chist-era

follow us
@corsmal

CORSMAL

Collaborative Object Recognition, Shared Manipulation And Learning

ORMR (2019-2022) - Object Recognition and Manipulation by Robots: data sharing and experiment reproducibility

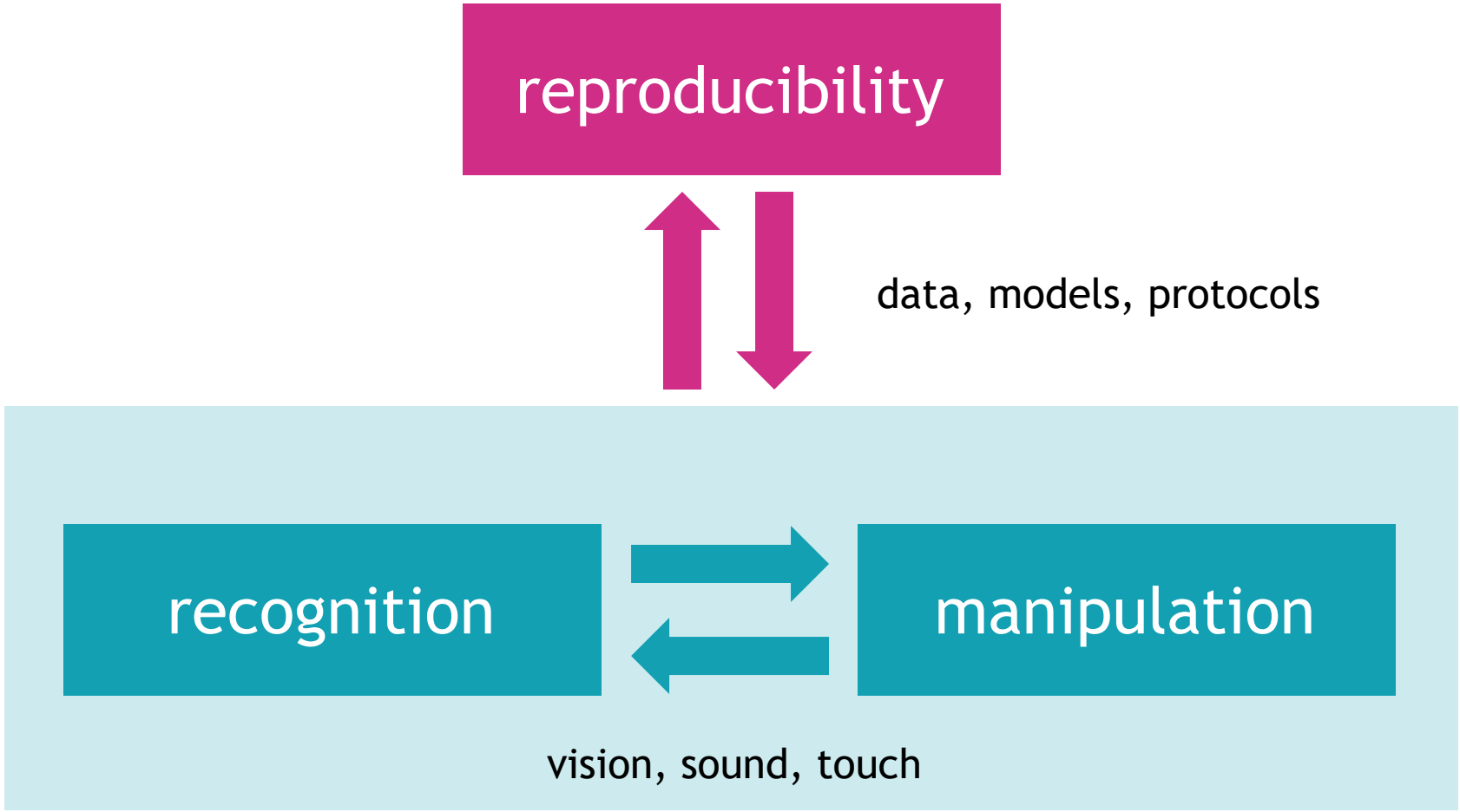
CHIST-ERA Conference 2020

Queen Mary University of London

SORBONNE UNIVERSITÉ

EPFL

Scope


reproducibility
data, models, protocols
recognition
manipulation
vision, sound, touch

Aims

- To create an open dataset and an evaluation protocol for the recognition and manipulation of previously unseen object instances

