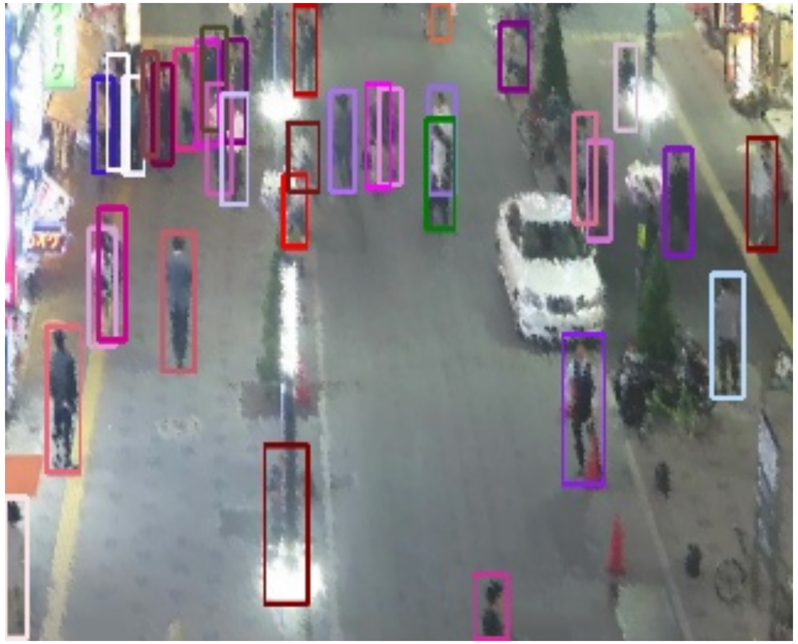
Multiple Person Tracking based on Gait Identification using Kinect and OpenPose

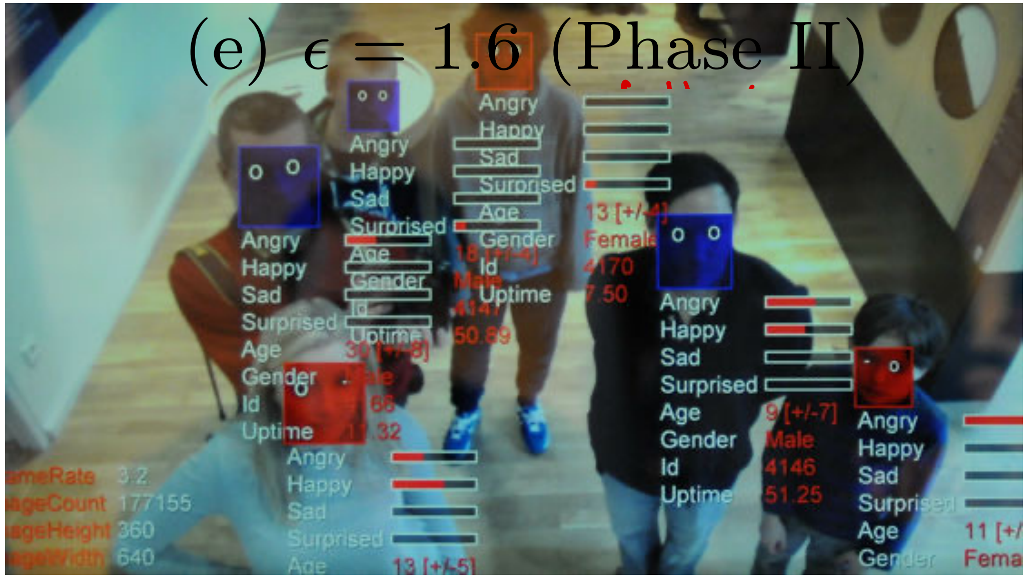
Ryotaro Toma, Terumi Yaguchi, and Hiroaki Kikuchi

Meiji University

Background

- Multiple human tracking
  - surveillance camera
  - crowd analysis
  - behavior understanding
  - human-computer interaction





Existing Studies

- Multi-Camera Tracking [Amosa 2023]
  - Most common approaches. Use multiple cameras to track individuals.
- Depth Sensors [Preis 2012, Mori 2019]
  - Sensors, such as Microsoft Kinect, provides most accurate depth information.
- Device-based Tracking [Muaaz 2017]
  - Devices such as Android smartphone record accelerometer and gyroscope data.

