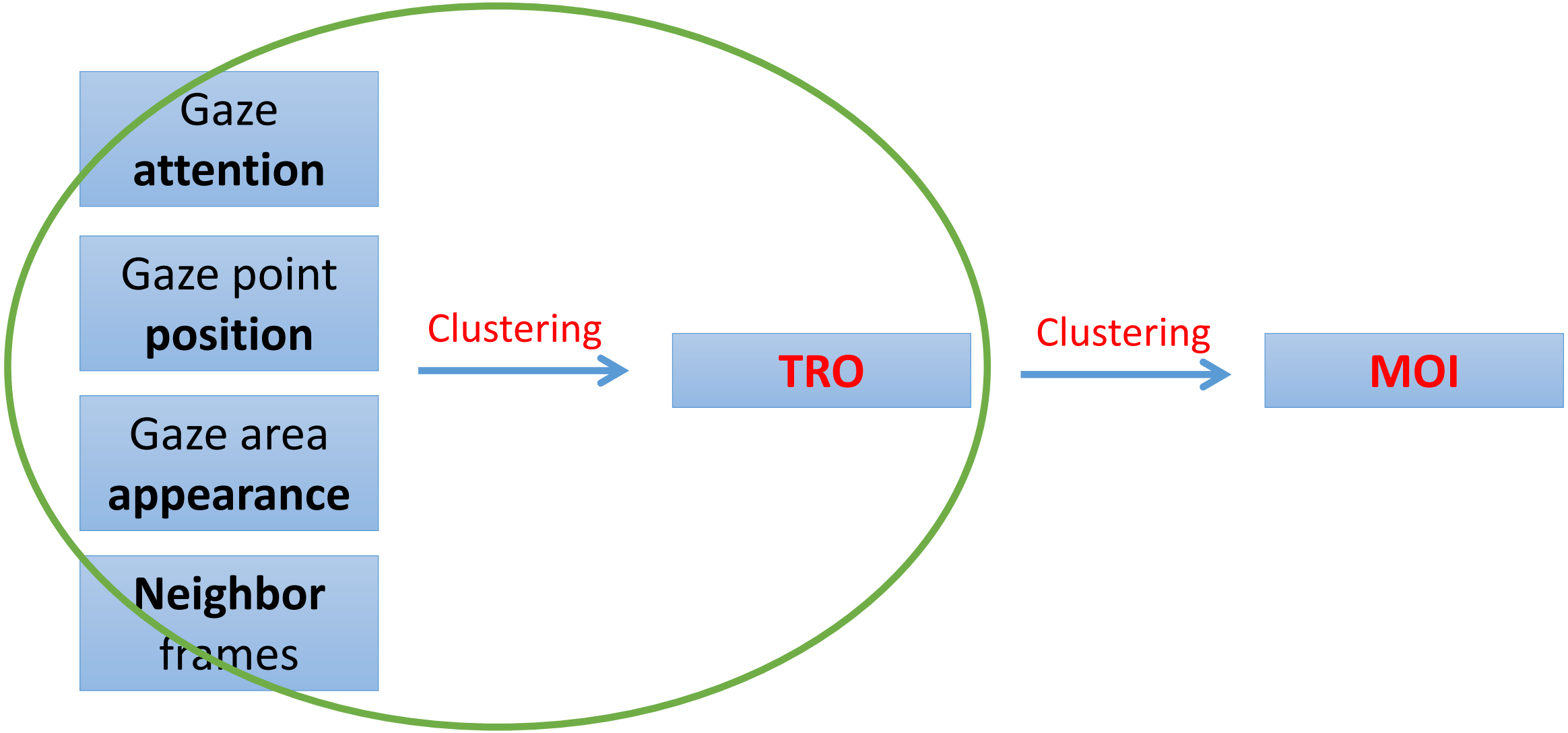
The experiments for "You-Do, I-Learn"

Presenter: Wenguang Mao

Instructor: Kristen Grauman

Author for the paper: Dima Damen

Recap of the Paper

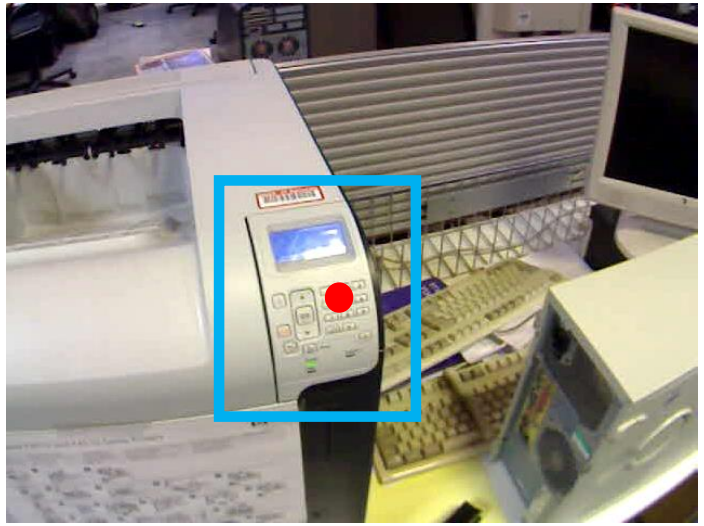


Experiment Setup

- Dataset: **Bristol Egocentric Object Interactions Dataset**

Experiment Setup

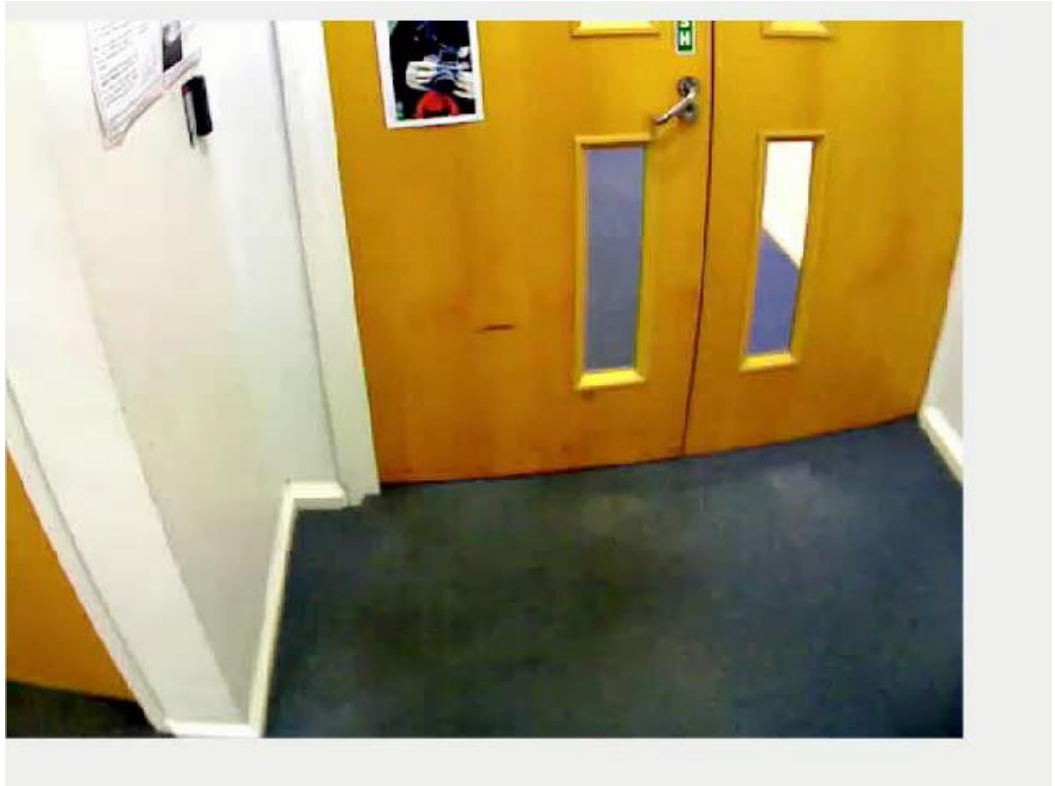
- Dataset: **Bristol Egocentric Object Interactions Dataset**
  - Egocentric videos at 6 locations
  - Gaze point on each frame
  - Gaze positions in 3D space
  - Gaze fixation on each frame
  - Ground truth positions of TROs
  - 3D map for each location, 3D positions of the camera for each frame, ……



- Code: VLFeat, Matlab toolboxes, and programs written by myself

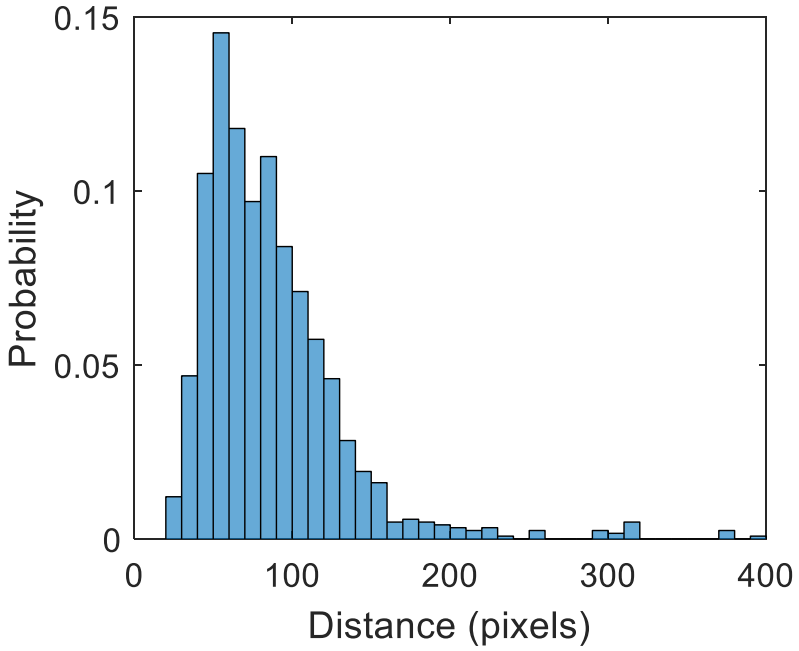
Why Need Gaze Info

- Given an egocentric image, which part of the image do you think I am focusing on?
  - Center of image?
  - Blue point: center of image
  - Red point: gaze point

