

2019 Intelligent Sensing Summer School (London) September, 2-6

## **The CORSMAL challenge**

Team #1

# Detection of unseen object through hand tracking and depth map segmentation

Luca Varotto

Raphael Leung

Paul Curuia Valentin

Centre for Intelligent Sensing

Queen Mary University of London

# Outline

---

- Initiative and goals
- Hand detection
  - Tracking
  - The setup
- Depth detection of hand
  - Results
  - Unseen object isolation
- Conclusion and final thoughts

# Our initiative and goals for the challenge

---

- We are aiming for accuracy and certainty:

We want to find a way for the robot to **certainly find** and **accurately grasp** the object no matter the **size, shape, colour or hold**.

Object  
Tracking

- Successful tracking of the hand
- Localisation of the object with the help of tracking
- Object size estimation

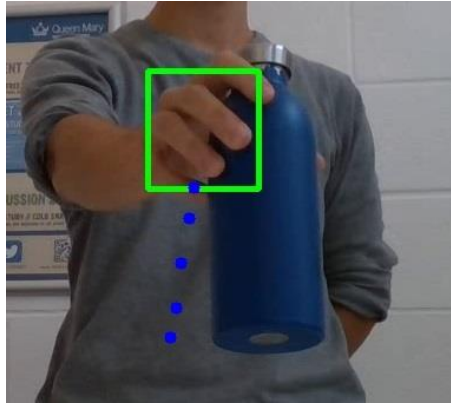
Object  
dimensions  
estimation

Grasp  
points  
estimation

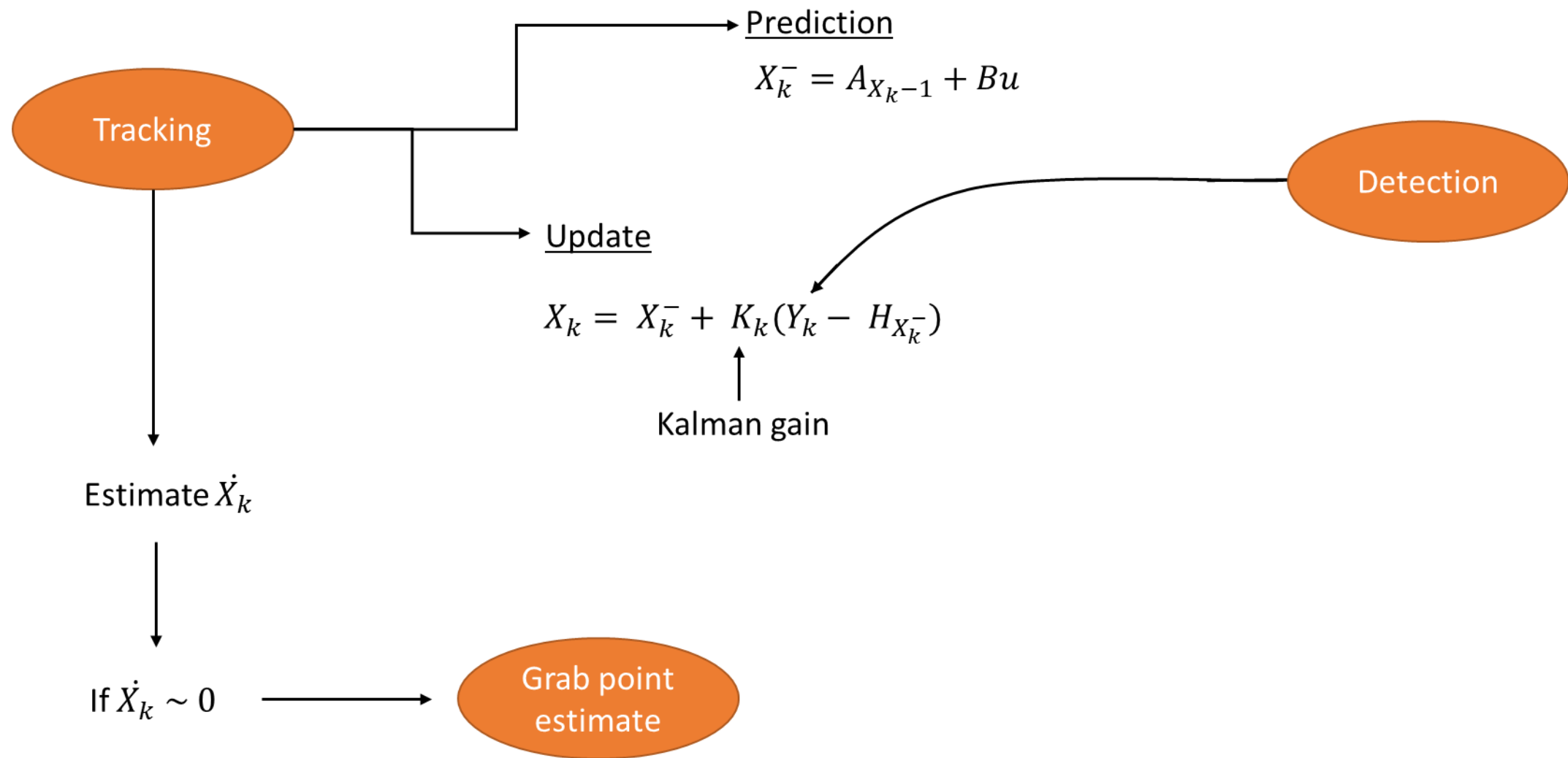
# Real-time Hand-Detector using Neural Networks (SSD) (Tensorflow Object Detection API)

