

CLUSTERING ALGORITHM OF VEHICLE MOTION TRAJECTORIES IN ENTRANCES AND EXITS OF FREEWAY

Zongyuan SUN¹ Dongxue LI²

¹Tongji University, 4800 Cao'an Road, Shanghai, 201804, P. R. of China

Northeast Electric Power Design Institute, 4368 Ren'min Street, Changchun 130021, P. R. of China

²Jilin Provincial Transportation Scientific Research Institute, Jinhua Street 908, Changchun, 130012, P. R. of China

KEYWORD: Freeway entrances and exits; Trajectories analysis; Hausdorff distance; Clustering algorithm;

ABSTRACT: In order to improve the research level of the intelligent traffic surveillance system, this paper presents a new hierarchical trajectory clustering algorithm in view of the features of trajectories in entrances and exits of freeway. The validity of trajectories is judged by means of the length of trajectories and vehicle speed firstly, and then the valid ones are encoded. The improved Hausdorff distance is proposed and applied to measure the similarity of temporal and spatial characteristics of trajectories. The fuzzy C-means hierarchical clustering algorithm of trajectories is further established in which trajectories are first clustered into different paths according to the position and geometric shape of trajectories, and then trajectories belonging to the same path are further clustered according to vehicle speed. Finally, the results of experiments in true scene have confirmed the validity and applicability of the algorithm.

INTRODUCTION

As the accident-prone areas, how to improve the security of entrances and exits has been an important research subject for freeway operation management. With the rapid development of computer hardware and software and visual technology, it is now possible to use the traffic control and surveillance system to improve the level of road safety and one of the most important efforts is to analyze and identify behaviors of vehicles based on their trajectories (Valera & Velastin, 2005, Morris & Trivedi, 2008). Through the analysis of trajectories in entrances and exits of freeway, we can not only detect the illegal vehicles, such as changing lanes frequently, speeding and retrograde violations, but also can predict the tendency of motion, assess their safety state in real time and further judge whether they might lead to accidents, if so, they might be able to warn drivers before danger occur and thus reduce the number of accidents. Trajectory clustering of vehicle is the basis for the subsequent behavior analysis and recognition. The satisfactory clustering results can not only provide the training samples for the construction of recognition models, but also affect the accuracy and reliability of trajectory analysis directly.

As the time-varying data series, the trajectory has the characteristics of dynamic, unequal length, high information content and so on. Trajectory clustering analysis aims at using unsupervised learning method to explore hidden patterns from a large amount of sequence data. At present, the trajectory clustering methods include: K-means clustering (Pan et al., 2009), hierarchical clustering (Hao et al., 2009) and neural network based clustering (Owens, 2000, Hu et al., 2004) and HMM (Hidden Markov Model clustering (Porikli, 2004, Saunier & Sayed, 2005) and so on. K-means clustering algorithm is simple, but the clustering results are generally easy to fall into the local optimal solution. Although the hierarchical clustering method is also relatively simple, the traditional one is not adaptive to the incomplete trajectory, and it needs to be improved. The clustering effect of NN and HMM is relatively good, but the calculation is complex and it is difficult to meet the real-time requirements. What is more, how to determine the number of HMM is still a problem to be studied. In

addition, none of the above methods proposes effective trajectory preprocessing method to overcome the defects of the tracking algorithm.

Aiming at the problems already existed in trajectory clustering algorithms; this paper proposes a hierarchical clustering algorithm based on the improved Hausdorff distance. Firstly, the validity of trajectories is judged by means of the length of trajectories and vehicle speed and the valid ones are encoded, then the improved Hausdorff distance is proposed and applied to measure the similarity between trajectories. Finally, the fuzzy C-means algorithm is used to realize the hierarchical clustering of effective trajectories. Experimental results on real scenes show that the trajectory clustering algorithm proposed in this paper can get satisfactory clustering effect and has good applicability for vehicle trajectories in entrances and exits of freeway.