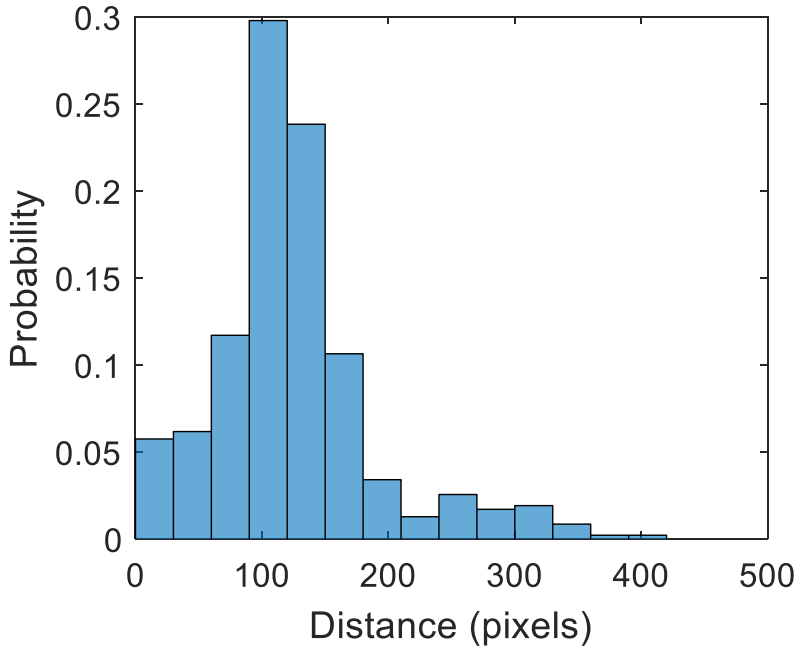


Why Need Gaze Info

- The distance between the center and the gaze point



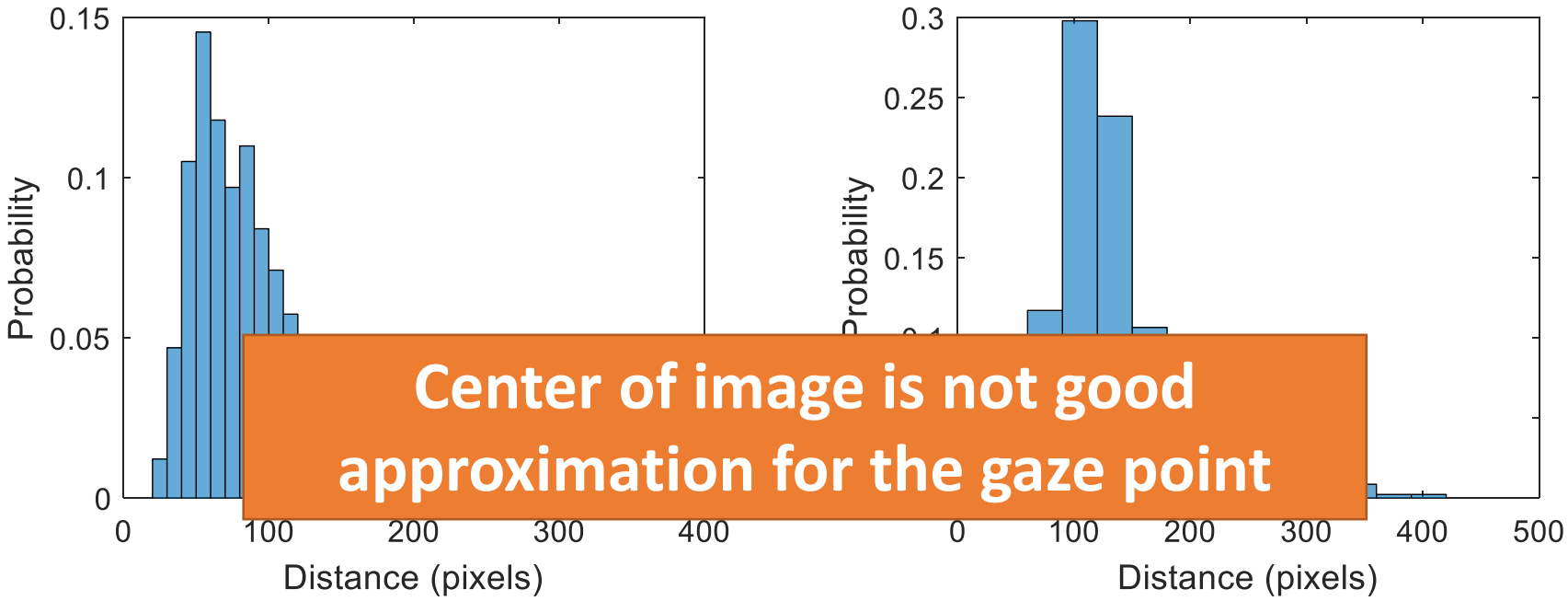
(a) Desk



(b) Door

Why Need Gaze Info

- The distance between the center and the gaze point

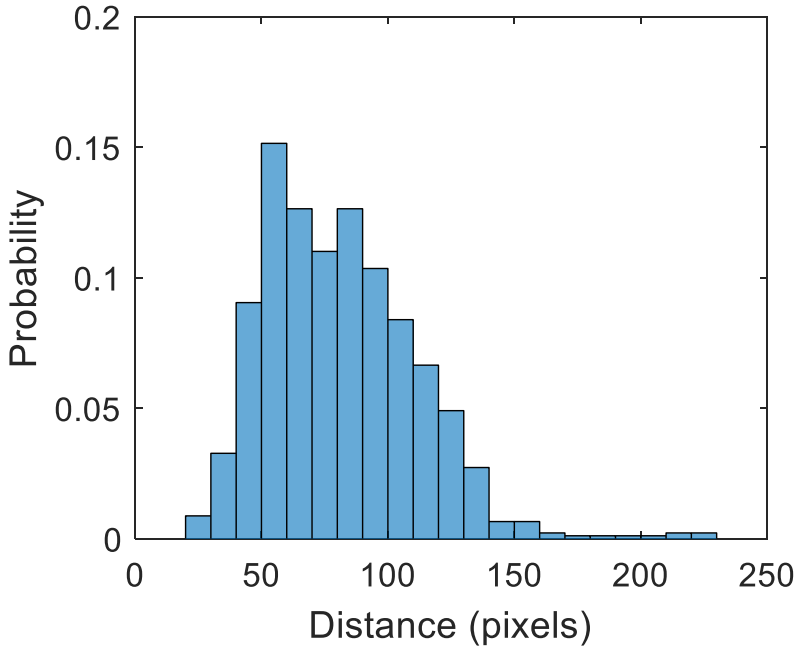


(a) Desk

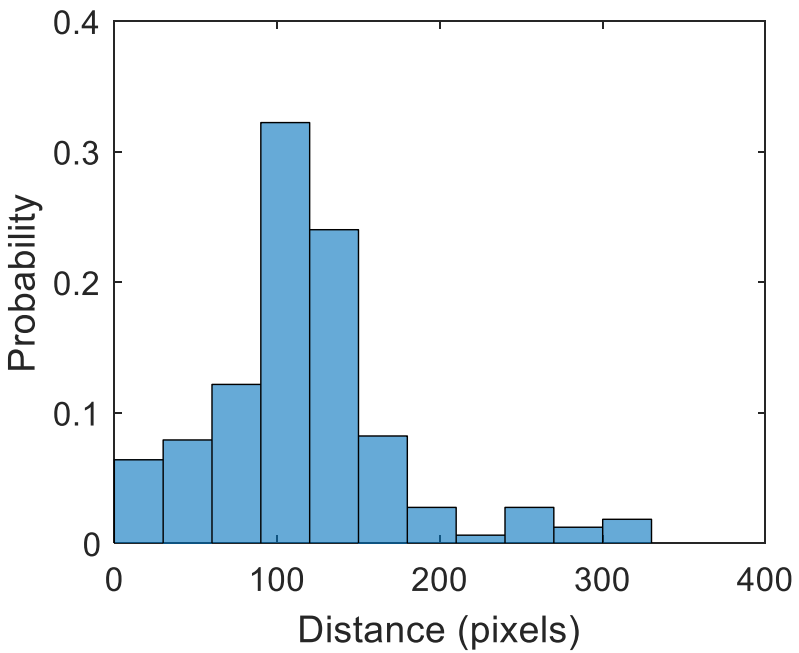
(b) Door

Why Need Gaze Info

- The distance between the center and the gaze point (**during gaze fixation**)



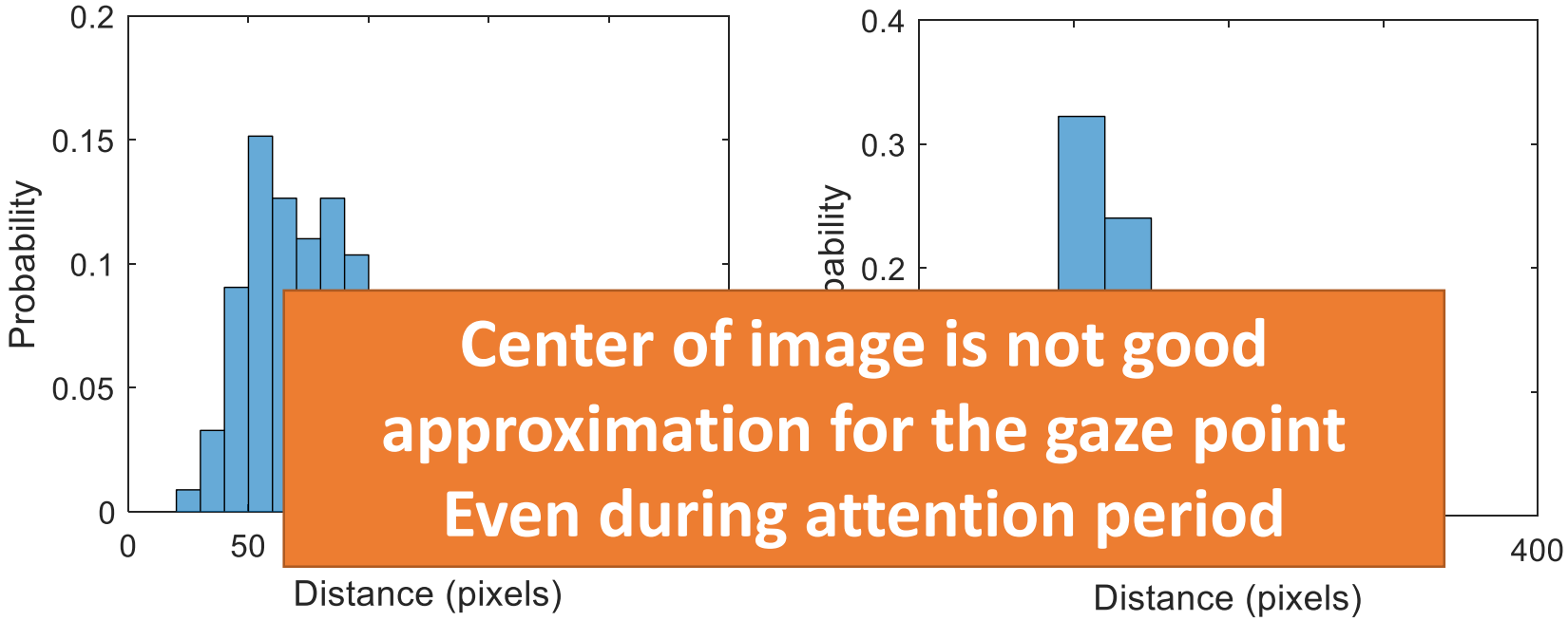
(a) Desk



(b) Door

Why Need Gaze Info

- The distance between the center and the gaze point (**during gaze fixation**)



