Big Trajectory Data Analysis for Clustering and Anomaly Detection

○Hirokatsu Kataoka †‡, Masaki Onishi‡, Kenji Iwata‡, Yutaka Satoh‡, Ikushi Yoda‡, Yoshimitsu Aoki†

† Keio University ‡ National Institute of Advanced Industrial Science and Technology (AIST)

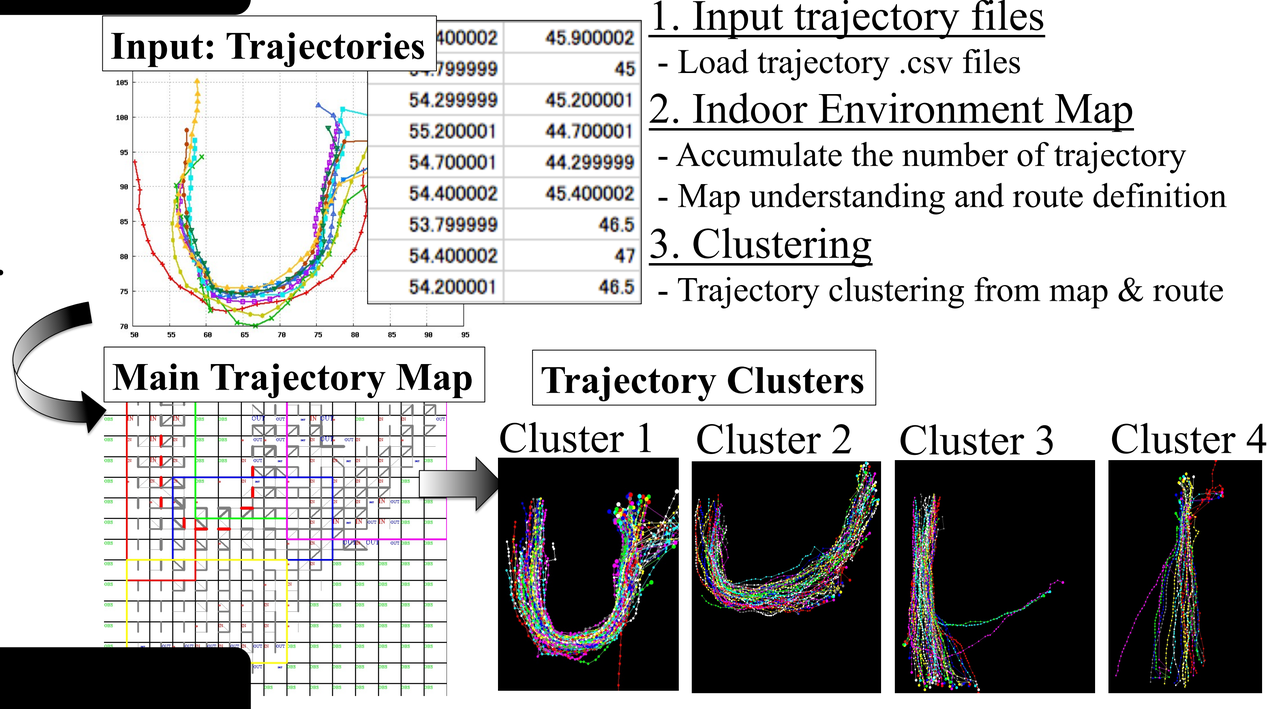
Introduction

Big Trajectory Data Analysis

- CV:Tracking Mobile terminal to capture trajectories
- Big data analysis in office/building/store with traj. data
- For congestion reducing and shopping pattern analysis…

Proposed Approach

- Environment analysis with Main Trajectory Map
- Clustering applying route information
- Anomaly detection



