MVA2013 IAPR International Conference on Machine Vision Applications, May 20-23, 2013, Kyoto, JAPAN

Big Trajectory Data Analysis for Clustering and Anomaly Detection

Hirokatsu Kataoka, Yoshimitsu Aoki
Keio University
Yokohama, Kanagawa, Japan
kataoka@aoki-medialab.org

Kenji Iwata, Yutaka Satoh, Ikushi Yoda, Masaki Onishi
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Ibaraki, Japan

Abstract

We've been developing a sensor that can acquire positional data. Recently, a position-based big data creation is easy task and trajectory analysis is the highest priority for "position-based service". Traffic congestion, marketing mining, and pattern analysis are the one of the examples in trajectory analysis field. In this paper, we propose the trajectory analysis approach for clustering and anomaly detection by using big trajectory data. To execute clustering, we understand an environment in front of the camera and set a cluster route from trajectory map. The experiment shows that the proposed method understands environment and performs clustering. Moreover, the approach classifies anomalies from big data.

1 Introduction

The researchers have been studying detection and tracking approach with computer vision and pattern recognition techniques. We are possible to acquire positional data by using tracking methods. The results of sensor development, we will capture effective information from big trajectory data of pedestrian and vehicle for location-based services [1]. Now we can apply "big trajectory data" for market, traffic, and surveillance, for example, the pattern analysis is necessary to understand marketing activities.

Recently, the trajectory information have been large-scale database, it can be called "big data". We believe big data analysis is needed to solve social problems. We describe the related works as below:

The one of important trajectory analysis is the clustering problem. Wang et al. proposed clustering method by using topic model [2]. The topic model was improved to model clusters in multiple views [3]. Wang et al. applied over 40,000 trajectories to model clusters in a parking lot. The method forms semantic regions from input trajectories. The clusters are divided to connect the regions. From the clusters, the method succeeded to detect abnormal trajectories. Lee et al. proposed partition and group algorithm [4]. They divided frequent regions and turning points from 570 trajectories. The R-tree is the algorithm to understand regions and clusters. Another example of abnormal trajectory detection is [5]. Piciarelli et al. divided trajectories into clusters. The non-tagging trajectories are classified as abnormal trajectories.

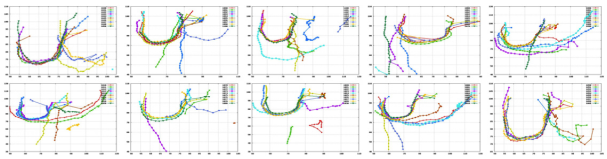

Figure 1. The samples of input trajectory data.

The related works applied only 100 - 10,000 (10×10^2 - 10×10^4) order trajectories. There is a big data that have over million order trajectories. To analyze detailed features or ambiguous trajectory clusters, we must input large-scale trajectory data. Large-scale trajectory data allows us to examine minor differences and cover the all patterns.

In this paper, we describe the approach to perform clustering and anomaly detection by using 100,000 (10×10^5) order trajectories. The samples of input trajectory are shown in Figure 1 The paper is consist of two steps, creating main trajectory map which shows regions in front of camera, and clustering with main trajectory map. As the result of clustering, we extract abnormal trajectories. The experiment shows clustering and anomaly detection from 300,000 (30×10^5) trajectories.

The rest of the paper is organized as follows. In Section 2, we describe the clustering and anomaly detection method. In Section 3, we show experimental results on large-scale trajectory data (30×10^5). Finally, Section 4 concludes the paper.

2 Trajectory Analysis

The trajectories are space-time position data tracked from stereo camera [7]. Figure 2 shows the proposed flow. The trajectory analysis is consist of two steps. The first step is to create trajectory map in front of the camera.

2.1 Main Trajectory Map

The creation of main trajectory map is given in Figure 3. To understand environment in front of camera, we create main trajectory map that indicates frequency and direction at each block. The block size is 5 pixels square. The main trajectory map is accumulated frequency and transition direction in block