Two Approaches

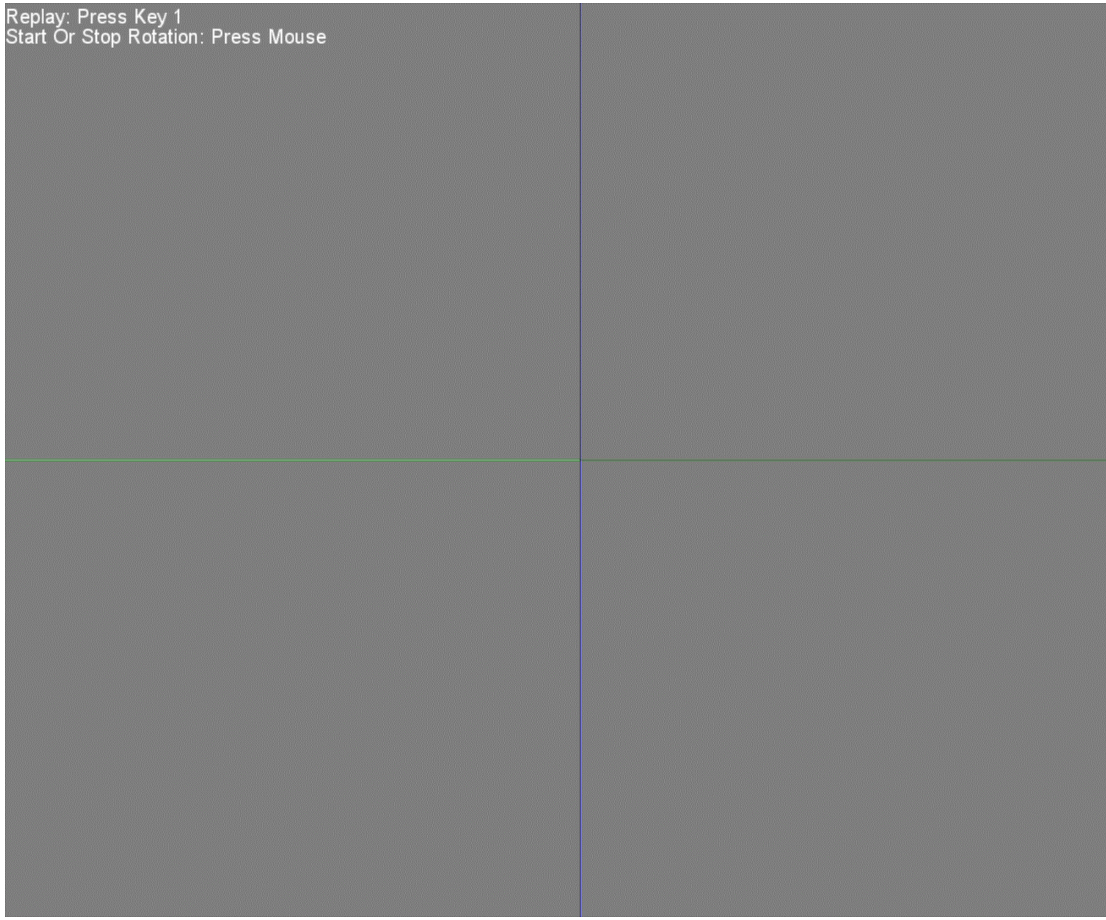
- Computer vision
  - OpenPose [20]

- Depth sensor
  - Kinect [14]

	Computer Vision	Depth Sensor
Pros	Rich information, wide range	Accurate 3D data, real-time detection
Cons	2D image (no depth information), Computing intensive	limitation of tracking ranges, # individuals (max 6 indviduals)