1. Input trajectory files
- Load trajectory .csv files

2. Indoor Environment Map
- Accumulate the number of trajectory
- Map understanding and route definition

3. Clustering
- Trajectory clustering from map & route

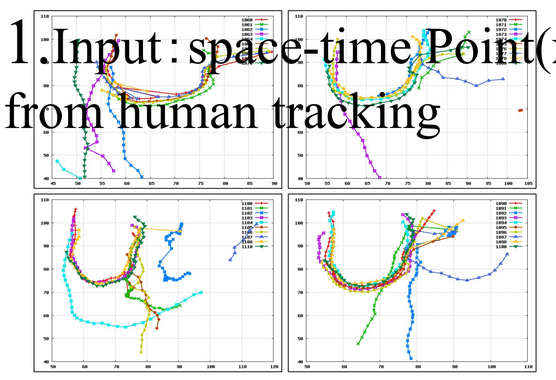
Big Trajectory Data Analysis

We have a large amount of trajectory data in a building

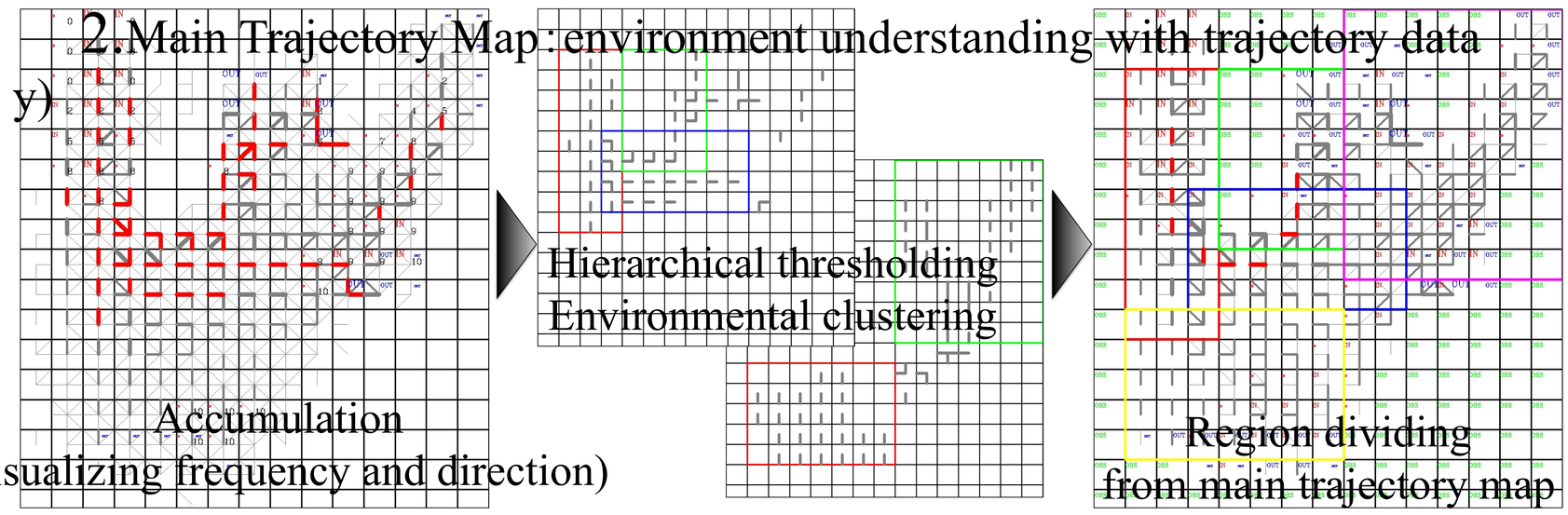
Given $DB = (T_1, T_2, \ldots, T_N)$, where $T_n = (P_1, P_2, \ldots, P_M)$, $P = (P_x, P_y)$

Main Trajectory Map: $M(i, j) = \Sigma^N_{n=1} \Sigma^M_{m=1} \{+1\}$ $(if\ P_x / s = i\ and\ P_y / s = j)$

1.Input : space-time Point(x, y) from human tracking



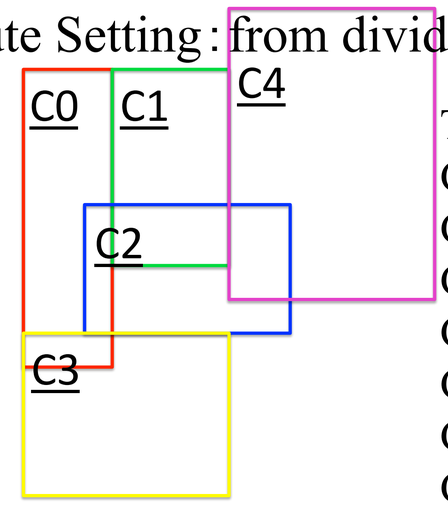
2.Main Trajectory Map : environment understanding with trajectory data



Environment Cluster Rectangle: $R_i = (x_i, y_i, w_i, h_i)$

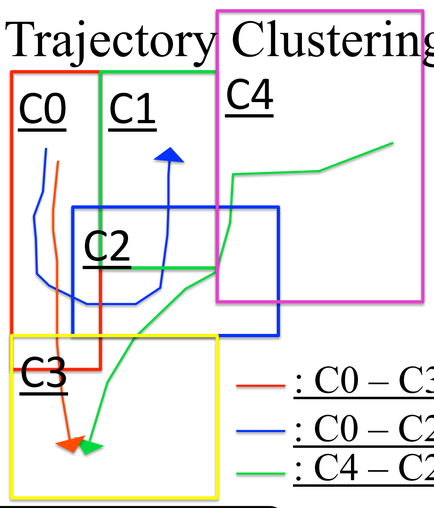
Trajectory Cluster: C = i (if trajectory route = R) where R is calculated route with trajectory data

3. Route Setting : from divided area



Trajectory Cluster
Cluster 0: C0 – C2 – C1
Cluster 1: C0 – C2 – C4
Cluster 2: C0 – C3
Cluster 3: C3 – C2 – C1
Cluster 4: C3 – C2 – C4
Cluster 5: C4 – C1
Cluster 6: C4 – C2 – C3