- To explore the fusion of multiple sensing modalities (touch + sound + vision) to accurately and robustly estimate the physical properties of objects in noisy and potentially ambiguous environments

The task

Human grasping — Human manipulation — Human-robot handover — Robot manipulation — Robot delivery


CORSMAL
Collaborative object recognition,
shared manipulation and learning

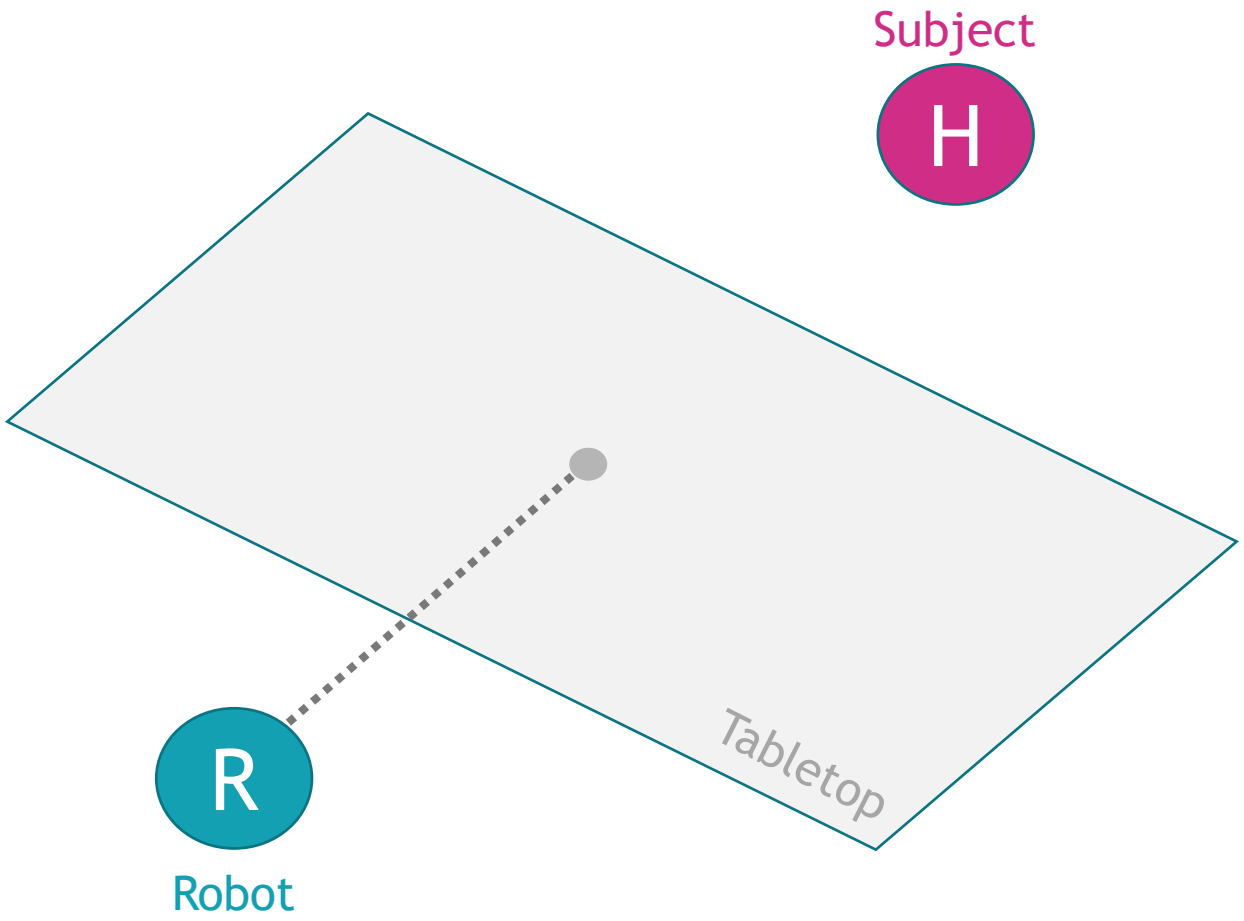
Challenges

Previously unseen object instance (shape, material, mass, and filling variability)