---

- This model was pre-learned through transfer learning
- Tracks the hand, not the object
- This hand detector contains high quality, pixel level annotations where hands are located across 4800 images. All images are captured from an egocentric view (Google glass) across 48 different environments (indoor, outdoor) and activities (playing cards, chess, jenga, solving puzzles etc).

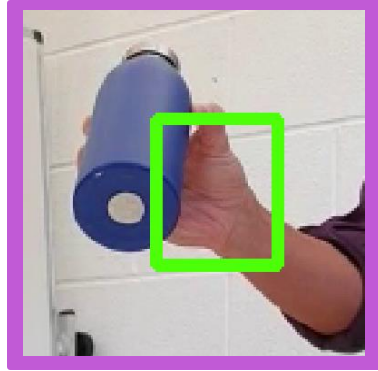


- we've implemented a Kalman filter (blue dot) for tracking

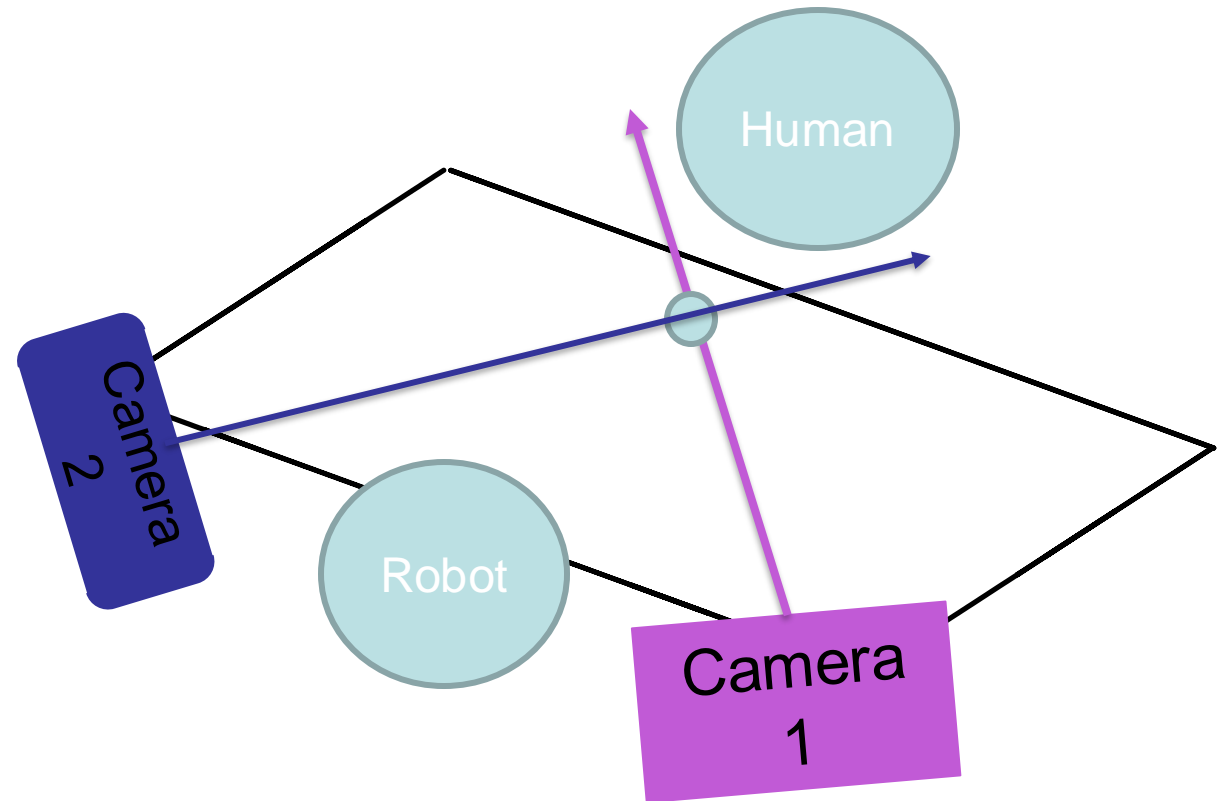
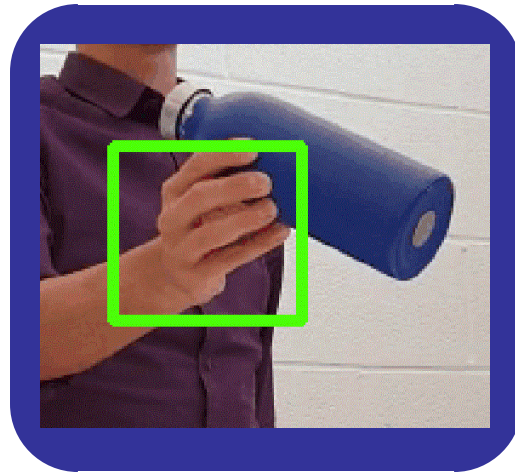