Trajectory preprocessing

Trajectory validity judgment

Since the process of vehicle tracking may be accompanied by occlusion, error, failure, and so on, the obtained trajectories are often full of noise, different from the starting point and truncated to a number of short fragments. If the original complex trajectory is not effectively preprocessed, all the above factors will seriously affect the subsequent trajectory classification, modeling and identification accuracy. Integrated various factors, this paper mainly applies the trajectory length and vehicle speed to determine the validity of trajectories. By preprocessing the trajectory, it can not only simplify the calculation complexity and reduce the amount of computation, but also improve the accuracy of the subsequent clustering and recognition.

The trajectory length

To calculate the number of coordinates of each complete trajectory N and set a threshold T_1 , if $N \geq T_1$, then keep the track and record the corresponding trajectory by the subscript S . Through this process, we can filter out a part of short trajectories which are influenced by the noise in the tracking process and produced due to tracking failure, so it will reduce the amount of calculation and the error of further analysis and control the overall situation of the trajectories.

Vehicle speed

Because of the short time interval between frames of the video sequence, and the points in the trajectory sampled at equal time interval, the vehicle speed can be characterized by the change of the position of adjacent frames. Assuming that the two-dimensional coordinates of cancrroids in the frame i is (x_i, y_i) , $i+1$ is (x_{i+1}, y_{i+1}) , Then the speed vector of the vehicle o at the time i is:

$$\vec{V} = (d_{xi}, d_{yi}) \quad (1)$$

$$d_{xi} = x_{i+1} - x_i \quad (2)$$

$$d_{yi} = y_{i+1} - y_i \quad (3)$$

When the length of the trajectory is valid, the position change of all adjacent frames in each trajectory $\vec{V} = (d_{xi}, d_{yi})$ and the statistical mean value of the characteristic values in the sliding window with length L \bar{X}_L , \bar{Y}_L are calculated in turn. Set the two thresholds T_x , T_y and determine whether $\bar{X}_L > T_x$, $\bar{Y}_L > T_y$ and $\bar{X}_L / \bar{Y}_L > T_q$ are satisfied at the same time, if the two conditions are satisfied at the same time, it is considered that the trajectory is valid; Otherwise the trajectory is invalid and removed from the samples. Due to the serious disturbance of noise and the influence of the tracking error of the target, a large number of trajectories obviously indicate that the speed is too fast or

change violently which are not in accordance with the laws of the movement of the vehicle. After effective judgment of the speed, we can further filter out parts of invalid trajectories, and finally get reliable samples.

Trajectory coding

The purpose of coding is to create a concise and compact expression for trajectory according to the demand of the clustering method, and finally get the sample library for clustering. From the information derived from the region-based vehicle tracking, we can acquire the training trajectories that are used to cluster. In this paper, we use the similar trajectory coding schemes which are used in (Johnson & Hogg, 1996, Stauffer et al., 2000, Sumpter & Bulpitt, 2000). The original trajectories are sampled at a fixed rate (once every Δt frame). Given a vehicle O , we represent the world coordinates of the cancrroids at the i th sampling as (x_i, y_i) . After sampling n times, we obtain a point sequence T^n that is composed of n pairs of world coordinates:

$$T^n = \{(x_1, y_1), (x_2, y_2), \dots, (x_{n-1}, y_{n-1}), (x_n, y_n)\} \quad (4)$$

In addition, we use the previously mentioned $\mathbf{v} = (d_{xi}, d_{yi})$ to represent the speed of the moving vehicle at time i . According to the purpose of this study, the movement of vehicle O is represented by the set F^n , which is composed of n flow vectors $F^n = \{f_1, f_2, \dots, f_i, \dots, f_{n-1}, f_n\}$, where $f_i = (x_i, y_i, d_{xi}, d_{yi})$.