(a) Desk

(b) Door

How Gaze Fixation Helps

- Do you think there is any TRO in the video clips
  - Red dot: gaze point

How Gaze Fixation Helps

- Do you think there is any TRO in the video clips
  - Red dot: gaze point

Gaze fixation helps identify a TRO

How Gaze Fixation Helps

- Do you think there is any TRO in the video clips
  - Red dot: gaze point

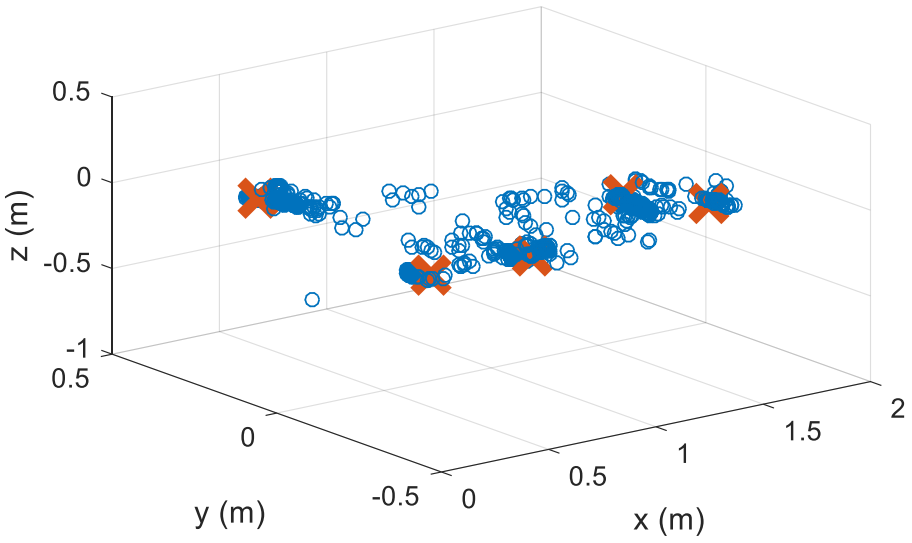
How Gaze Fixation Helps

- Do you think there is any TRO in the video clips
  - Red dot: gaze point

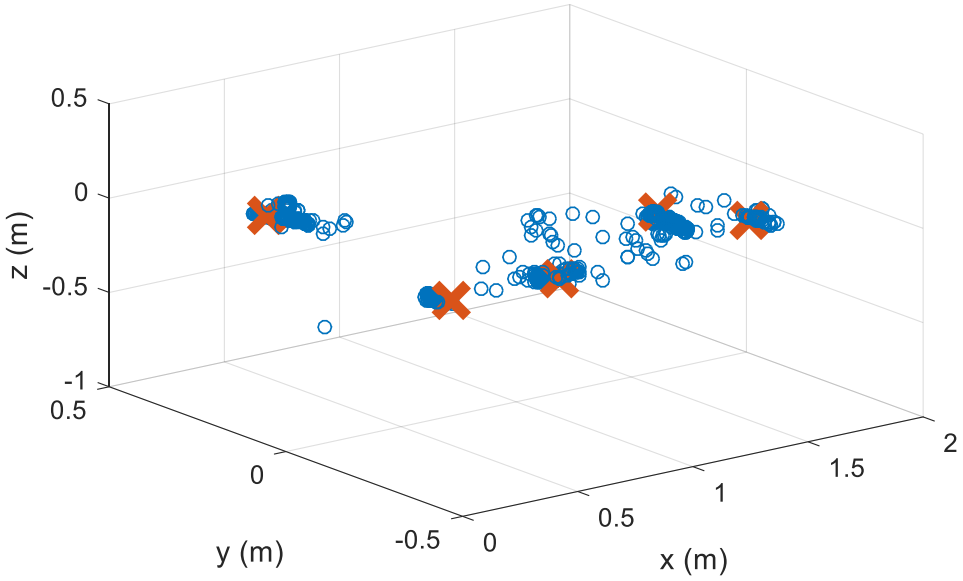
Gaze fixation alone is far from enough to find TROs

How 3D Positions of Gaze Help

- Blue circles: 3D positions of gazes in a video
- Red cross: ground truth positions of TRO



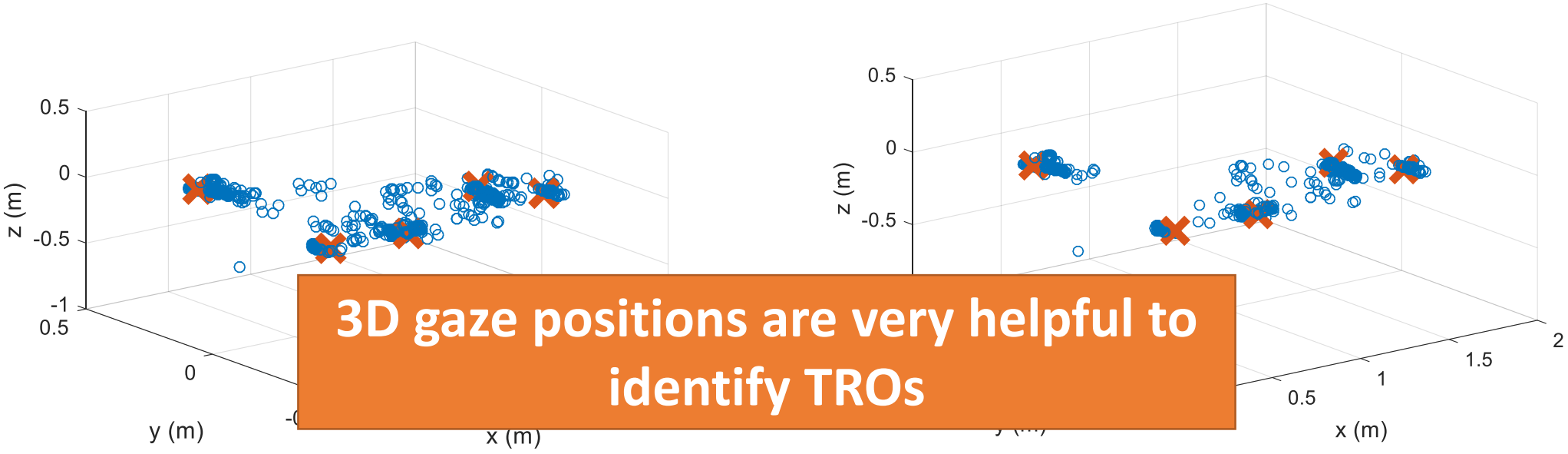
(a) Without gaze fixation filtering



(a) With gaze fixation filtering

How 3D Positions of Gaze Help

- Blue circles: 3D positions of gazes in a video
- Red cross: ground truth positions of TRO



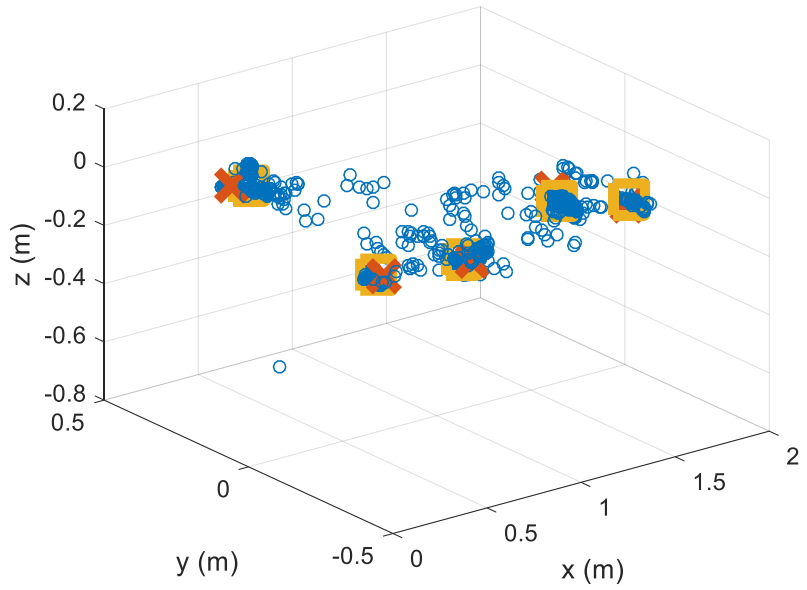
3D gaze positions are very helpful to identify TROs