## The setup

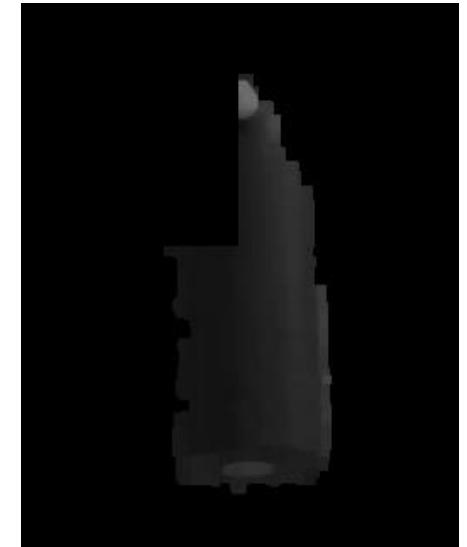
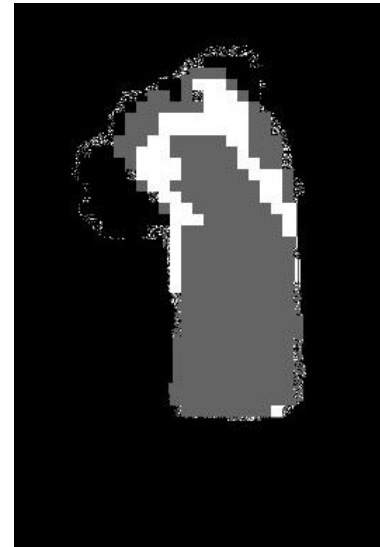
- The pipeline is implemented in only one of the cameras