It is worth noting that, the acceleration of vehicle can be obtained through the trajectory analysis, but taking into account the noise and the efficiency of the clustering algorithm, here we do not consider it. At the same time, in order to balance the relative contribution between the position and the speed elements in the calculation of the flow vector similarity, they all need to be normalized. After transformation, each component of the flow vector f_i will be in the range of $[0, 1]$, that is $(x_i, y_i, d_{xi}, d_{yi}) \in [0, 1]$. By tracking vehicles, a set of trajectories $S = \{F_1, F_2, \dots, F_l, \dots, F_M\}$ is acquired, where M denotes the number of sampled trajectories and F_l the l th trajectory.

Trajectory Similarity Measures

In order to make a comparison between the trajectories and obtain the meaningful clustering results, it is necessary to determine the appropriate distance metric to measure the similarity between samples before clustering. Trajectory similarity measure is mainly used to measure the degree of similarity between the two trajectories, which is the basis of trajectory clustering. The current commonly used trajectory similarity measures include: Euclidean distance (Euclidean), Hausdorff distance (Hausdorff), the longest common sequence (LCSS), dynamic time warping (DTW) and so on (Xu, 2005). Euclidean distance can only be used to deal with the trajectory of equal length and is prone to the noise and sequential variation; LCSS and DTW can effectively measure the similarity of the trajectory, but they are more suitable for the trajectory of the flexible object due to the shape change (Zhang et al., 2006). The Hausdorff distance can't consider trajectory direction, but in view of the research object of this paper is the vehicle trajectories of freeway entrances and exits which are different in length, but in the same direction. Therefore, based on the comprehensive analysis of the advantages and disadvantages of these indicators, we propose and apply the improved Hausdorff distance to measure the similarity between the trajectories. The traditional Hausdorff distance is a measure of the similarity between a pair of points set which constitute the trajectories (Junejo, 2004, Khalid & Naftel, 2005, Atev, 2006). When there is a great difference between the lengths of the two trajectories, the traditional Hausdorff distance will increase with the length of the trajectory, and cannot truly reflect the degree of similarity between trajectories. Therefore, given the two trajectory F_i, F_j , this paper proposes and symmetrically defines the improved Hausdorff distance as follows:

$$D_H(F_i, F_j) = \min(D_h(F_i, F_j), D_h(F_j, F_i)) \quad (5)$$

$$D_h(F_i, F_j) = \max_{f_{j,l}} \left(\min_{f_{i,k}} d_E(f_{i,k}, f_{j,l}) \right), (\forall k, l) \quad (6)$$

$$D_h(F_j, F_i) = \max_{f_{j,l}} \left(\min_{f_{i,k}} d_E(f_{j,l}, f_{i,k}) \right), (\forall l, k) \quad (7)$$

Where $d_E(f_{i,k}, f_{j,l})$ denotes the Euclidean distance between flow vectors $f_{i,k}, f_{j,l}$ in feature space.

HIERARCHICAL CLUSTERING OF TRAJECTORIES BASED ON FUZZY C-MEAN ALGORITHM

The aim of trajectory clustering is to classify the similar trajectories into the same class. Compared with the ordinary c-means clustering algorithm, because the fuzzy c-means clustering uses the fuzzy theory to guide the work, it is more close to the reality and often be done with less workload and better results in practical applications (Hu et al., 2006).

Due to the complexity and variety of trajectory samples, the calculation of distance between the large numbers of trajectories is relatively heavy. In order to improve the efficiency of clustering, this paper proposes a hierarchical clustering method based on fuzzy C-means algorithm, which uses the different space and time characteristics of trajectories. As trajectories belonging to different paths must be assigned to different clusters, firstly trajectories are clustered into different paths according to the geometric shape of trajectories, and then trajectories belonging to the same path are further clustered according to vehicle speed.

Path clustering

The path is the space curve which is characterized by the position of each successive point of the trajectory. Although the path contains only the coordinates of the position, it fully reflects the most important geometrical and spatial characteristics of the trajectory. Therefore, the corresponding paths for trajectories are clustered to obtain the preliminary results firstly. In the path clustering, only coordinates of the points on each trajectory are kept for clustering and all corresponding samples are input to the clustering algorithm. The “centre” trajectories are used to present each cluster and the cluster results of original trajectories are obtained finally.