(a) Without gaze fixation filtering

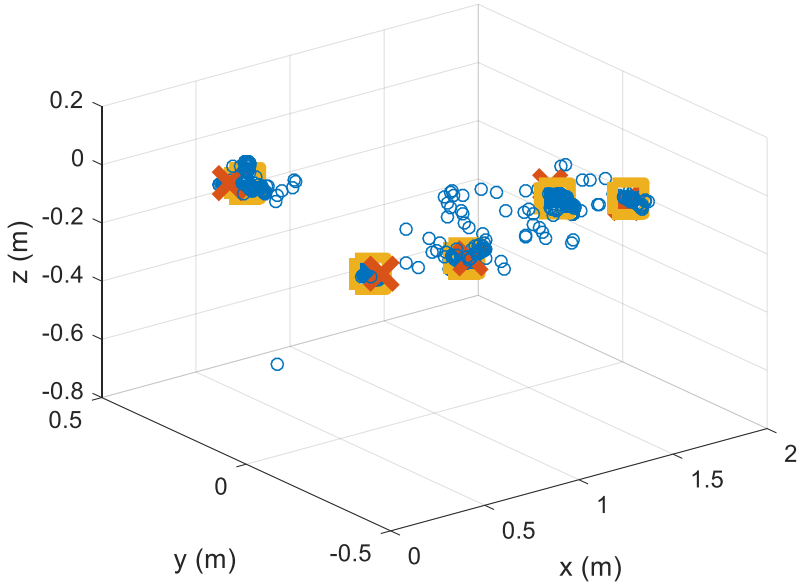
(a) With gaze fixation filtering

Clustering for Gaze 3D positions

- Right number of clusters (kmeans)
  - Yellow square: cluster center



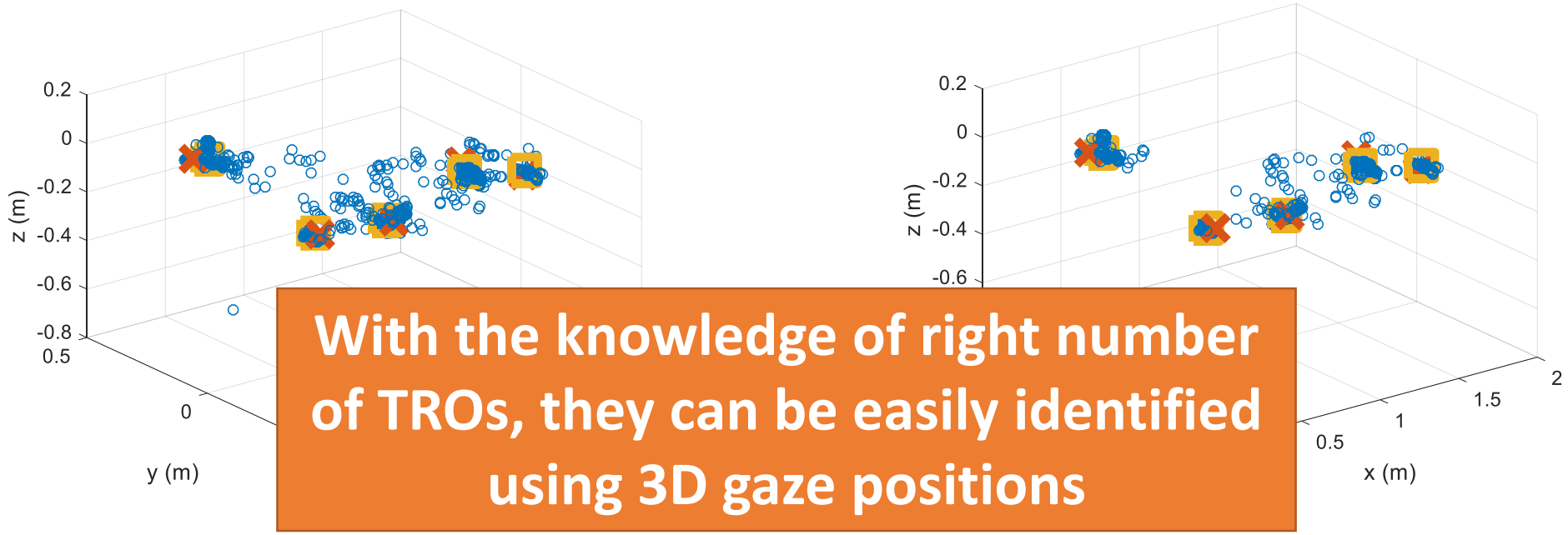
(a) Without gaze fixation filtering



(b) With gaze fixation filtering

Clustering for Gaze 3D positions

- Right number of clusters (kmeans)
  - Yellow square: cluster center



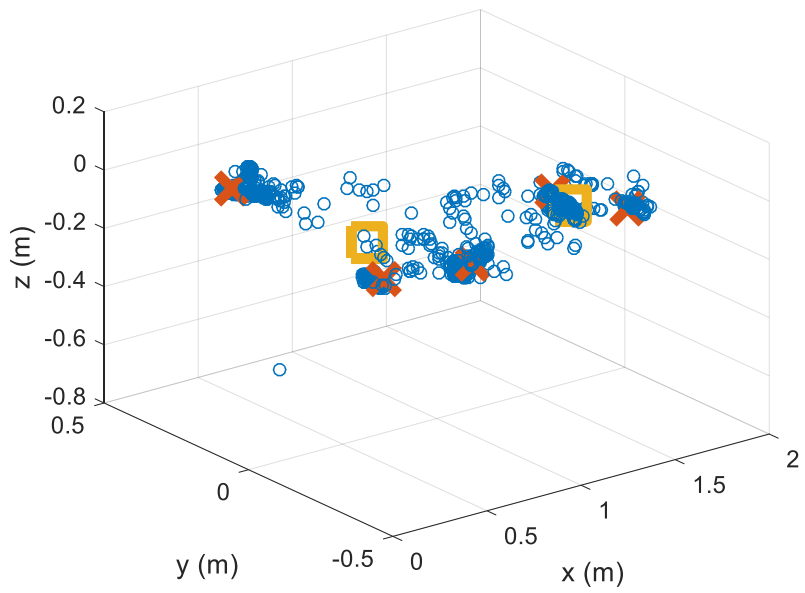
With the knowledge of right number of TROs, they can be easily identified using 3D gaze positions

(a) Without gaze fixation filtering

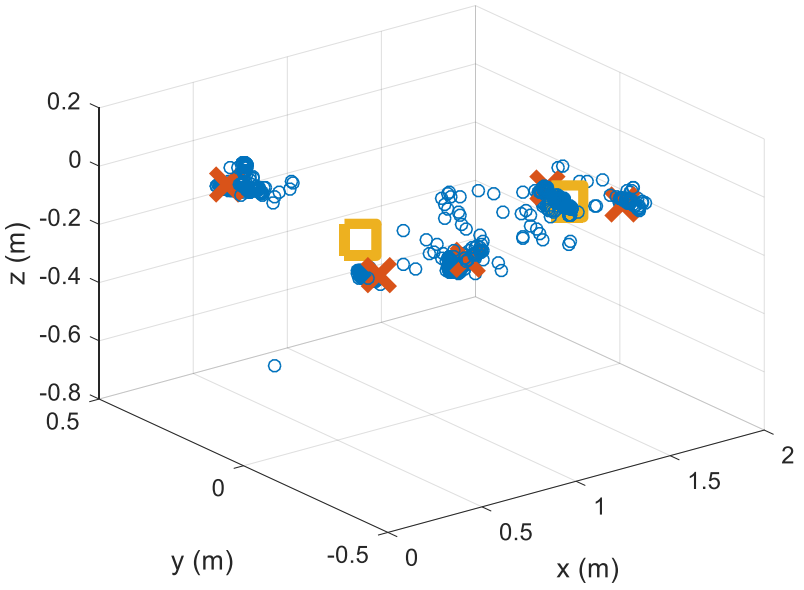
(b) With gaze fixation filtering

Clustering for Gaze 3D positions

- Too less clusters
  - Yellow square: cluster center



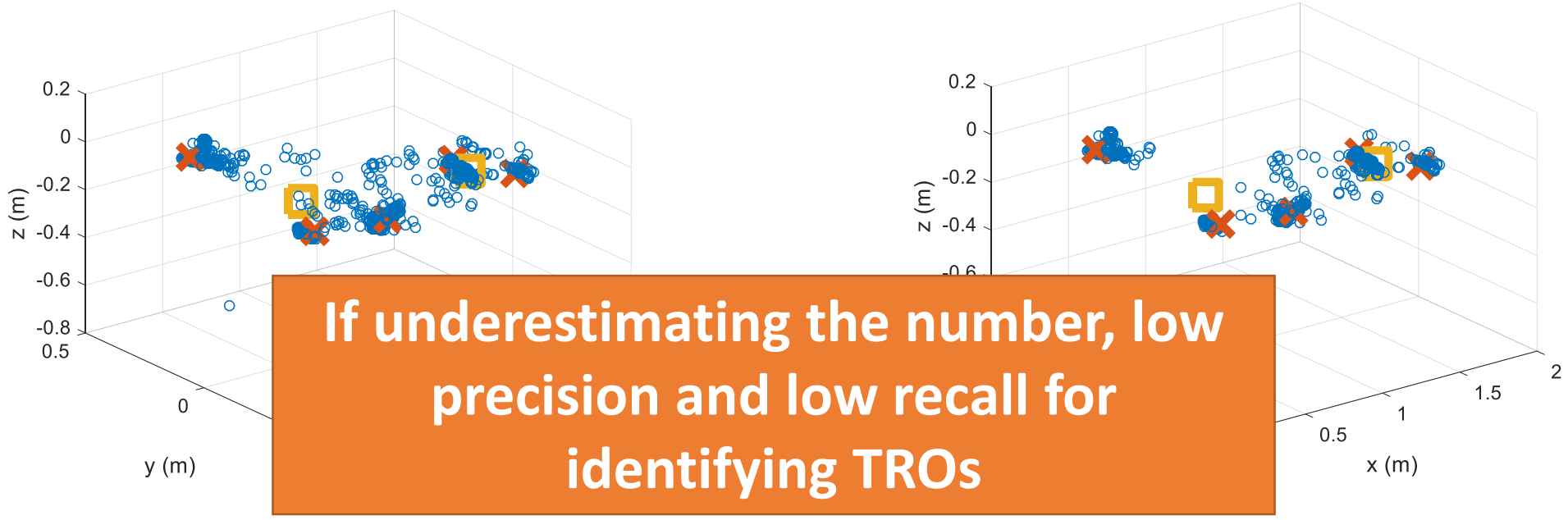
(a) Without gaze fixation filtering



(b) With gaze fixation filtering

Clustering for Gaze 3D positions

- Too less clusters
  - Yellow square: cluster center



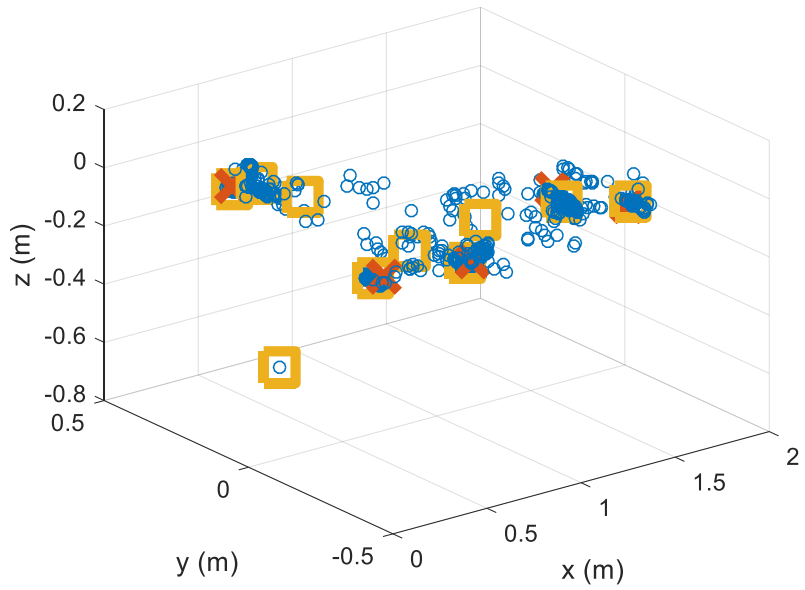
If underestimating the number, low precision and low recall for identifying TROs