Variability in grasp type and handover location

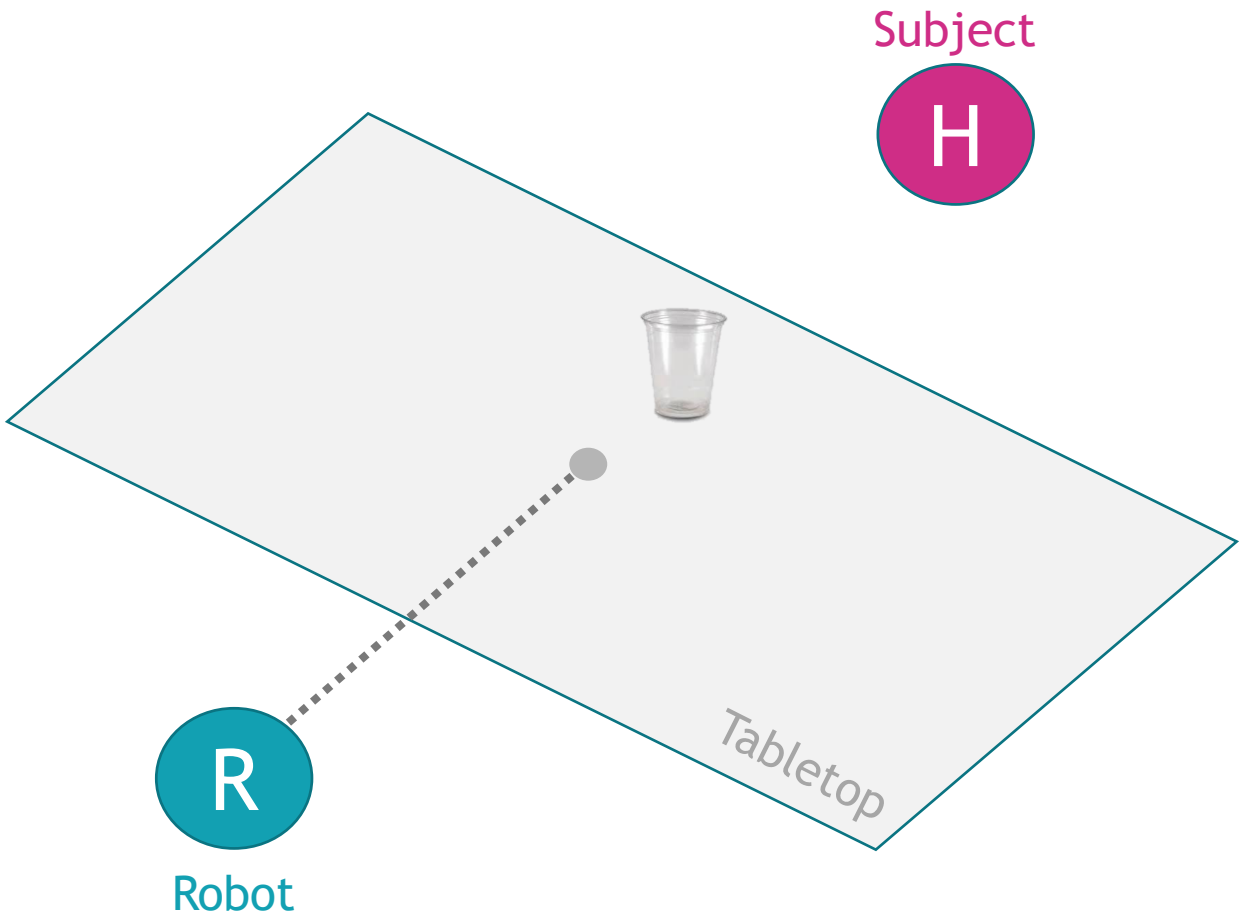
The setup

*The robot should be located so that the tabletop centre is at 40%-50% of the robot reachability