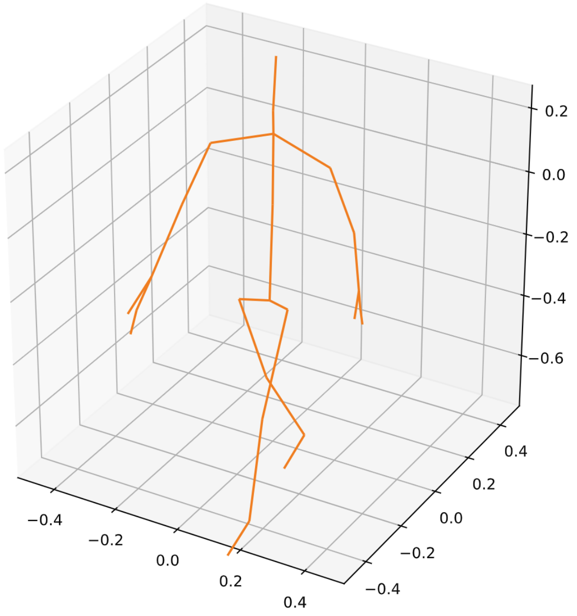
Challenges (Feature from time-series)

- 1. Walking video is **unstable** time-series data. It is not trivial to get stable biometric features.
  - Human body varies over individuals. It is not easy to find common features for attributes.

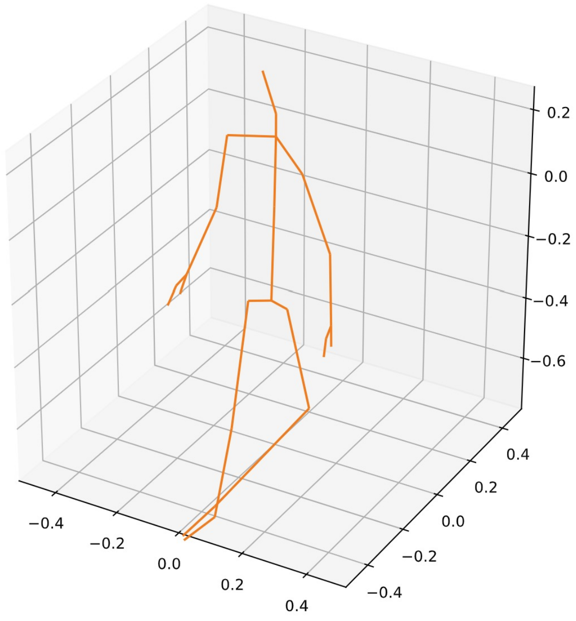


Challenges (Sensing errors)

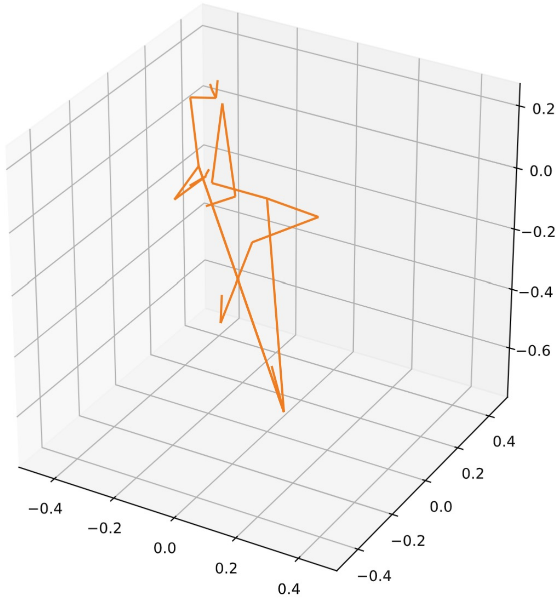
- 2. Devices fails to detect motions



Normal



Partially malfunction frames



Completely malfunction frames

Research Questions

- Can we have stable features based on unstable dynamic time-series data?
- Which approach is superior than others? Computer vision or Depth sensor?
- Is it feasible to combine two approaches to composite the pros and cons?