In this paper, the path set corresponding to the original valid trajectories is represented as $S' = \{F'_1, F'_2, \dots, F'_l, \dots, F'_M\}$, $F'_l = \{f'_1, f'_2, \dots, f'_i, \dots, f'_n\}$. The purpose of path clustering is to cluster S' into K different subset $q = \{q_1, q_2, \dots, q_j, \dots, q_K\}$, where θ_j denotes the j th subclass and K denotes the number of cluster which can be determined in advance or according to certain criteria. The above task is performed in five steps:

Step 1: Randomize the initial K clustering centers

Step 2: Input all samples $S' = \{F'_1, F'_2, \dots, F'_l, \dots, F'_M\}$ and calculate the improved Hausdorff distance from each sample F'_l to all the K clustering centers

$$D_H(F'_l, q_j) = \min(D_h(F'_l, q_j), D_h(q_j, F'_l)) \quad l=1, 2 \dots M, j=1, 2 \dots K. \quad (8)$$

Step 3: Compute the memberships of each sample to all the K clustering centers

$$R_{lj}(t) = \frac{1/D_H(F'_l, q_j)}{\sum_{j=1}^K (1/D_H(F'_l, q_j))}. \quad (9)$$

Step 4: Adjust each cluster center according to the formula

$$q_j(t+1) = q_j(t) + \frac{\sum_{l=1}^M R_{lj}(t) \times (D_H(F'_l, q_j))}{\sum_{l=1}^M R_{lj}(t)} \quad (10)$$

Step 5: Verify the stability condition of the cluster

$$\max_{\substack{1 \leq j \leq K \\ 1 \leq i \leq n}} \{ |q_{ji}(t+1) - q_{ji}(t)| \} < e \quad (11)$$

If the stability condition is satisfied or the predefined number of iterations is achieved, then the learning process terminates; otherwise, go to Step 2 for another loop of learning. Corresponding to the clustering result of the intermediate trajectories, the set of original trajectories $S = \{F_1, F_2, \dots, F_l, \dots, F_M\}$ is clustered into K subsets:

$$S = \{ \{F_{1,1}, \dots, F_{1,M_1}\}, \dots, \{F_{i,1}, \dots, F_{i,M_i}\}, \dots, \{F_{K,1}, \dots, F_{K,M_K}\} \} \quad (12)$$

Where: M_i denotes the number of original trajectories in the i th path subclass.

Speed clustering

Although trajectory pattern and geometric path are similar, there are many differences between them. Compared with the latter, the former also contains the dynamic time information of trajectory, and its external performance is speed difference. For example, even if the vehicle is moving along the same path, they may also be attributed to different trajectory patterns, as a vehicle may speed up and down or run normally. The purpose of trajectory clustering is to identify and predict the motion of vehicles, and using the time information to further classify the trajectories will no doubt improve the accuracy of the above work. Therefore, it is necessary to use temporal information to further cluster the path subset $S_i = \{F_{i,1}, \dots, F_{i,M_i}\}$, $i=1, 2 \dots K$.

Since each flow vector of the original trajectory already contains the speed element, the similarity between the trajectories belonging to the same path subclass is still measured by calculating the improved Hausdorff distance, and the fuzzy C-means algorithm is used to cluster each subclass again. It is worth noting that since the scale of each subset S_i is much less than that of the whole original trajectories; the clustering procedure based on the speed information is more accurate and efficient than that based on the original samples directly.

After speed clustering, all intermediate trajectories are clustered into C clusters. According to the correspondence between the original trajectories and the intermediate trajectories, the original trajectories $S_i = \{F_{i,1}, \dots, F_{i,M_i}\}$ are clustered into C sets of trajectories:

$$S = \{ \{F_{1,1}, \dots, F_{1,M'_1}\}, \dots, \{F_{j,1}, \dots, F_{j,M'_j}\}, \dots, \{F_{C,1}, \dots, F_{C,M'_C}\} \} \quad (13)$$

Where M'_j denotes the number of original trajectories in the j th speed subclass.