*Initial cup location at a pre-define location

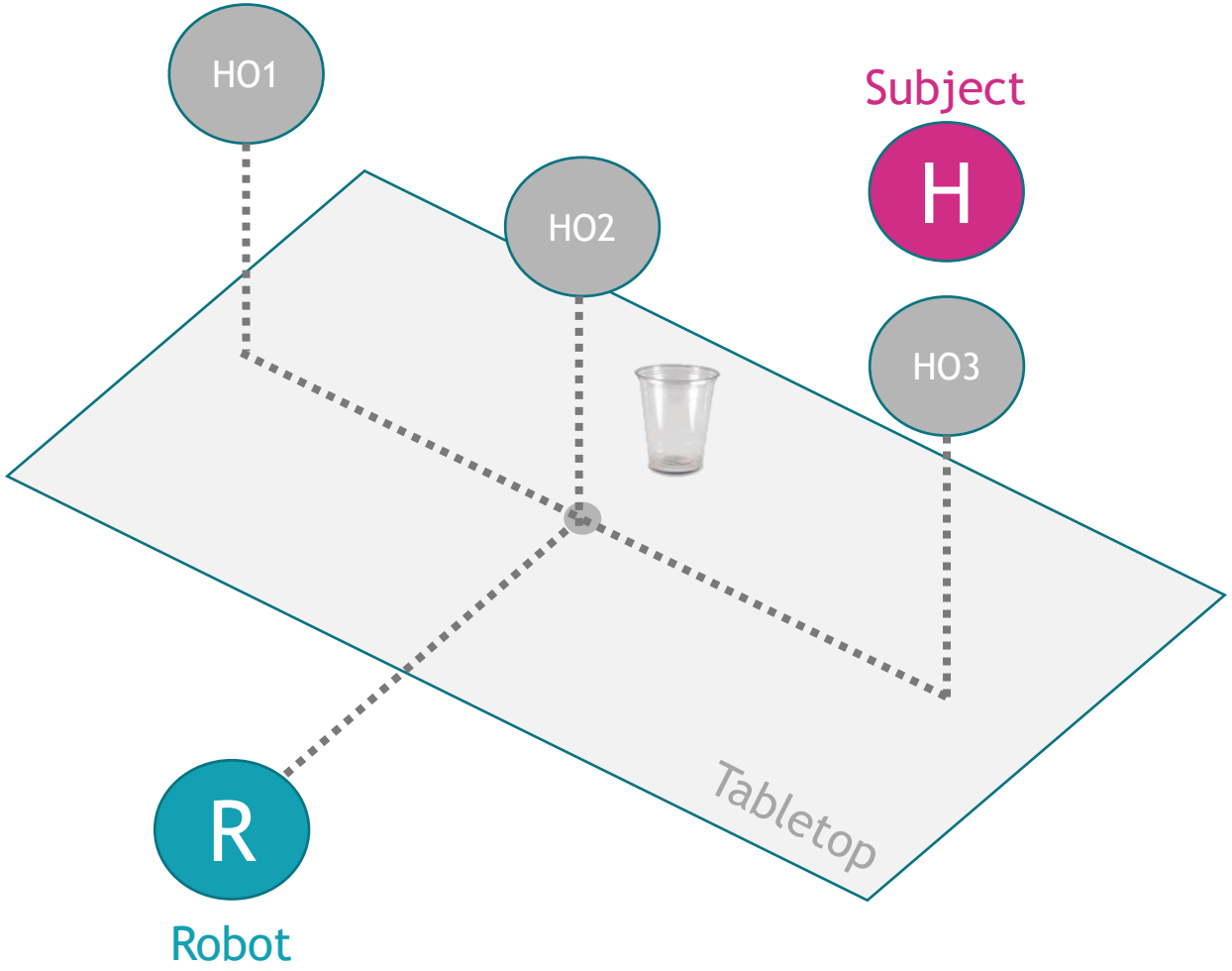
The setup



CORSMAL
Collaborative object recognition, shared manipulation and learning

The setup

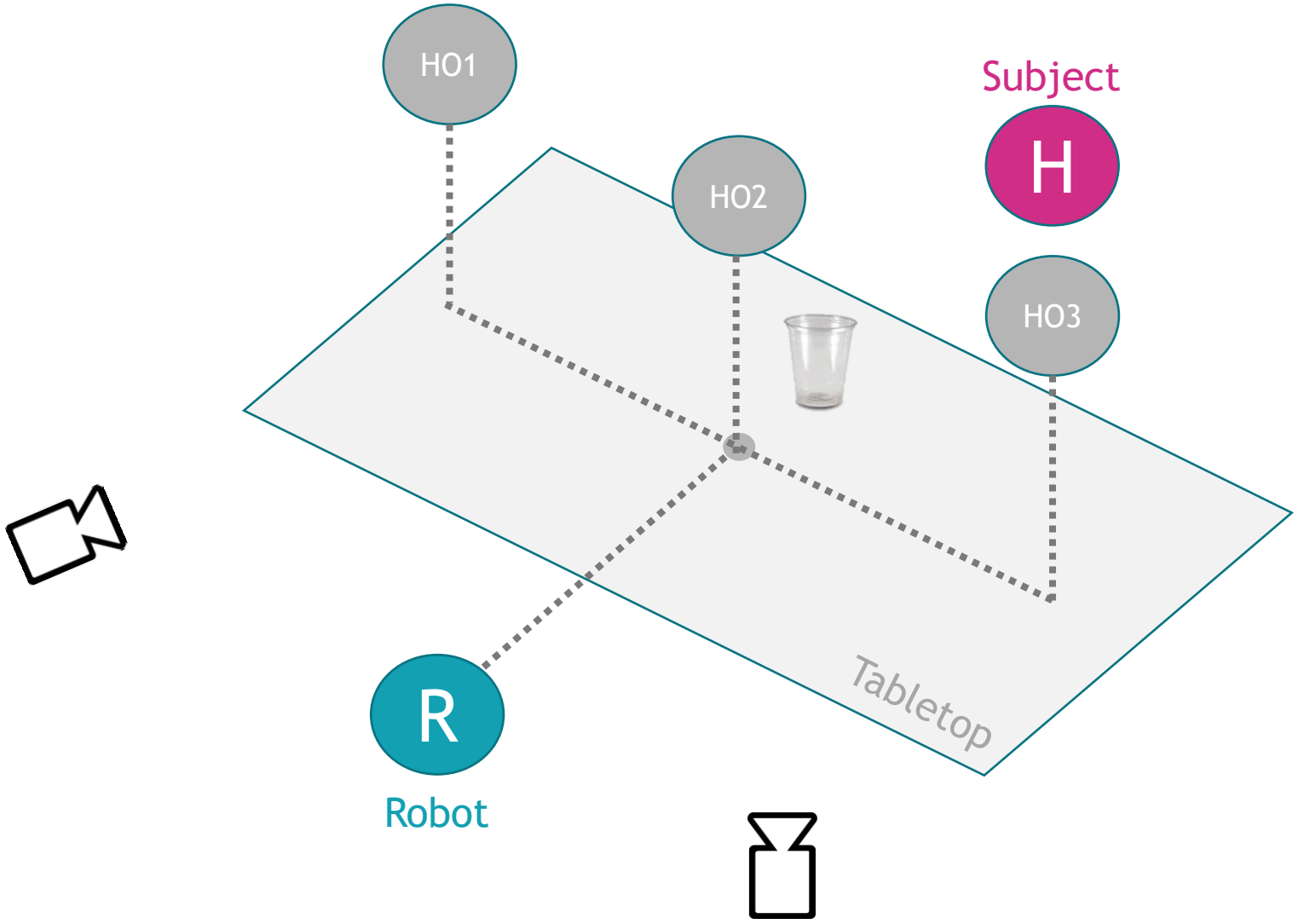
HOx Handover location x



CORSMAL
Collaborative object recognition, shared manipulation and learning

The setup

Sensing:
- up to two cameras
- [optional] force sensors
- [optional] tactile sensors
- [optional] proximity sensors



CORSMAL
Collaborative object recognition, shared manipulation and learning

The objects

	Cup 1	Cup 2	Cup 3	Cup 4
				
Deformability	High	Medium	Medium	None
Transparency	Medium	Low	High	High

Grasp types

	Cup 1	Cup 2	Cup 3	Cup 4
Grasp 1				
Grasp 2				
Grasp 3				

The CORSMAL benchmark

Objects: 4 cups with different transparencies and deformabilities

Filling: empty or 90% (rice)

Human subjects: 4

Human grasp types: bottom of cup, top of cup, natural (i.e. unconstrained)

Handover locations: in front, front-left, or front-right of robot

Total unique configurations: 4 x 2 x 4 x 3 x 3 = 288

Benchmark for Human-to-Robot Handovers of Unseen Containers with Unknown Filling
Sanchez-Matilla, Chatzilygeroudis, Modas, Ferreira Duarte, Xompero, Frossard, Billard, Cavallaro
IEEE Robotics and Automation Letters (RA-L), vol. 5, no. 2, Apr. 2020

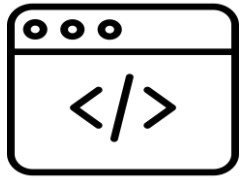
Evaluation scores

Vision scores	Robotic scores	Global scores
Object dimensions Object fullness Object mass	Human-hand pose End-effector pose Object mass	Delivery location Residual filling Manoeuvring time

Benchmark for Human-to-Robot Handovers of Unseen Containers with Unknown Filling
Sanchez-Matilla, Chatzilygeroudis, Modas, Ferreira Duarte, Xompero, Frossard, Billard, Cavallaro
IEEE Robotics and Automation Letters (RA-L), vol. 5, no. 2, Apr. 2020