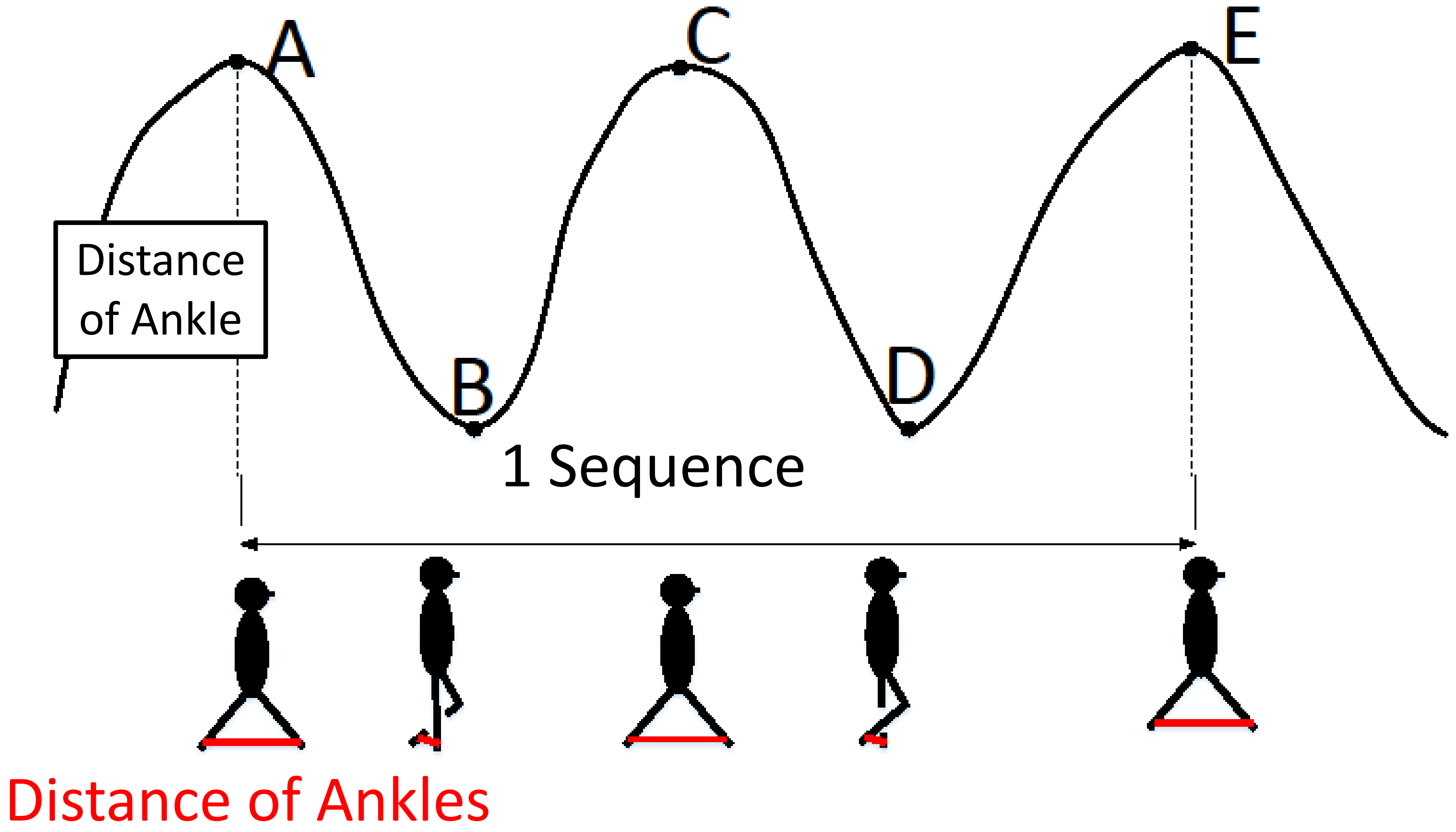
Our Approaches

- **Dynamic Time Warping** (DTW) distance [15]
  - Robust metrics quantifying the distance between two time-series data of 3D points.
  - Based on DTW distance, we identify walking cycle and track multiple walking individuals
- Multiple-person **Experiment**
  - Evaluate the baseline accuracy of gait tracking.
  - Compare two tracking approaches and figure out the tradeoff between accuracy and the coverage.