- But there will be situations where the hand detection isn't possible

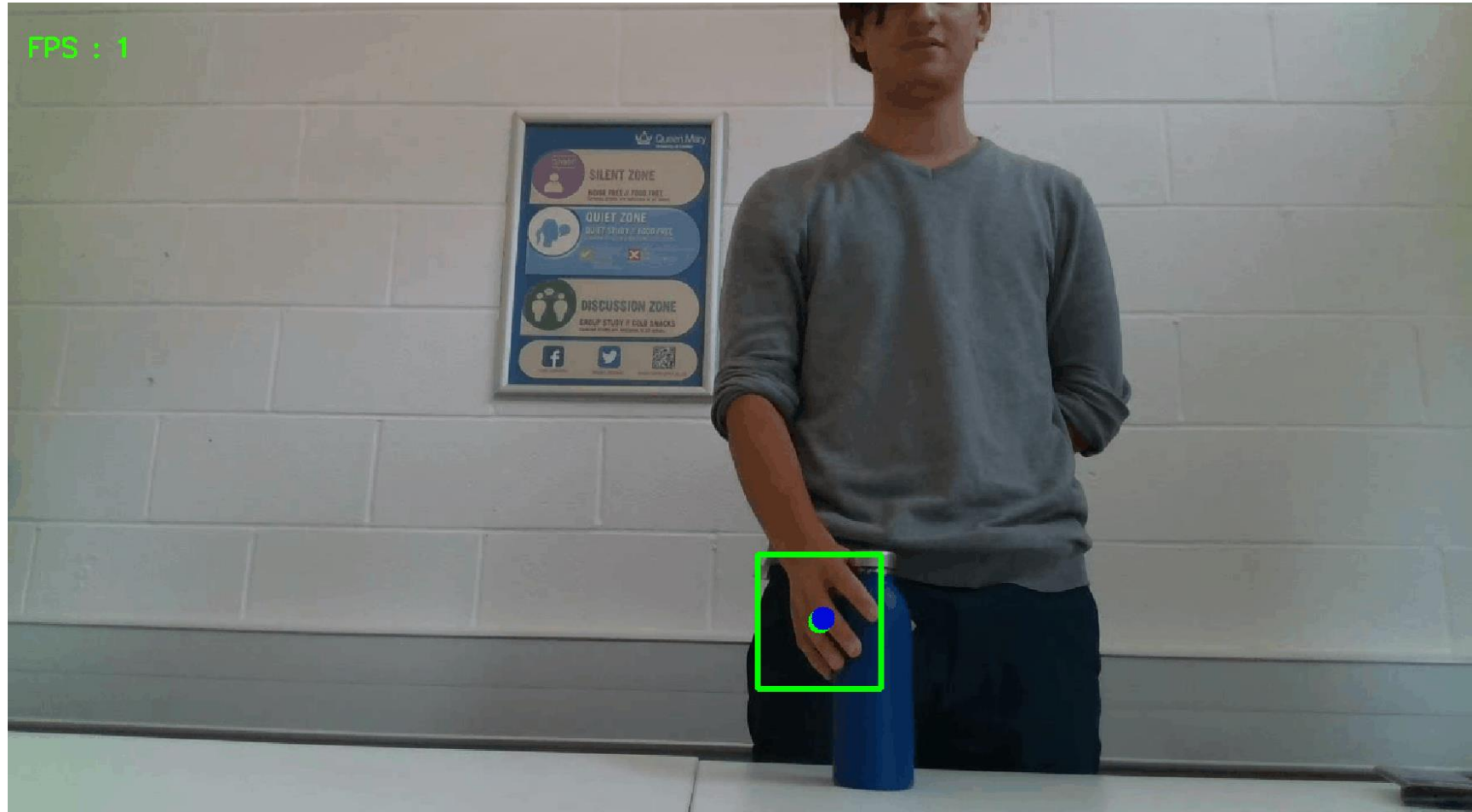


## Combining the depth camera