(a) Without gaze fixation filtering

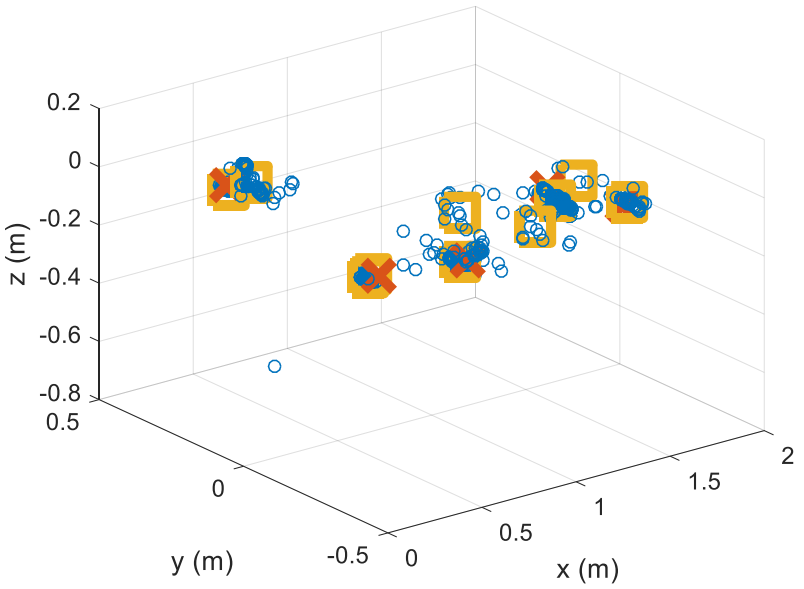
(b) With gaze fixation filtering

Clustering for Gaze 3D positions

- Too much clusters
  - Yellow square: cluster center



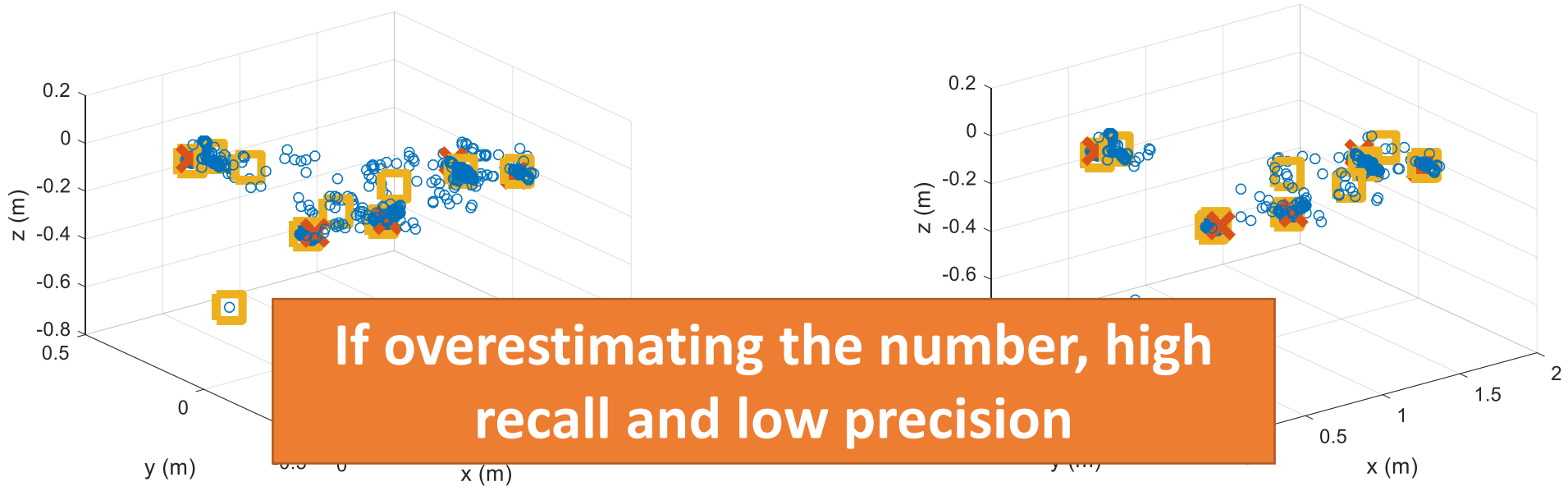
(a) Without gaze fixation filtering



(b) With gaze fixation filtering

Clustering for Gaze 3D positions

- Too much clusters
  - Yellow square: cluster center



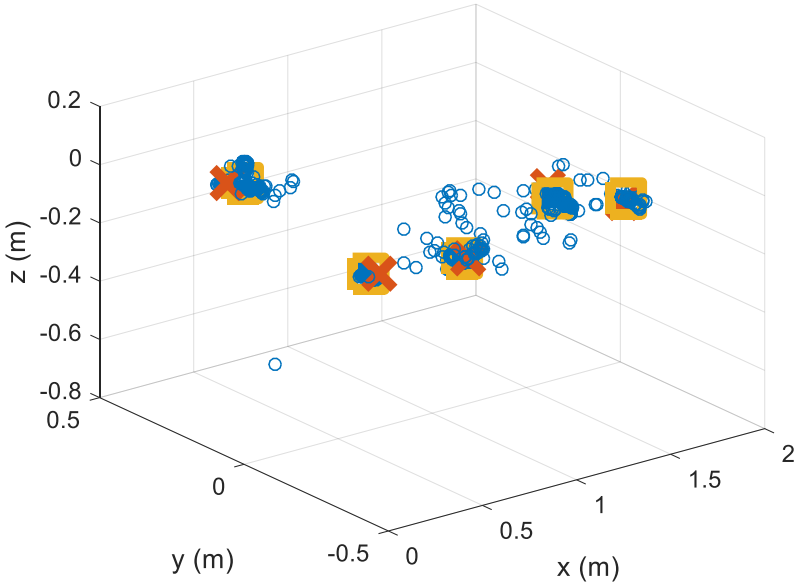
If overestimating the number, high recall and low precision