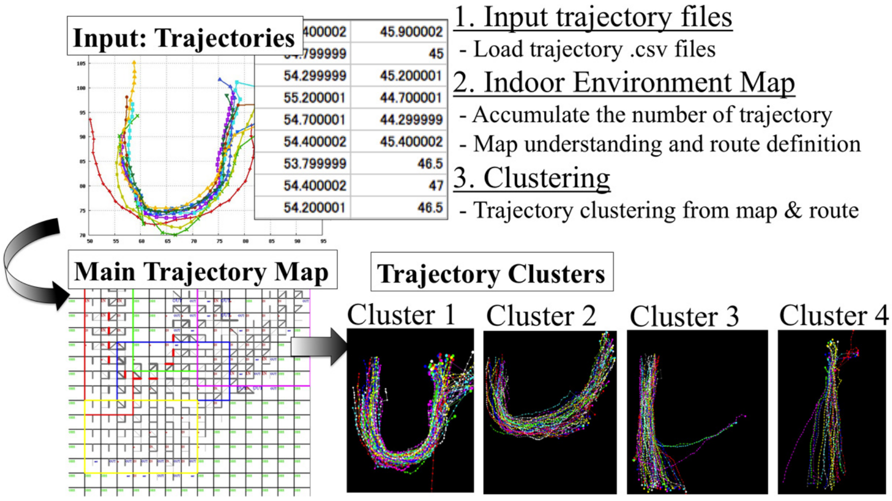

Figure 2. The proposed flow: Trajectory is tracked in front of escalator with Onishi et al. [7] tracking method. The method is consist of main trajectory map creation and trajectory clustering.

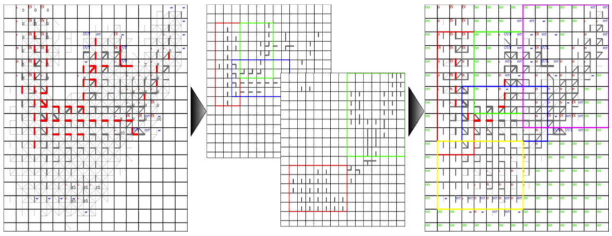

Figure 3. Main trajectory map: (left) Trajectories are accumulated in main trajectory map. Frequency and direction is the indicator. (center) The blocks are binarized to execute labeling algorithm. Two thresholding values are prepared in this experiment. (right) Regions are collected from main trajectory map with labeling algorithm.

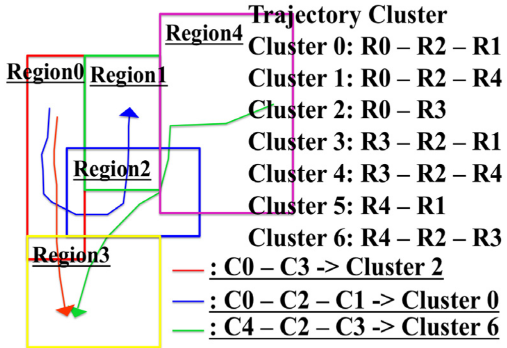

Figure 4. The cluster routes for clustering.

of map. We set 10 dimensions for block representation. The dimensions include human's entrance and exit (IN/OUT) and eight directions. And more, we set direction visualization five indication depending on frequency.

Each block is binarized to collect regions where have the same direction with the value of main trajectory map. The binarized values are collected applying labeling algorithm that can capture a region from pixels. Labeling algorithm calculates the number of pixels and the position of regions. In this approach, we set the labeling algorithm as eight directions. And the proposed approach changes thresholding value and binarizes in main trajectory map. From the labeling algorithm, we capture number of pixels, center of gravity, and maximum rectangle.

2.2 Trajectory Clustering and Anomaly Detection

The cluster routes are set from main trajectory map. In each region (region 0 - 4 in Figure 4) has the information of IN/OUT and relationship between regions (ambulant route or not). A route of cluster can be established to divide trajectories into 7 clusters in Figure 4.

A trajectory of real scene has an ambiguous route on a floor. We deal with the problem simple matching approach, for example, R0 - R2 - "R4" - R1 into Cluster 0 in Figure 4. And we adjust entrance and exit in the environment.

Anomalies are "not" normal trajectories. We extract abnormal trajectories from large-scale trajectory data with cluster routes in the main trajectory map. The conditions of anomalies are below:

- A trajectory comes in a region where cannot enter.
- A trajectory strides over between regions that cannot cross.
- An extremely long / short trajectory.

The results are given in experiment of trajectory analysis.

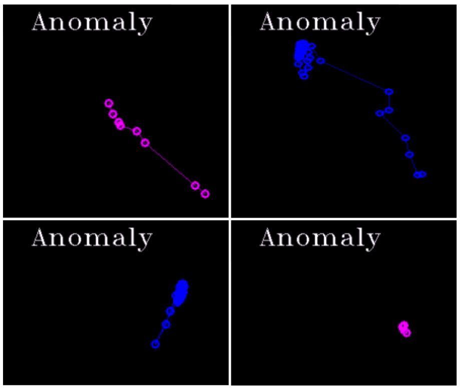

Figure 6. Anomaly detection: short/long trajectories and abnormal trajectories are extracted as anomalies.