video with hand tracking



# Video result blue bottle

---

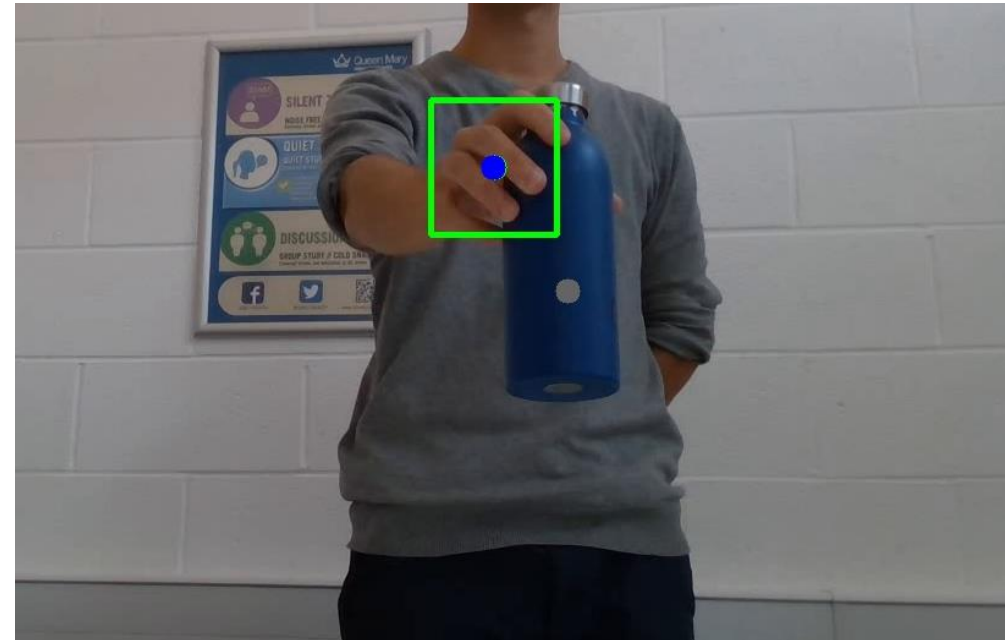
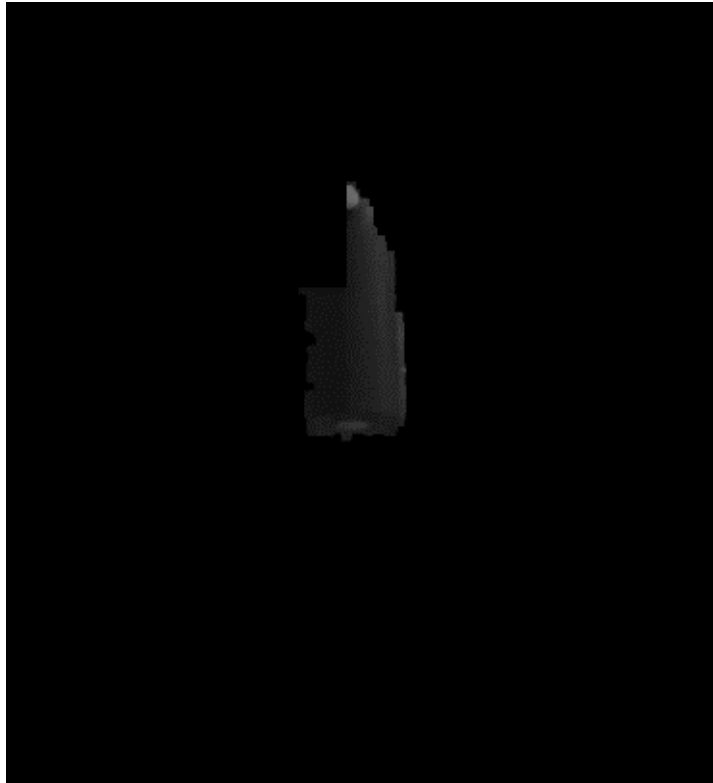