Proposed Tracking

- Tracking based on gait
  - 1. cycle extraction
  - 2. DTW Distance
  - 3. Human identification

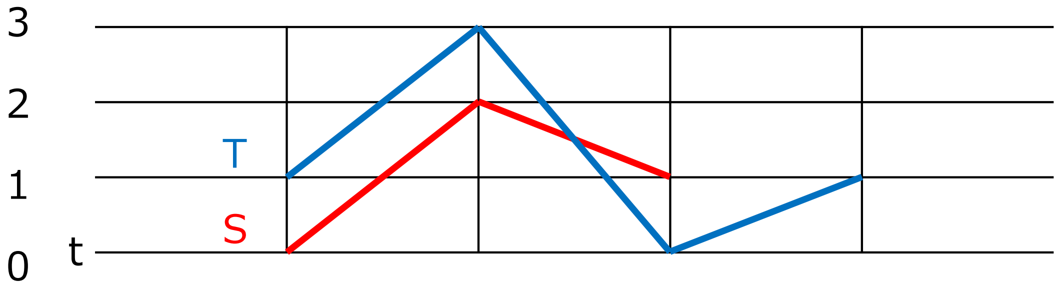
1. Cycle extraction



2. DTW Distance

- DTW
  - Similarity between two time-series data T and S
  - Robust against inconsistent dimensions
  - Dynamic programing to find the shortest distance between T and S

$$f(i,j) = \|p_i - q_j\| + \min\begin{cases} f(i,j-1) \\ f(i-1,j) \\ f(i-1,j-1) \end{cases}$$