3 Experiment

In this experiment, we validate the setting of threshold value, clustering method, and anomaly detection approach. Pedestrians are tracked with Onishi et al. [7] approach in compound building. The input data is space-time (x, y) coordinates. We apply over 300,000 trajectories at most to suppose big-data analysis. In experiment one, we apply 100 - 100,000 trajectories to create main trajectory map. In experiment two, we implement clustering and anomaly detection by using main trajectory map and route information.

Experiment 1 Main Trajectory Map Creation: We examine the relationship between number of sample or thresholding value and main trajectory map. We change the value of sample 100, 1,000, 10,000, 100,000, and the thresholding value N/1,000, N/100, N/20 (N is the number of sample). The results are shown in Figure 5. We cannot eliminate noise if the thresholding value is low, however, the high thresholding value deletes the direction and frequency in map. From the main trajectory map, we believe a large amount of trajectory data allow us to perform clustering easily. The region clustering is given in Figure 3. We got the main trajectory map from 10,000 trajectories. The five regions are extracted to calculate routes in the main trajectory map. Moreover, the obstacle regions are recognized from main trajectory map. Wall, escalator, and door are understood as obstacle area. Low frequency area is captured as obstacle area.

Experiment 2 Clustering and Anomaly Detection: In this experiment, we divide trajectory data into 7 clusters with the routes of main trajectory map. Figure 7 shows the results of clustering. The most frequency sample is cluster one, approximately 40% of all trajectories. In this situation, we captured trajectories in front of escalator. A lot of people moved with escalator. Other situations are moving the same floor, or up with escalator and move to door. Cluster 5 and 7 in Figure 7 have variance because of region 3 and 4 are the large area.

Cluster 1 - 4 are over 90% accuracy, however cluster 5 - 7 are 70 -80% accuracy. The regions between 1 and 4 has entrance and exit, therefore, it is difficult to understand the first/last route of trajectory. In anomaly detection, we detect abnormal trajectories with the length of trajectory, the main trajectory map, and the routes. We confirmed tracking miss because of over detection comes from changing light source or similar shapes to human head. The tracking method [7] reports tracking accuracy is 98.6%, however, a lot of over detection and tracking. Other anomalies are extremely long trajectories crossing many regions on main trajectory map, and IN (entrance) and OUT (exit) are the same.

4 Conclusion and Future Works

We analyzed big trajectory data for clustering and anomaly detection. 300,000 trajectories and main trajectory map are applied to perform clustering and detect anomalies.

In the future, we'd like to model anomalies. To the best of our knowledge, there're no related works about anomaly modeling because of the number of data. Currently, we hold over 50,000,000 trajectory samples in a building. We believe that the anomaly mining is realized with a big trajectory data.

References

[1] Zheng. Yu, Zhou. Xiaofang : "Computing with Spatial Tra-jectories", Springer, 2011.

[2] Wang, X., Ma, K.T., Ng, G., Grimson, E.: Trajectory analysis and semantic region modeling using a nonparametric Bayesian model. In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2008.

[3] Wang, X., Tieu, K., Grimson, E.: Correspondence-free activity analysis and scene modeling in multiple camera views. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), 2009.

[4] Jeung, H., Yiu, M.L., Zhou, X., Jensen, C.S., " Path prediction and predictive range querying in road network databases, The VLDB Journal 19(4), 585-602, 2010.

[5] Claudio Piciarelli, Christian Micheloni, Gian Luca Foresti : "Trajectory-Based Anomalous Event Detection", IEEE Transactions on Circuits and Systems for Video Technology,Vol. 18, No. 11, pp. 1544 - 1554, 2008

[6] J. G. Lee, J. Han, "Trajectory Clustering: A Partition-and Group Framework", ACM SIGMOD international confer-ence on Management of data, pp.593-604. 2007.

[7] M. Onishi, I Yoda, "Dynamic Trajectory Extraction from Stereo Vision Using Fuzzy Clustering", IEEJ Trans. Elec-tronics, Information and System Society, 128(9), pp.1438-1446, 2008.