4. Trajectory Clustering : clustering using route

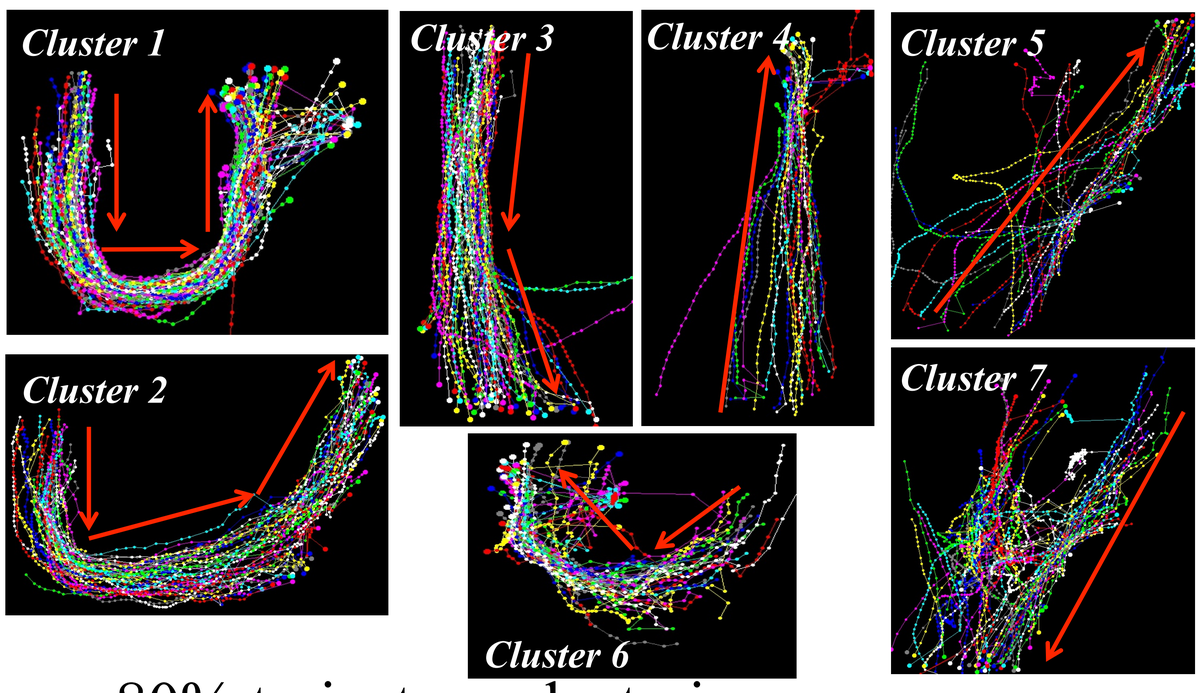
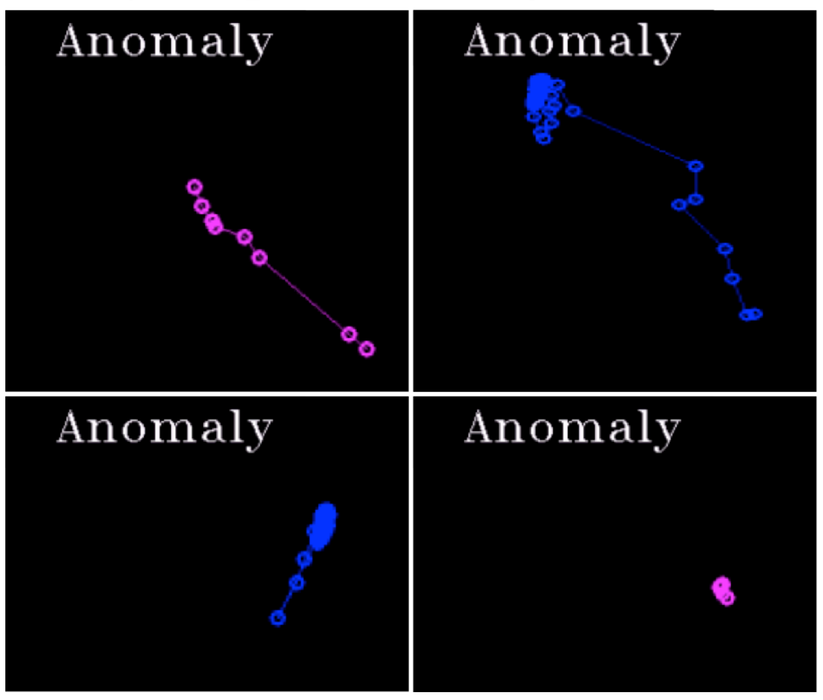


- Recording area transition as a route
- Considering difference of route length
- Clustering dealing with In/Out

: C0 – C3 -> Trajectory Cluster 2
: C0 – C2 – C1 -> Trajectory Cluster 0
: C4 – C2 – C3 -> Trajectory Cluster 6

Trajectory Analysis Experiment

Input: 300,000 (3.0×10^5) trajectories; Approach: Main Trajectory Map, Clustering



80% trajectory clustering accuracy