(a) Without gaze fixation filtering

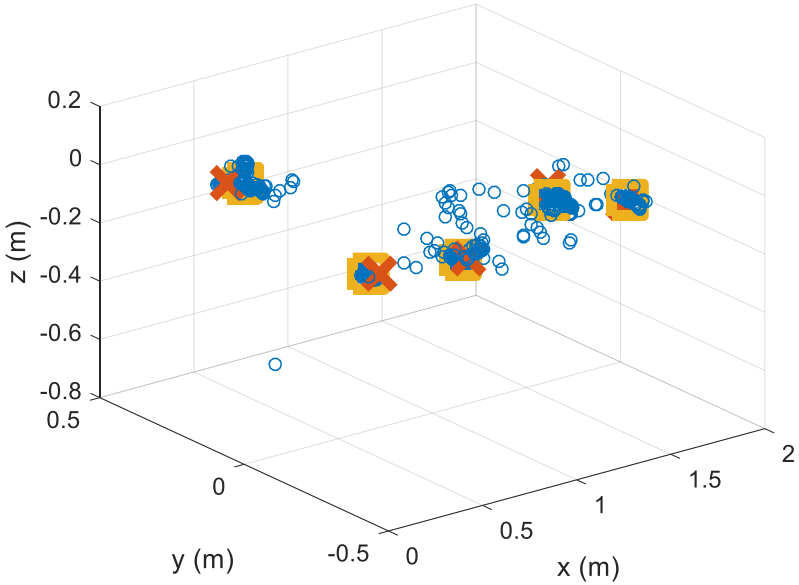
(b) With gaze fixation filtering

Spectral Clustering

- Right number of clusters



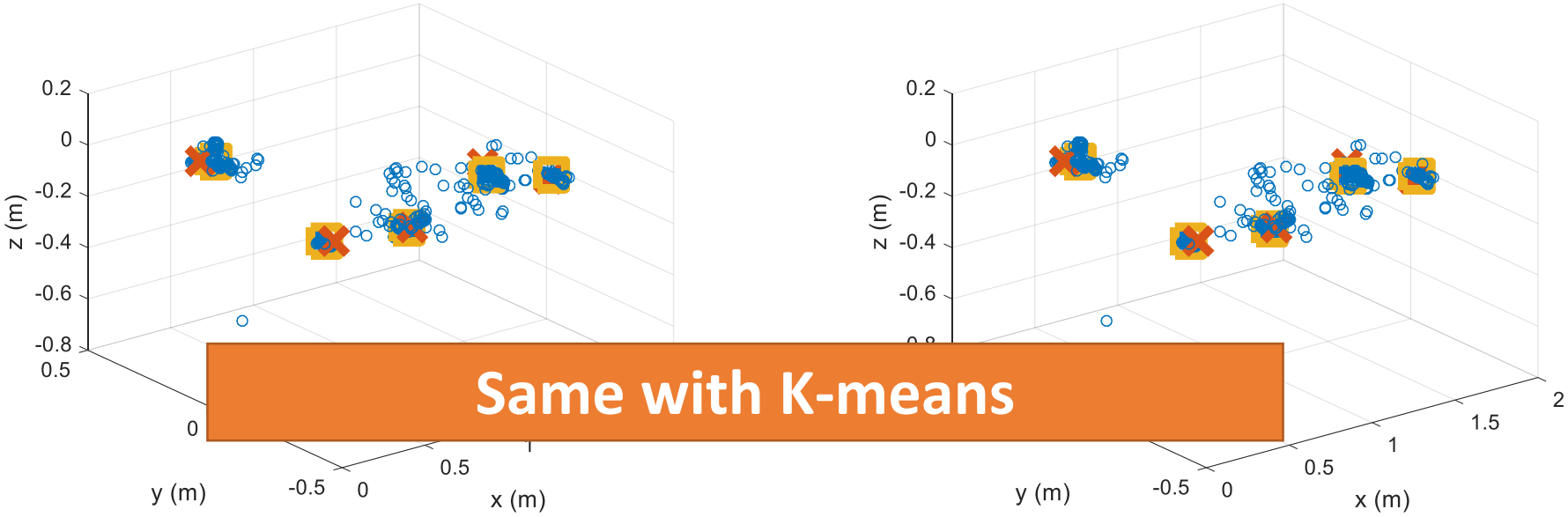
(a) kmeans



(b) spectral

Spectral Clustering

- Right number of clusters

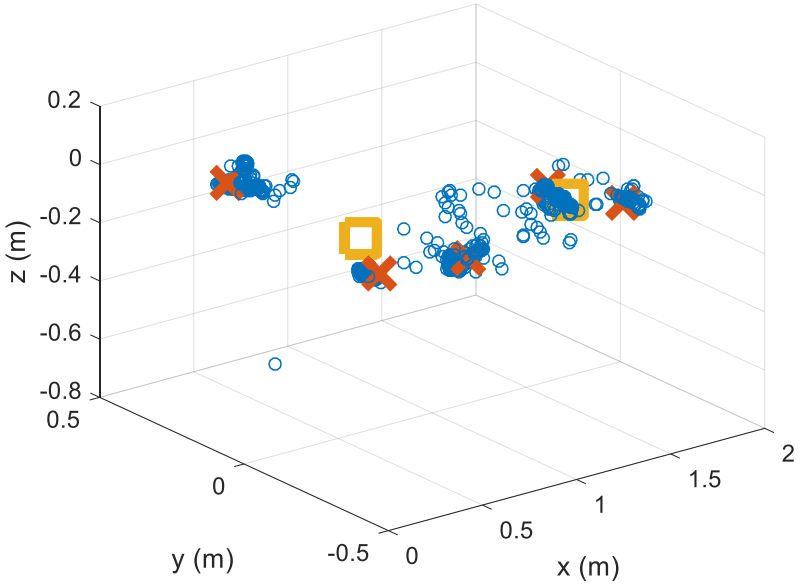


(a) kmeans

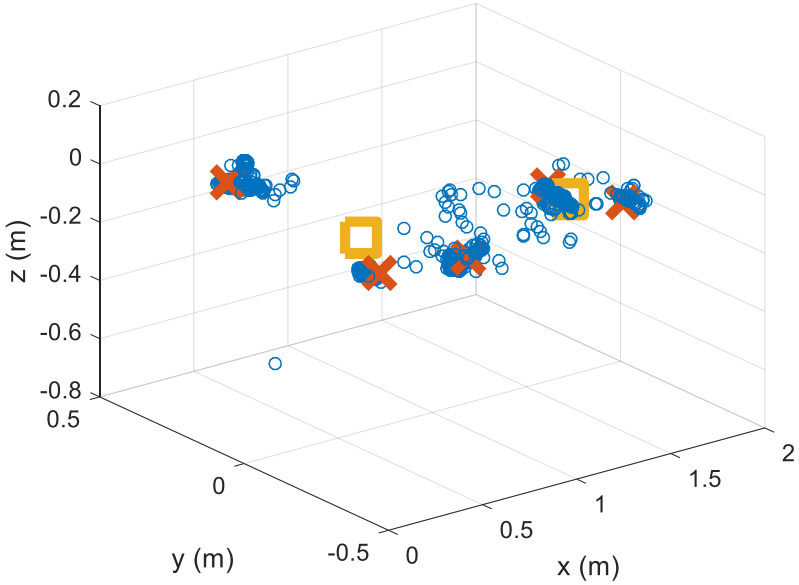
(b) spectral

Spectral Clustering

- Too less clusters



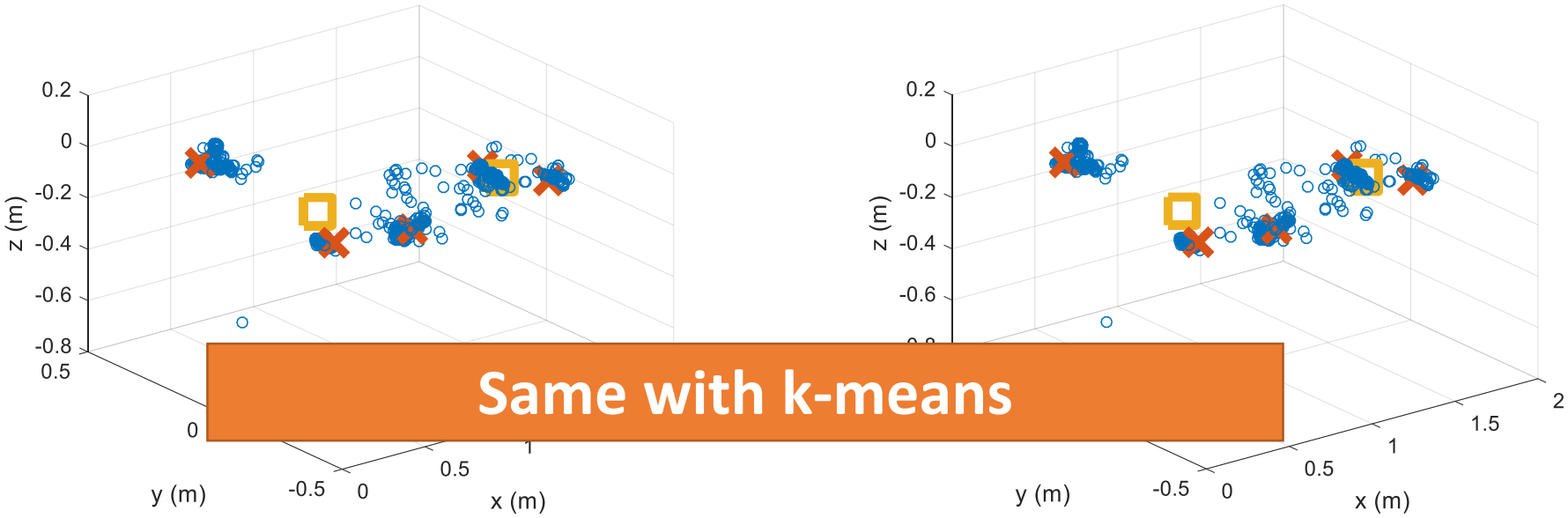
(a) kmeans



(b) spectral

Spectral Clustering

- Too less clusters

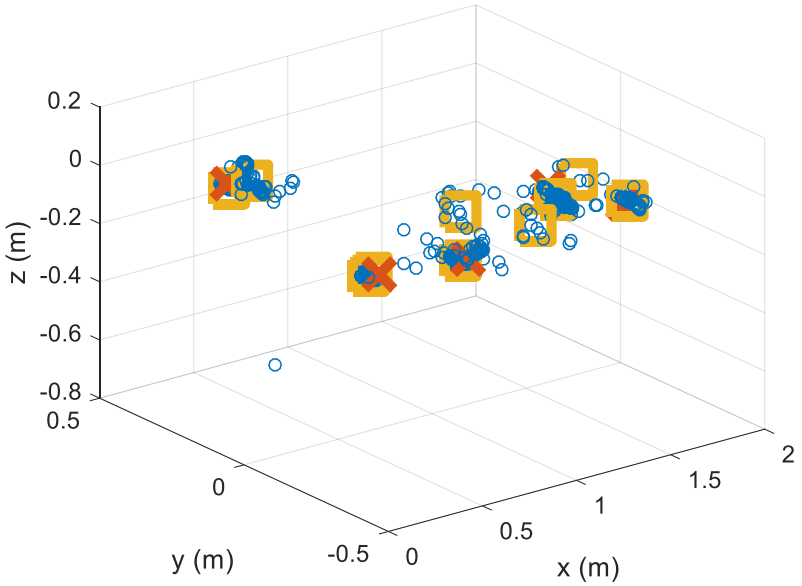


(a) kmeans

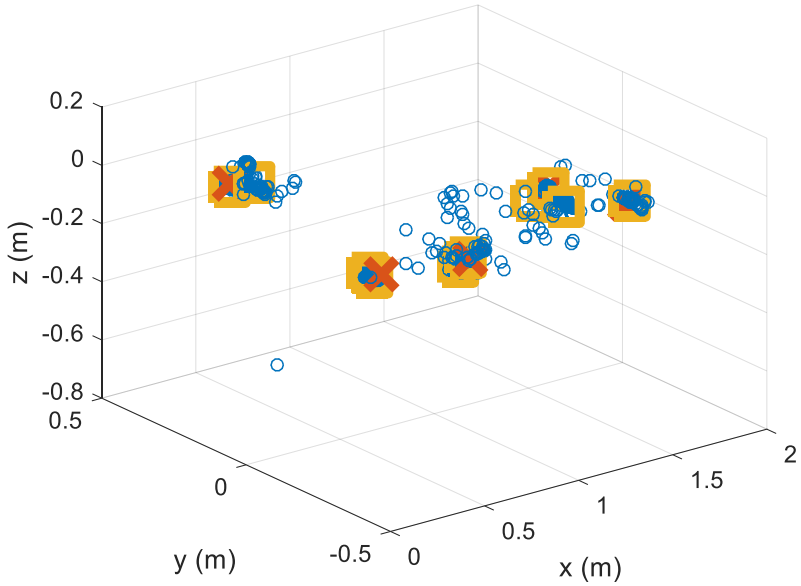
(b) spectral

Spectral Clustering

- Too much clusters



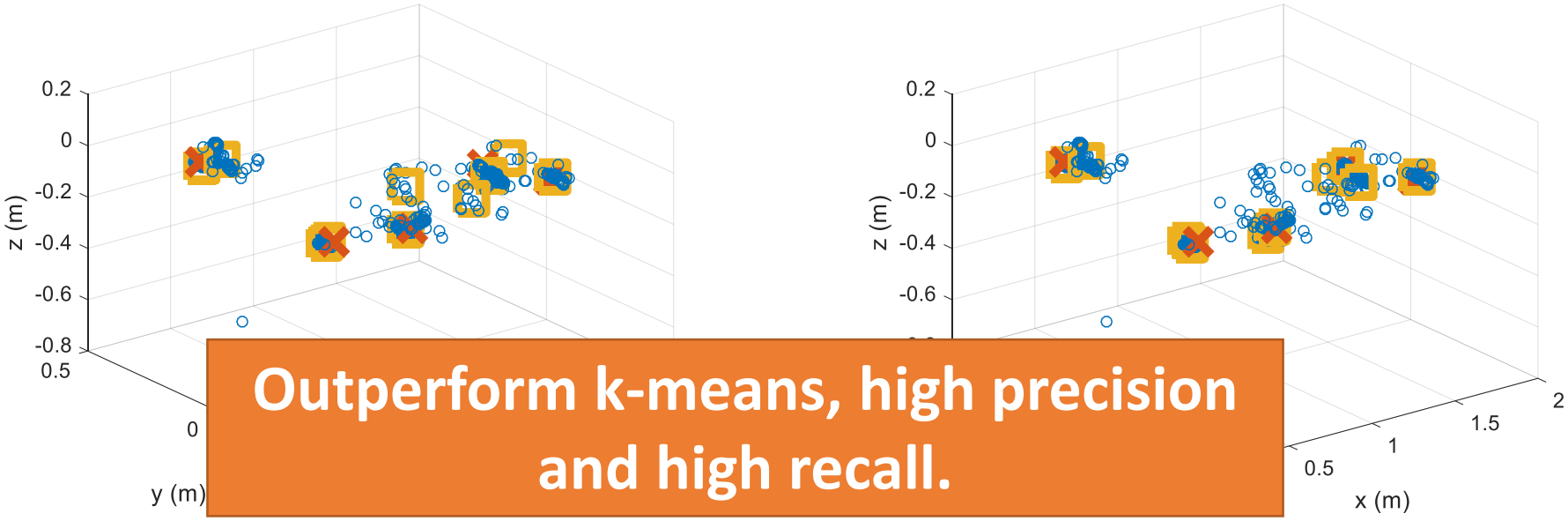
(a) kmeans



(b) spectral

Spectral Clustering

- Too much clusters



Outperform k-means, high precision and high recall.