**CORSMAL**  
Collaborative object recognition,  
shared manipulation and learning



# Video result blue bottle

---

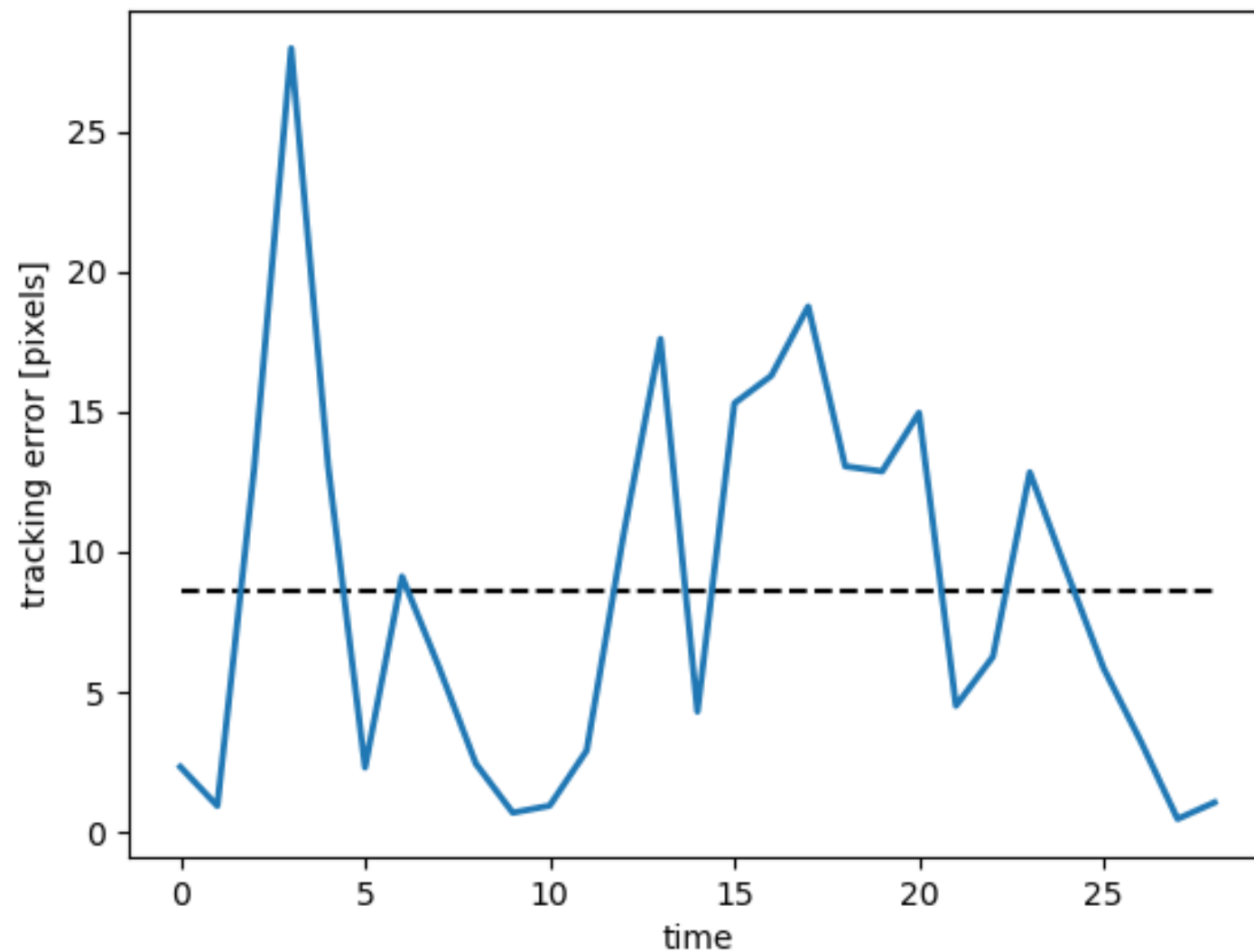


Grasp point found !