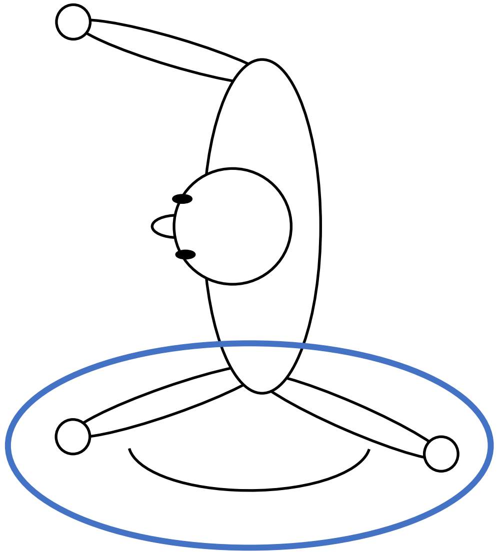
Two time-series data T and S

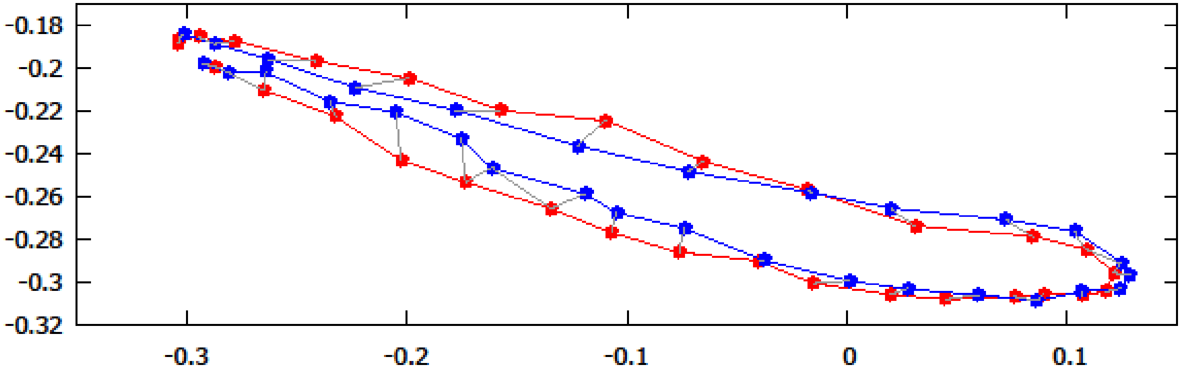
result

1	∞	2	4	3	3
2	∞	2	2	4	5
0	∞	1	4	4	5
	0	∞	∞	∞	∞
S／T		1	3	0	1

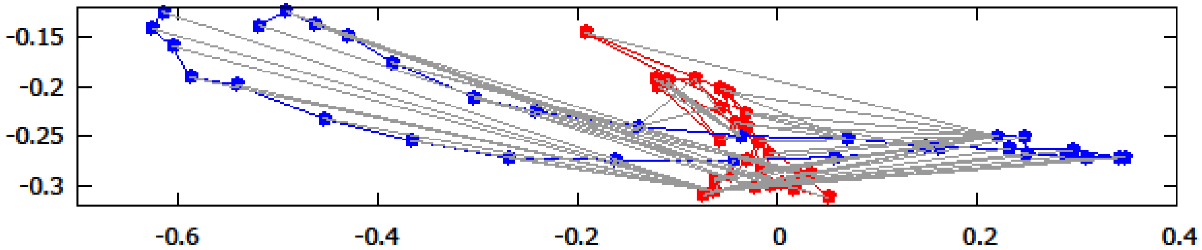
Distance Matrix

3. Test if data are same or different persons





With own data(distance=0.45m)

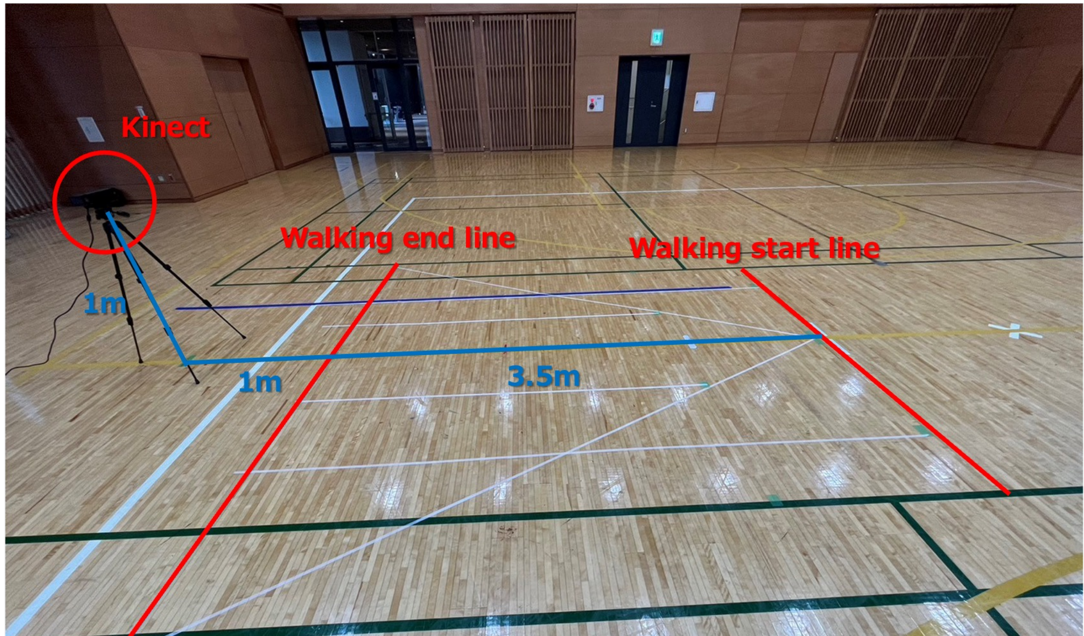


With other person's data(distance=12m)