EXPERIMENTAL RESULTS

In order to verify the validity and applicability of the proposed clustering algorithm, we tested it in an entrance and exit of freeway scene which both are parallel type and have two main lanes (as shown in Fig. 1 and 2 respectively). The images are of size 520×240 pixels and sampled at 25 frames per second. During the 15 minutes, 447 cars passed through the entrance and 352 cars passed through the exit. When the trajectories are preprocessed, the threshold value of each parameter is set as follows: $T_1=15$, $T_x=0.4$, $T_y=1.5$, $T_\theta=1.732$. During the testing time interval, 529 and 405 valid motion trajectories were produced using our tracking algorithm, which are shown as Fig. 3 and 4 respectively.



Figure 1. The freeway entrance.



Figure 2. The freeway exit.



Figure 3. Trajectories from the entrance scene.

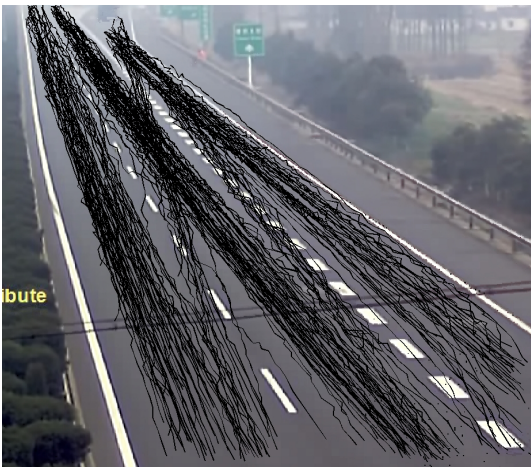


Figure 4. Trajectories from the exit scene

Fig.5 and 6 show the clustering results, respectively, in the entrance and exit using our algorithm. For simplicity, the results are represented by the geometric mean of each cluster with different color. As shown in the figures, the clustering results not only have strong regularity, but also are consistent with the actual motion patterns of vehicles at the entrance and exit of the freeway and reflect the spatial and temporal distribution of the corresponding sample trajectories in the entrance and exit area. It is seen that the method proposed in this paper has good applicability for the clustering of vehicle trajectories in the freeway entrances and exits.