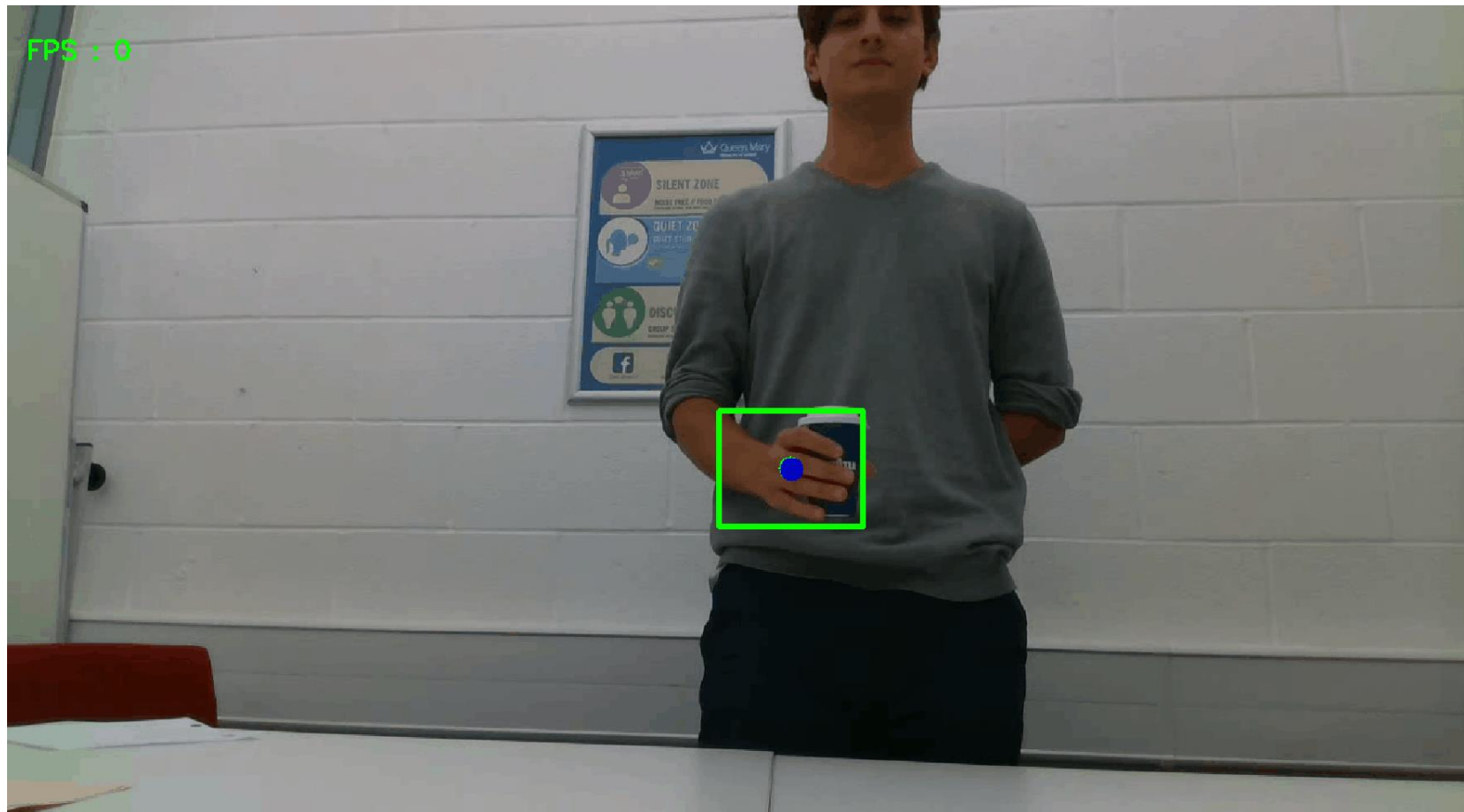
**CORSMAL**  
Collaborative object recognition,  
shared manipulation and learning