(a) kmeans

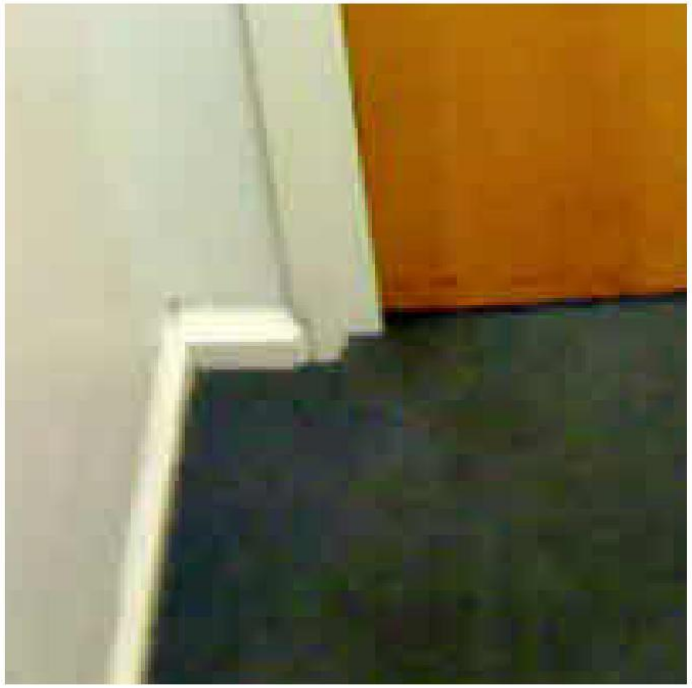
(b) spectral

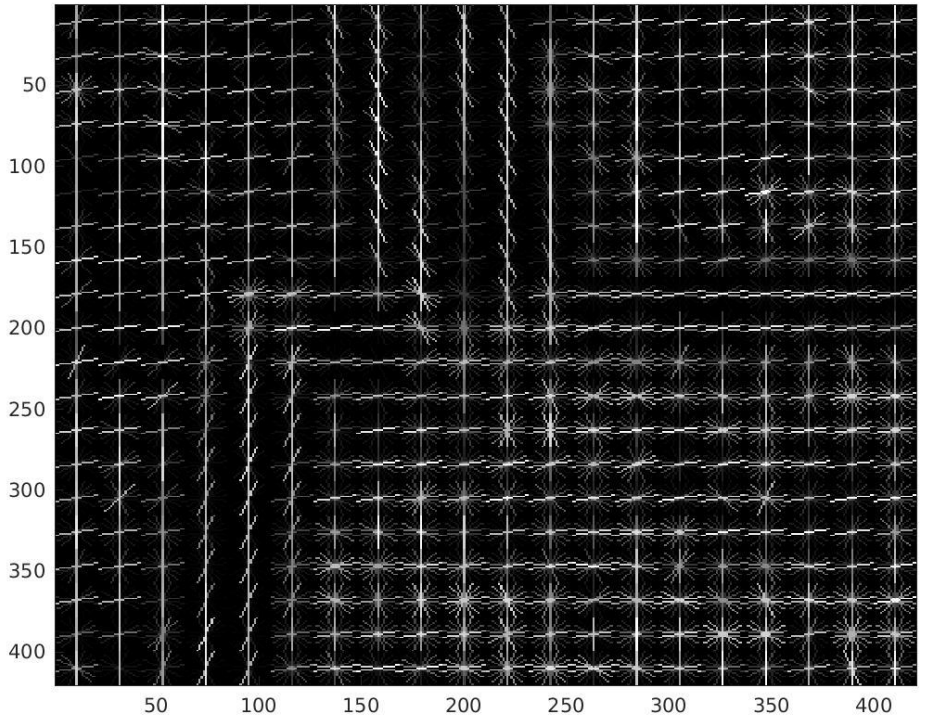
What is the Limitation of Gaze Positions

- Can we only use 3D gaze positions?
  - No, because of **moving** TRO
- How to solve this problem?
  - **Appearance**

Appearance

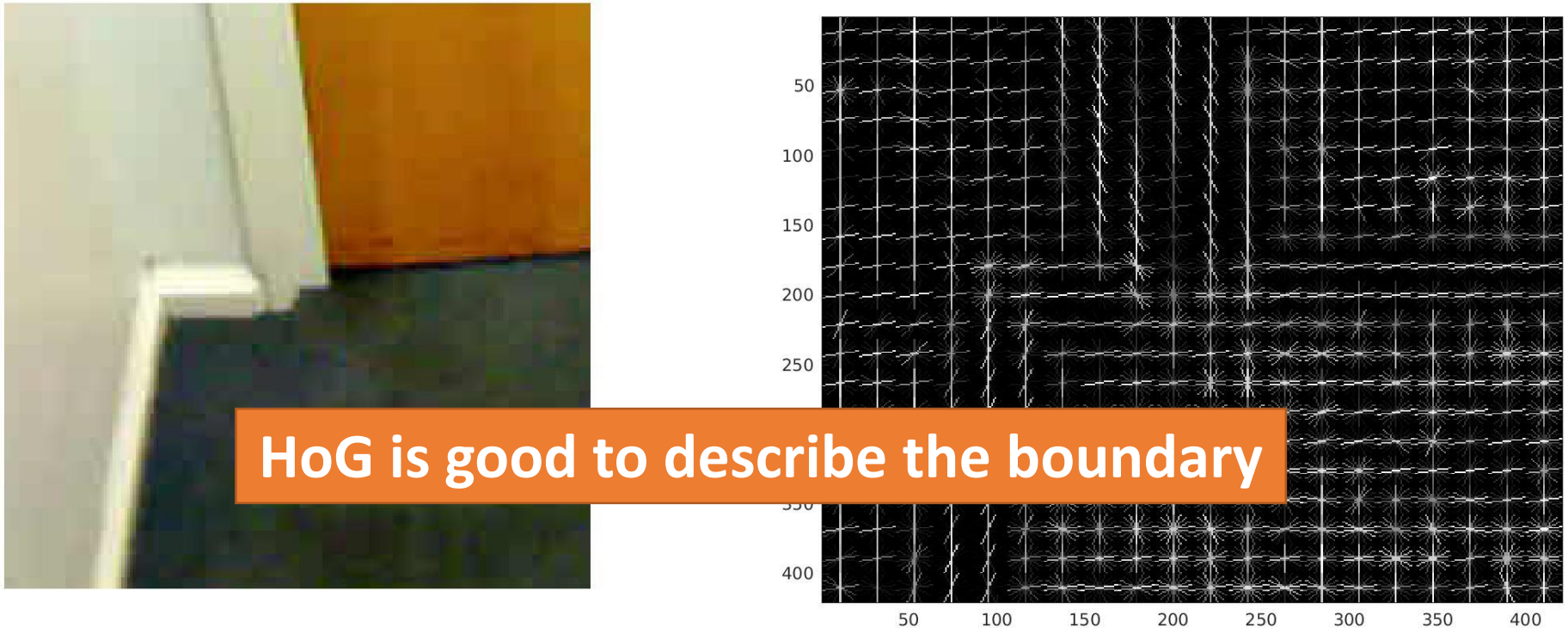
- How HoG features represent an image





Appearance

- How HoG features represent an image

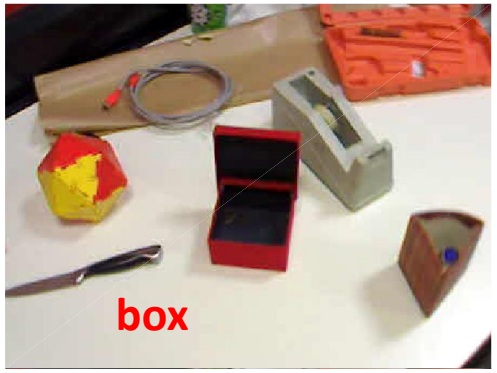


Identify TROs based on Appearance

- Extract HoG from the region near the gaze point for each frame
- Generate BoW representation for each frame
- Perform clustering on frames
- Use the frame ***closest to the center*** to represent each cluster
- Compare the appearance of center frames with the ground truth

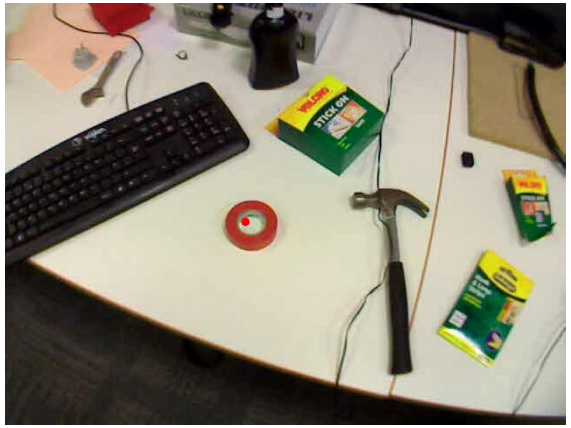
Appearance

- Five TROs around the desk

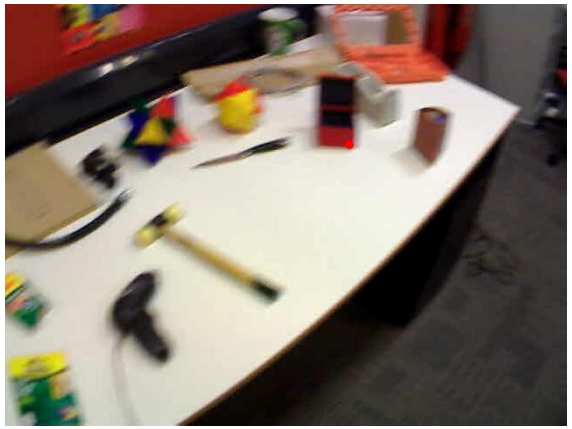


Results

Success (box)



Success (tape)



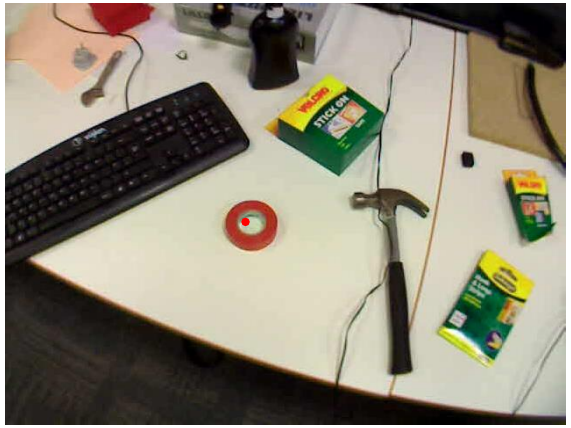
Duplicated (box)

Success (charger)

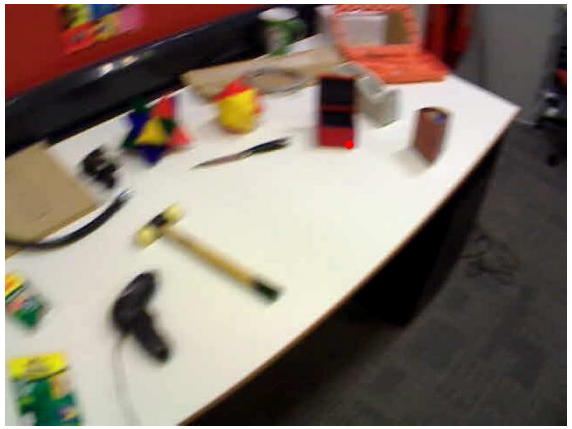
Failure

Results

Success (box)



Success (tape)