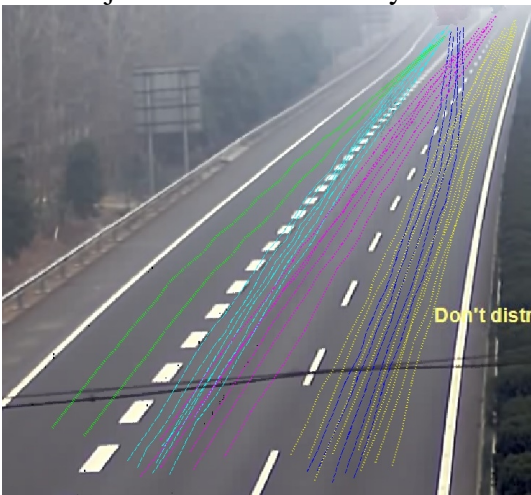


Figure 5. The clustered result in the freeway entrance

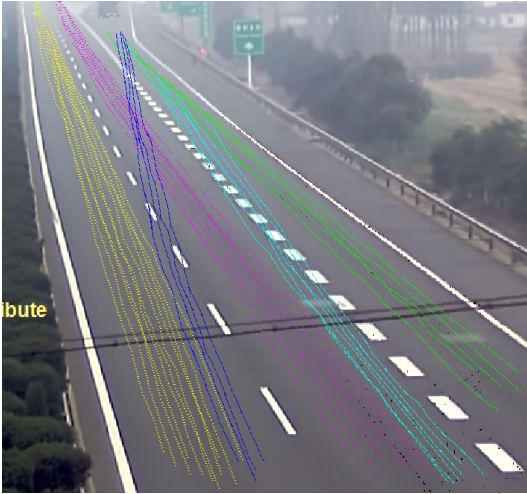


Figure 6. The clustered result in the freeway exit

CONCLUSIONS

According to the features of trajectories in the entrances and exits of freeway, we present a new hierarchical trajectory clustering algorithm. First, we use the length of trajectories and vehicle speed information to justify the validity of trajectories and obtain the trajectory samples. Second, the improved Hausdorff distances are proposed and applied to measure the similarity of temporal and spatial characteristics for trajectories. Finally, the fuzzy C-means trajectory hierarchical clustering algorithm based on the improved Hausdorff distances is established. Through experiments in true scene, the results have confirmed the validity and applicability of the proposed algorithm for the clustering of vehicle trajectories in the entrances and exits of the freeway.

REFERENCE

- [1] Atev S, Masoud O, and Papanikolopoulos N. "Learning traffic patterns at intersections by spectral clustering of motion trajectories," in Proc. IEEE Conf. Intell. Robots Syst., Beijing, China, Oct. 2006, pp.4851–4856.
- [2] Jiuyue HAO, Chao LI, Lei GAO, et al. Moving object trajectory clustering method in intelligent surveillance video [J]. Journal of Beijing University of Aeronautics and Astronautics, 2009, 35(9): 1083—1087(in Chinese).
- [3] Johnson N, Hogg D. Learning the distribution of object trajectories for event recognition. Image and Vision Computing, 1996, 14(8): 609~ 615.
- [4] Junejo I N , Javed O , Shah M .Multi feature path modeling for video surveillance[C] Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04), 2004:716-719.
- [5] Khalid S , Naftel A. Evaluation of matching metrics for trajectory-based indexing and retrieval of video clips[C] Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05), 2005:242-249.
- [6] Morris B T, Trivedi M M. A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2008, 18(8):1114-1127.
- [7] OWENS J, HUNTER A. Application of the self-organizing map to trajectory classification[C]: In Proceedings of the 3rd IEEE International Workshop on Visual Surveillance, 2000: 77-83.
- [8] Porikli F. Trajectory distance metric using hidden markov model based representation[C]// ECCV PETS Workshop. 2004.
- [9] Qiming PAN, Yongmei CHENG, Tao YANG, et al. Trajectory recognition classification and recognition of moving objects[J]. Fire Control&Command Control, 2009, 34(11): 79-83(in Chinese).

- [10] Rui Xu. Survey of clustering algorithms[J]. Neural Networks IEEE Transactions on, 2005, 16(3):645 - 678.
- [11] Saunier N. and T. Sayed. Automated Road Safety Analysis Using Video Data. In Transportation Research Board Annual Meeting Compendium of Papers, Washington, D.C., January 2007.
- [12] Stauffer C, Eric W, Crimson L. Learning patterns of activity using real time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747~ 757.
- [13] Sumpter N, Bulpitt A. Learning spatio-temporal patterns for predicting object behavior. Image and Vision Computing, 2000, 18(9): 697~ 704.
- [14] Valera M, Velastin S A. Intelligent distributed surveillance system: A review[c].Proceedings of the Vision, Image and Signal Processing, 2005:192-204.
- [15] Weiming Hu, Dan Xie, Tieniu. Tan, and Maybank S. “Learning Patterns of Activity Using Fuzzy Self-Organizing Neural Network,” IEEE Trans. Systems, Man, and Cybernetics—Part B, vol. 34, no. 3,pp. 1618-1626, June 2004.
- [16] Weiming Hu, Xuejuan Xiao, Zhouyu Fu, et al. A system for learning statistical motion patterns [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, 28(9):1450-1464.
- [17] Zhang Zhang, Kaiqi Huang, Tieniu Tan. Comparison of Similarity Measures for Trajectory Clustering in Outdoor Surveillance Scenes. International Conference on Pattern Recognition IEEE Computer Society, 2006:1135-1138.