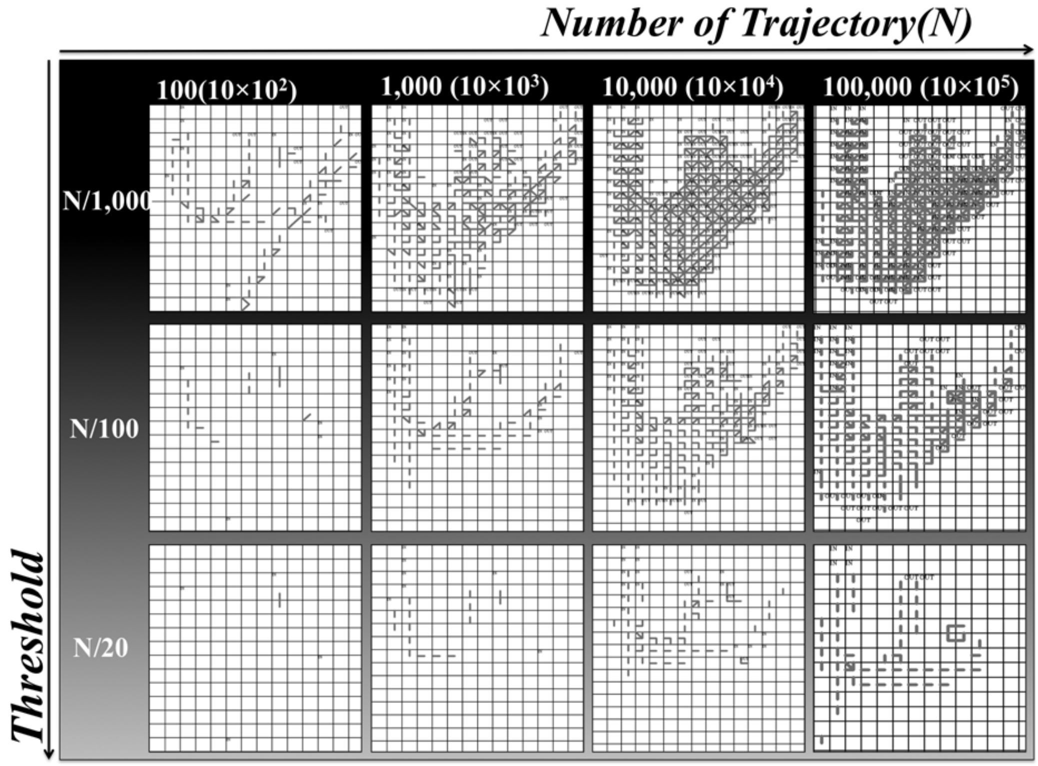

Figure 5. The relationship between number of sample or thresholding value and main trajectory map: The number of sample is 100, 1,000, 10,000, 100,000 from left to right. And the thresholding value is N/1,000, N/100, N/20 (N is the number of sample) from top to bottom.

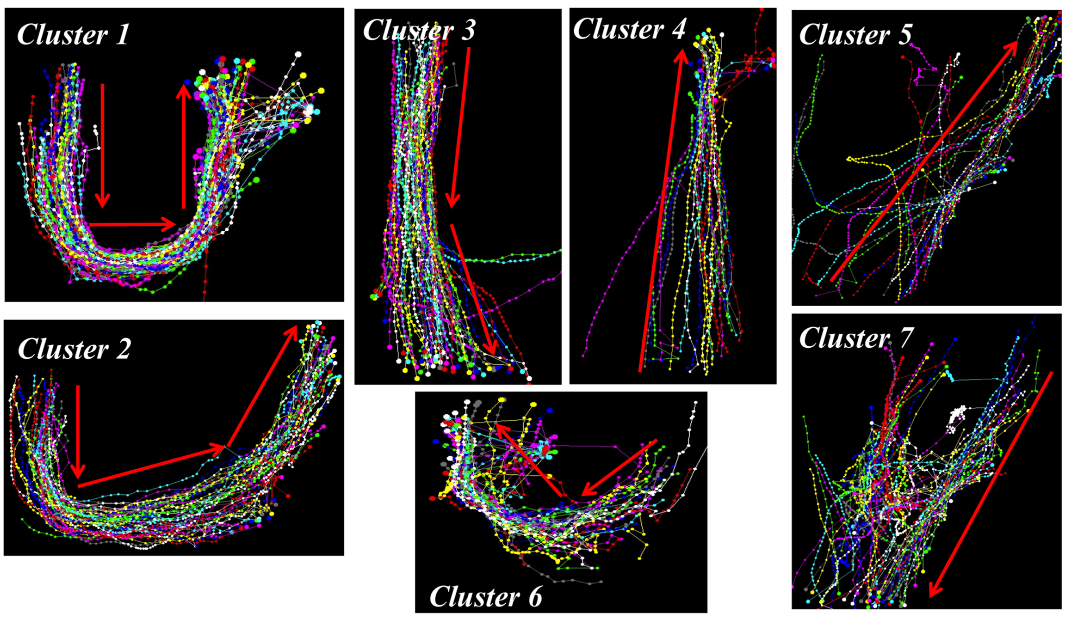

Figure 7. Trajectory clustering with main trajectory map.