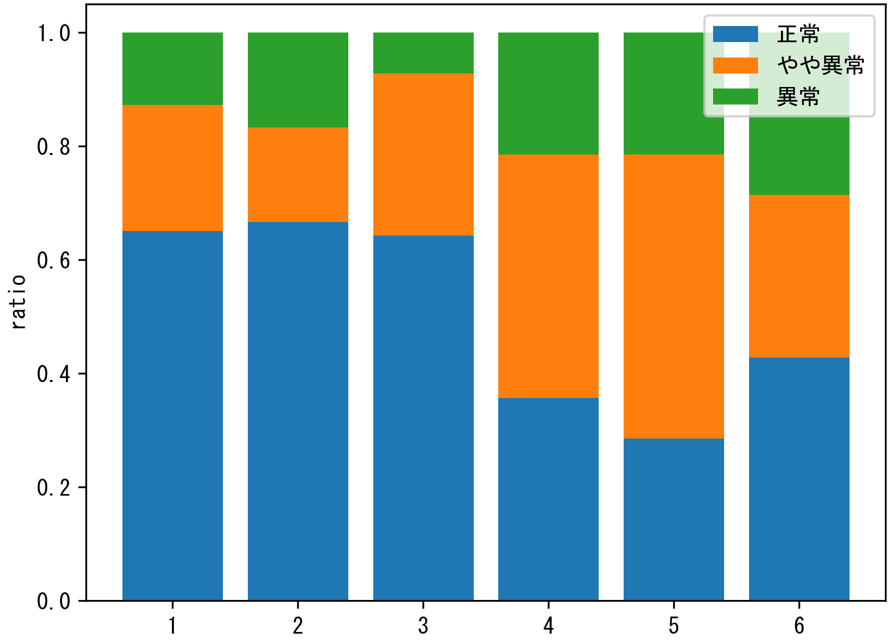
Experiment

- Objectives
  - 1. Quantify Kinect detection accuracy
    - Success Detection rate with regard to orientation, front and side
  - 2. Multiple Person Tracking
    - Success Tracing rate with regard to the number of individuals (n = 1,..,6)
- Data

item	value
date	July 16, 2022
venue	gymnastic hall, Meiji University
age	20's
population	7 (4 male, 3 female)