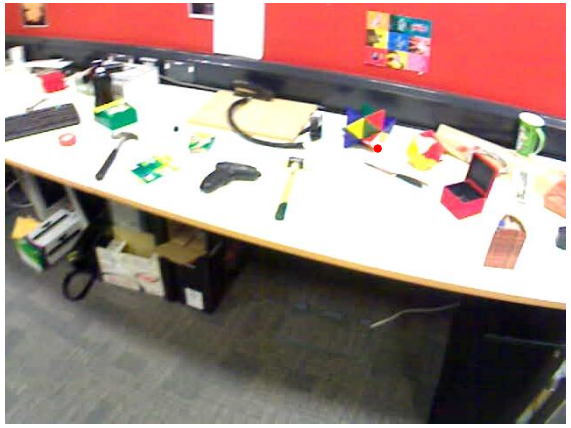
Duplicated (box)

Success (charger)

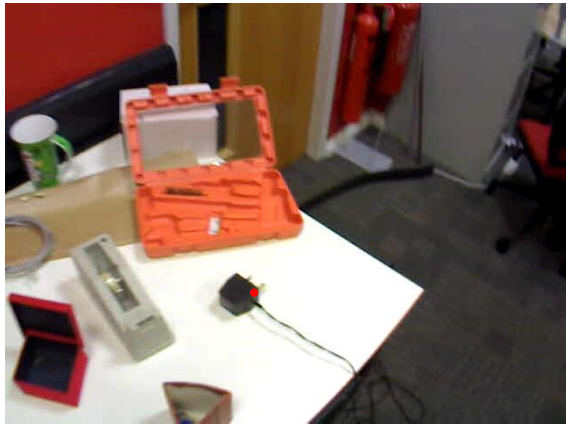
Failure

Missing two TROs, the appearance is not as effective as the position

Using Neighbor frames



Failure



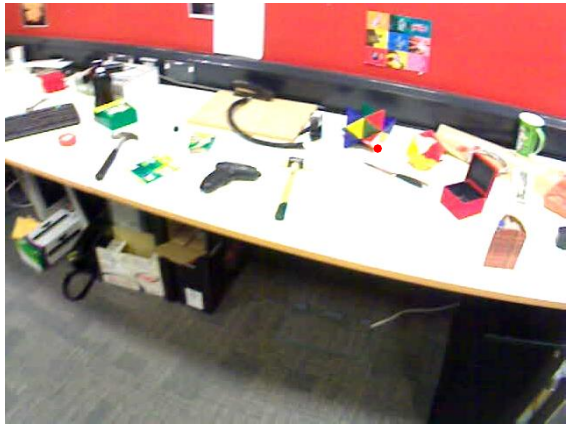
Success (charger)

Success (box)

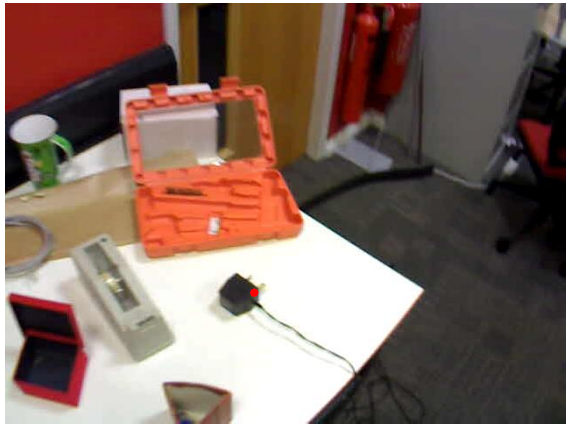
Success (driver)

Success (tape)

Using Neighbor frames

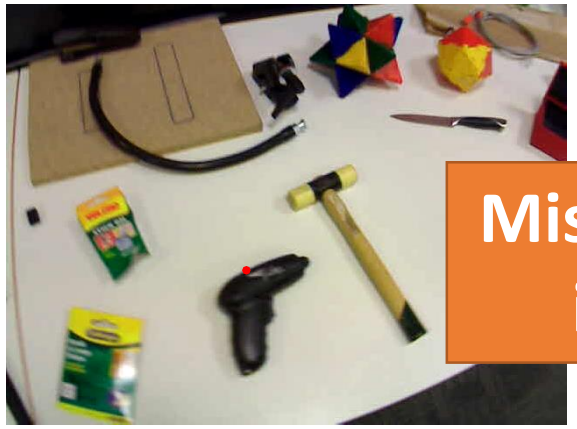


Failure



Success (charger)

Success (box)



Success (driver)

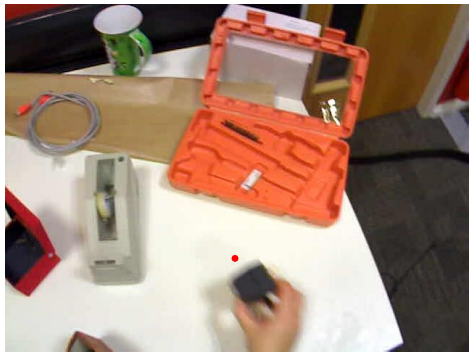
Success (tape)

Missing one TRO, using neighbor frames is helpful to improve performance

Over-Estimating No. of Clusters

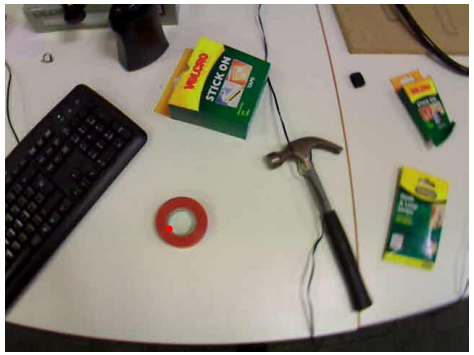
Success (driver)

Failure



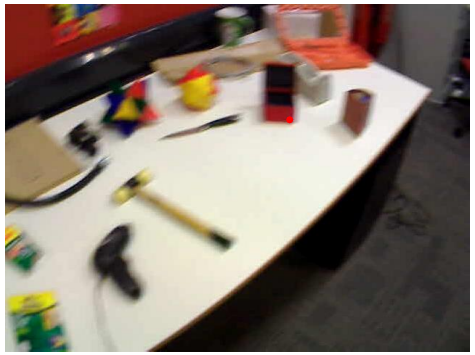
Success (charger)

Success (box)



Success (tape)

Duplicated (box)



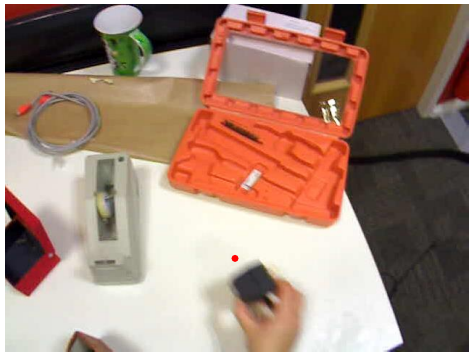
Duplicated (box)

Duplicated (driver)

Over-Estimating No. of Clusters

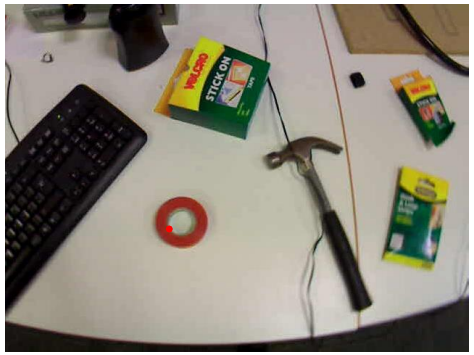
Success (driver)

Failure



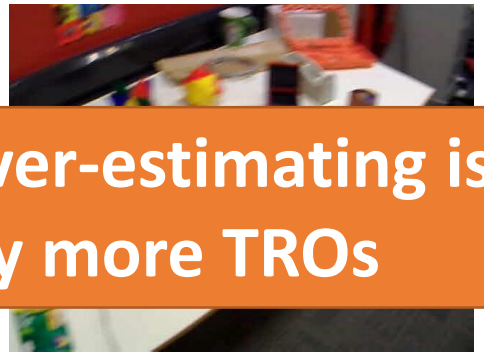
Success (charger)

Success (box)



Success (tape)

Duplicated (box)



Duplicated (box)

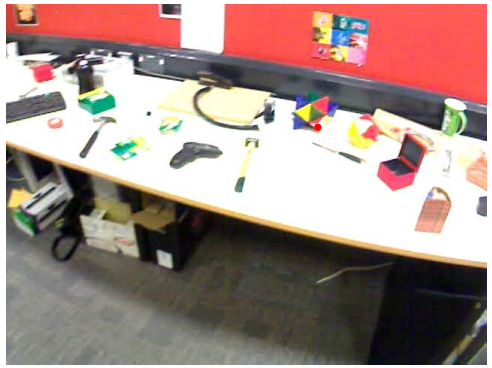
Duplicated (driver)

Missing one TROs, over-estimating is helpful to identify more TROs

Also Using Neighbor frames

Success (socket)

Success (tape)

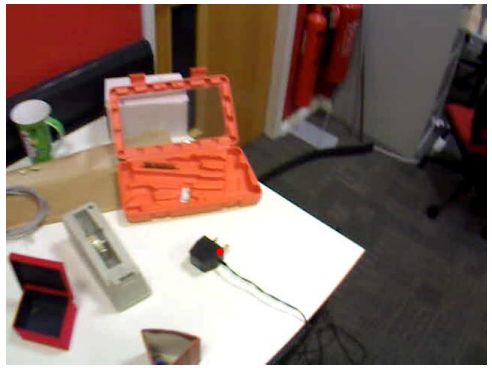


Failure

Duplicated (socket)

Success (driver)

Success (box)



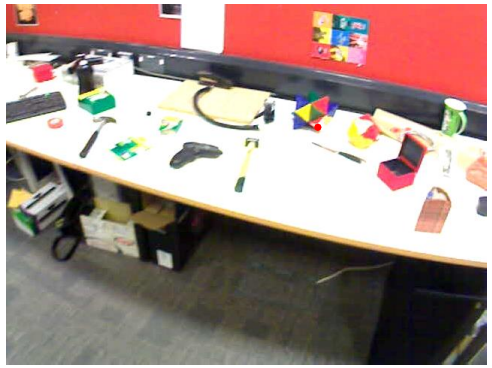
Success (charger)

Duplicated (box)

Also Using Neighbor frames

Success (socket)

Success (tape)



Failure

Duplicated (socket)

Success (driver)

Success (box)

Success (charger)

Duplicated (box)

Finding all TROs

Conclusion

- Gaze information is important and necessary for egocentric videos, and the center of image is not a good approximation
- Gaze fixation is helpful for identifying TROs, but itself is not enough
- 3D positions of gaze give rich information for TROs, but clustering method and the estimation on the number of TROs is critical
  - Use **spectral clustering** and do not worry about overestimating
- Appearance is another important feature for identifying TROs
  - Using **neighbor frames** is beneficial to improve performance
  - **Over-estimating** No. of TROs is helpful to reduce false negative