# Video result blue bottle - Graph



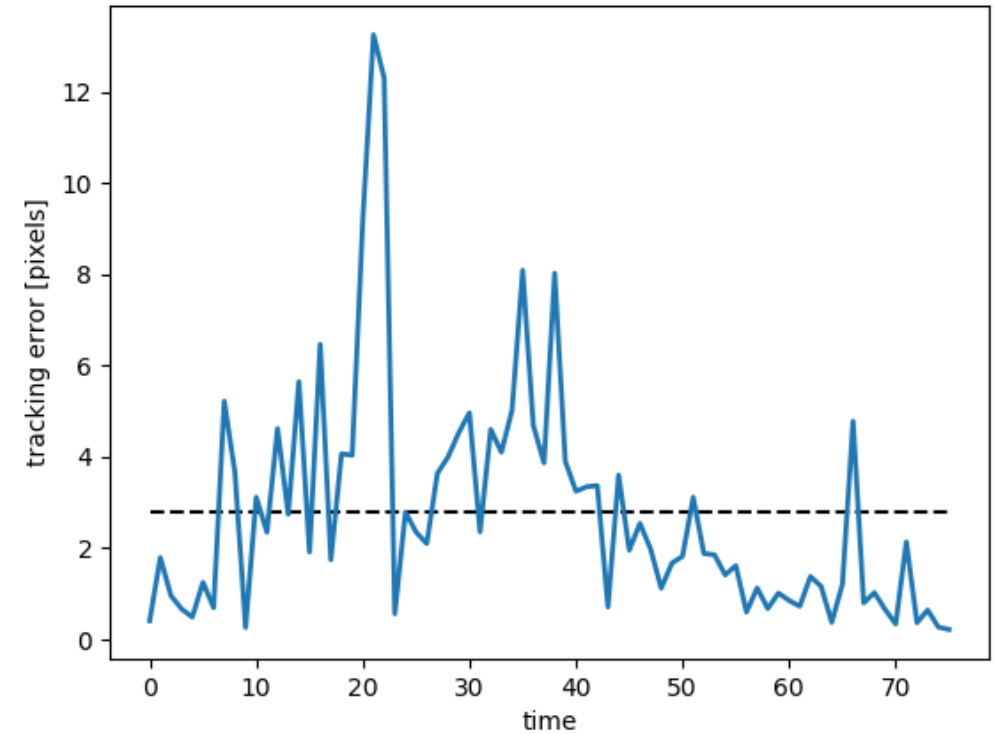
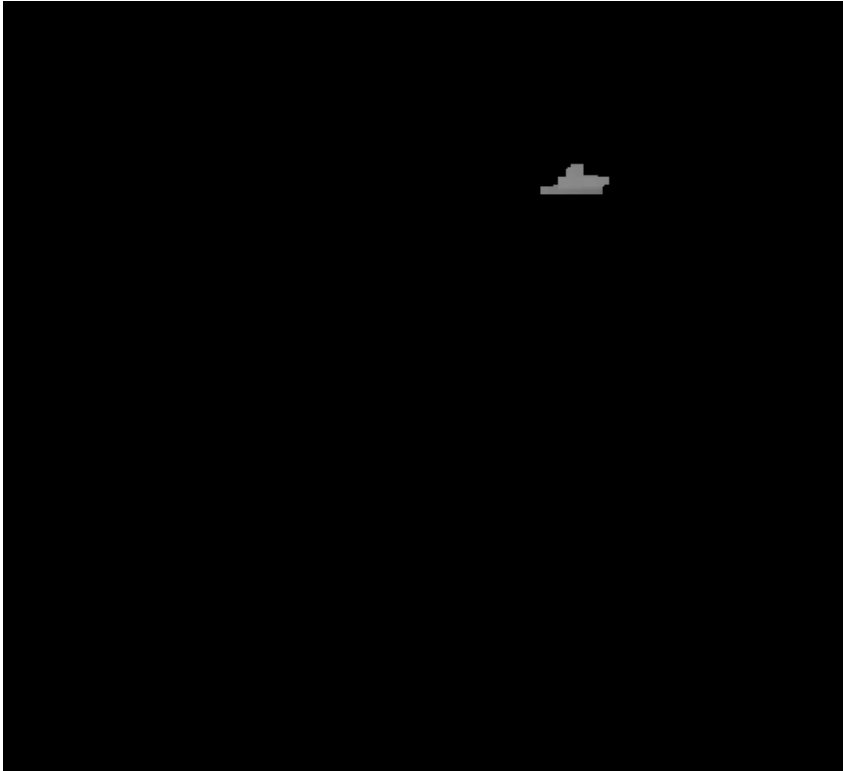
# Video result coffee cup 1

---



**CORSMAL**  
Collaborative object recognition,  
shared manipulation and learning

# Video result coffee cup 1



Grasp point not found!



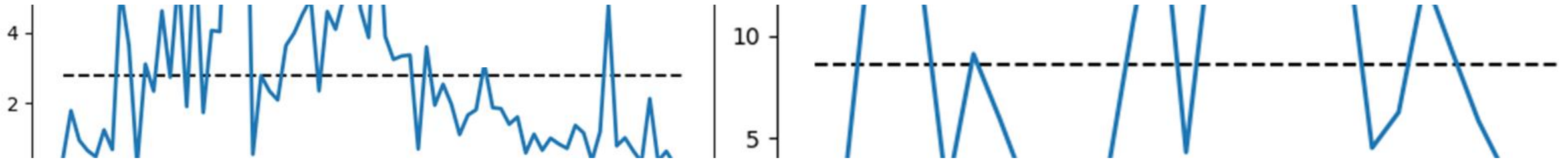
CORSMAL

Collaborative object recognition,  
shared manipulation and learning

# Performance

---

- SSD detector has a mAP of 0.9686
- Performance of Kalman shows a tracking errors average of under 10 to 3 pixels



- Estimation of grasping point is 100% correct

# Conclusions

---

- Successful hand detection & tracking
- Successful hand & bottle isolation → estimation of grasping point

## Future works:

- If we had more time we would improve:
  - Object dimension estimate
  - Refinement of the grasping moment estimation by looking at depth between person and object

# References

---

- For the hand detector:

<https://github.com/victordibia/handtracking?fbclid=IwAR1gvVLpy8EoHxdyrZP3QujrgV3FJ4lipgT76fx6zTdYD0FQdUd6waoG5ds> By Victor Dibia

- .