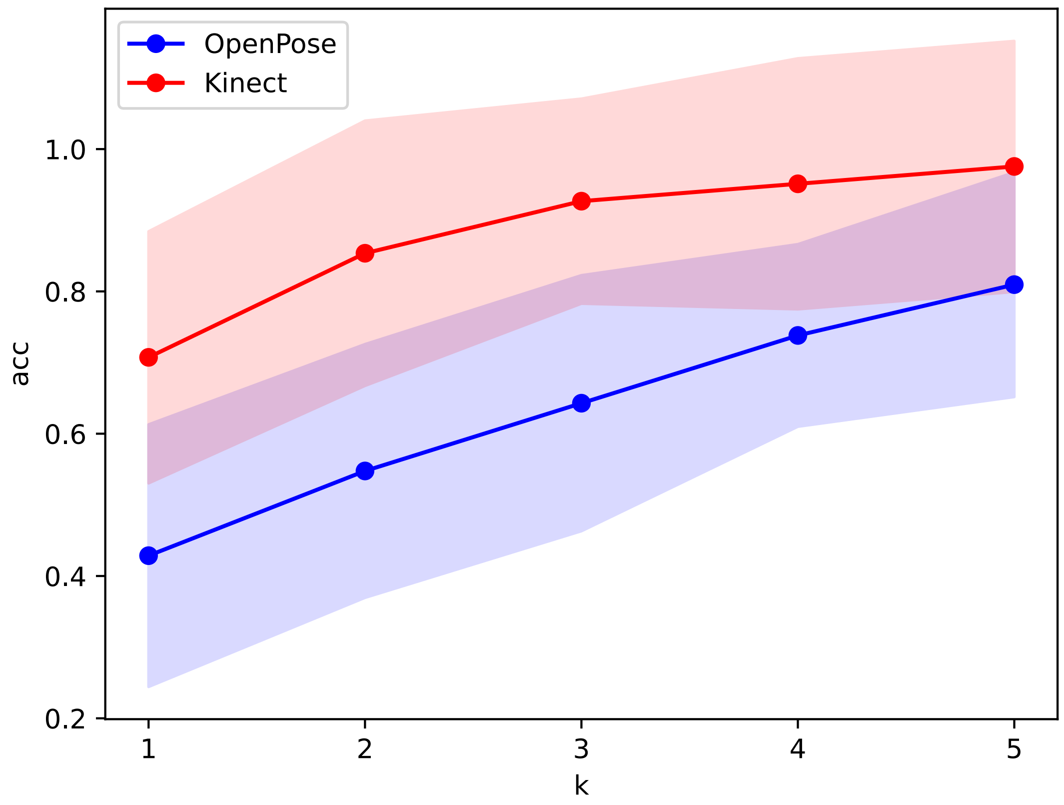
Result 1



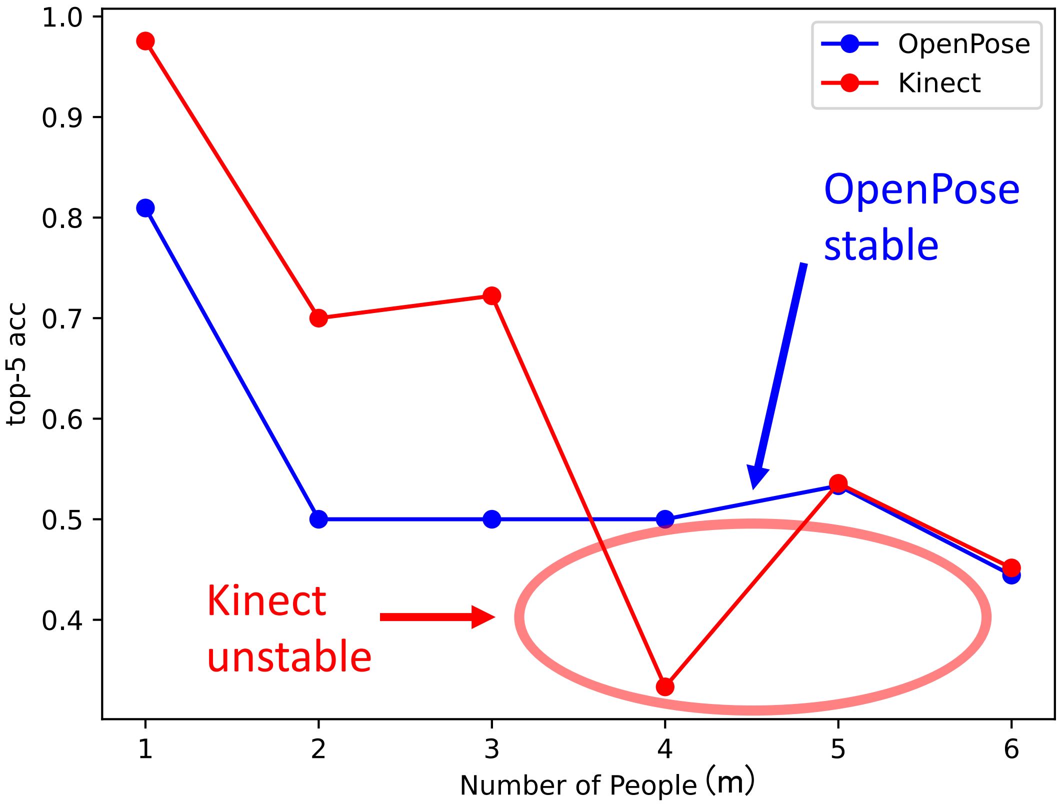
Success rate wrt n

orientation	success	partially malfunction	malfunction
front	1.0 (21/21)	0.0 (0/21)	0.0 (0/21)
side	0.48 (20/42)	0.33 (14/42)	0.19 (8/42)
total	0.65 (41/63)	0.22 (14/63)	0.13 (8/63)

Result 2



Top-k accuracy



Top-5 accuracy wrt number of individuals

Discussion

- Depth sensor’s robustness
  - We find a significant decline in tracking accuracy of Kinect as the number of individuals increased.
  - The maximum number of individuals is 3, less than the specification of Kinect.
- OpenPose’s robustness
  - We find that OpenPose exhibits greater robustness in the number of individuals being tracked.
  - Tracking accuracy depends condition in environment, such as brightness and obstacles.
- Privacy Concerns
  - Regulation (GDPR and CCPA) forbid the collection of personal information without consent. Tracking based on gait information raises privacy concerns.

Conclusions

- We have examined the performance of multiple human tracking using two approaches: OpenPose and Kinect.
- By employing the DTW distance metric, we have demonstrated the feasibility of tracking multiple humans based on time-series 3D point data.
- Our experimental results indicate that the depth sensor, Kinect, is capable of accurately tracking multiple individuals. However, its accuracy diminishes when there are more than three individuals walking simultaneously or when their walking orientations differ.
- Consequently, we conclude that person tracking is influenced by factors such as the orientation of walking and the number of individuals being tracked.