position difference. To measure the shape similarity in sign dataset, the alignment based distances (LCSS, DTW) are more appropriate, while the Euclidean and PCA+Euclidean distance are brittle to the distortion in time axis and other noise. However, in outdoor surveillance scenes the impairment of such noise is much less, when we measure the position similarity using the Euclidean or the PCA+Euclidean distance.

5. Conclusions

In this paper, six similarity measures have been assessed for trajectory clustering in outdoor surveillance scenes. The experimental results have shown that the Hausdorff and HMM distance are not competent for the task. And the advantage of the LCSS and DTW distance to measure the shape similarity in trajectories cannot be fully utilized in outdoor surveillance scenes, while the high computational cost weakens their competitive ability with other measures. Compared with the Euclidean distance, the PCA +Euclidean distance produces better results with lower cost due to the optimally compact representation, while

Table 5: Comparison of shape complexity and divergence of different dataset

	Surveillance	ASL
average δ	0.6840	0.0380
divergence	8.1430	0.1959

it has the limitation on distinguishing speed variation. To solve this problem, explicit speed feature is needed to form new trajectory representation.

In this paper, the experiments are conducted in a labeled dataset without any outliers. For practical application, the outliers may weaken the clustering result. How to solve the problem is our future work.

6. Acknowledgment

This work is supported by research grants from the National Basic Research Program of China (No.2004 CB318100), the National Natural Science Foundation of China (No.60335010,No.60121302),the International Cooperation Program of Ministry of Science and Technology of China (No.2004DFA06900).

7. References

[1] J.Lou, Q.Liu, T.Tan, W. Hu, "Semantic Interpretation of Object Activities in a Surveillance System", in *Proc. 16th Intl. Conf. on Pattern Recognition* (ICPR'02), vol 3, 2002.

[2] I.Junejo, O.Javed, M. Shah, "Multi Feature Path Modeling for Video Surveillance", in *Proc. 17th Intl. Conf. on Pattern Recognition* (ICPR'04), vol 2, pp 716-719, 2004.

[3] F.M.Porikli, T.Haga, "Event Detection by Eigenvector Decomposition Using Object and Frame Features", in *Proc. Conf. on Computer Vision and Pattern Recognition Workshop* (CVPRW'04), vol 7, pp 114-121, 2004.

[4] D.Buzan, S.Sclaroff, G.Kollios, "Extraction and Clustering of Motion Trajectories in Video", in *Proc. 17th Intl. Conf. on Pattern Recognition* (ICPR'04), vol 2, 2004

[5] Z.Fu, W.Hu, T.Tan, "Similarity Based Vehicle Trajectory Clustering and Anomaly Detection", in *Proc. Intl. Conf. on Image Processing* (ICIP'05), vol 2, pp 602-605, 2005

[6] F.I.Bashir, A.A.Khokhar, D.Schonfeld, "Segmented trajectory based indexing and retrieval of video data", in *Proc. Intl. Conf. on Image Processing* (ICIP'03), vol 2, pp.623-626, 2003

[7] E.J.Keogh, M.J.Pazzani, "Scaling up Dynamic Time Warping for Datamining Application", in *Proc. sixth ACM SIGKDD Intl. Conf. on Knowledge discovery and data mining* (KDD'00), pp285-289, 2000

[8] M.Vlachos, G.Kollios, D.Gunopulos, "Discovering Similar Multidimensional Trajectories", in *Proc. 18th Intl. Conf. on Data Engineering* (ICDE'02), pp.673-685, 2002

[9] D.Xie, W.Hu, T.Tan, J.Peng, "A Multi-Object Tracking System for Surveillance Video Analysis", in *Intl Conf. Pattern Recognition* (ICPR'04), 2004

[10] M.Maila, J.Shi, "A Random Walks View of Spectral Segmentation," *AI and STATISTICS* (AISTATS'01), 2001

[11] http://www.cse.unsw.edu.au/~waleed/tml/data/