Benchmark: resources

http://corsmal.eecs.qmul.ac.uk/benchmark.html

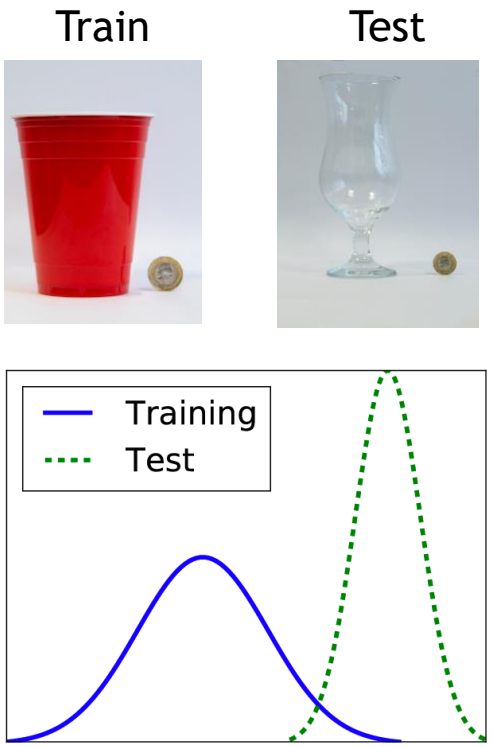


https://github.com/CORSMAL/Benchmark

CORSMAL
Collaborative object recognition, shared manipulation and learning

Robustness of classifiers

- Objective
  - train on a constrained set → generalize in the physical world
  - however: train & deployment data distribution shift
    => poor generalization performance
- Identified problem
  - deep learning models may learn spurious correlations
- Solution
  - to eliminate unnecessary features that might cause such correlations
  - to mimic the human visual system: filter high frequencies in natural images (no need for data augmentation!) to improve robustness to specific distribution shifts



Redundant features can hurt robustness to distribution shift
Ortiz Jimenez, Modas, Moosavi Dezfooli, Frossard
Robustness & Uncertainty in Deep Learning Workshop, ICML, July 2020

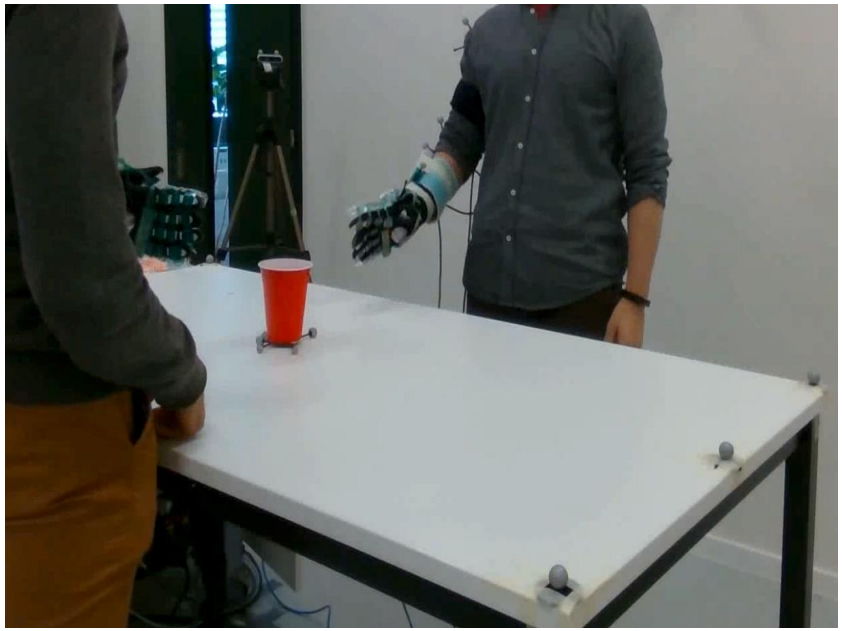
From human action understanding to robot action execution: how the physical properties of handled objects modulate non-verbal cues

Ferreira Duarte, Chatzilygeroudis, Santos-Victor, Billard

ICDL-EpiRob2020, October 2020

Human manipulation

- Actions
  - grasping from a table
  - manipulation
  - placing on the table
- Data (3 hours, 6 pairs of human subjects)
  - force sensors on the hand
  - position & orientation of shoulder, elbow, wrist
  - egocentric camera
  - two external RGB cameras
  - head-mounted eye-tracker (on one participant)
- Containers
  - empty
  - full (90% water)

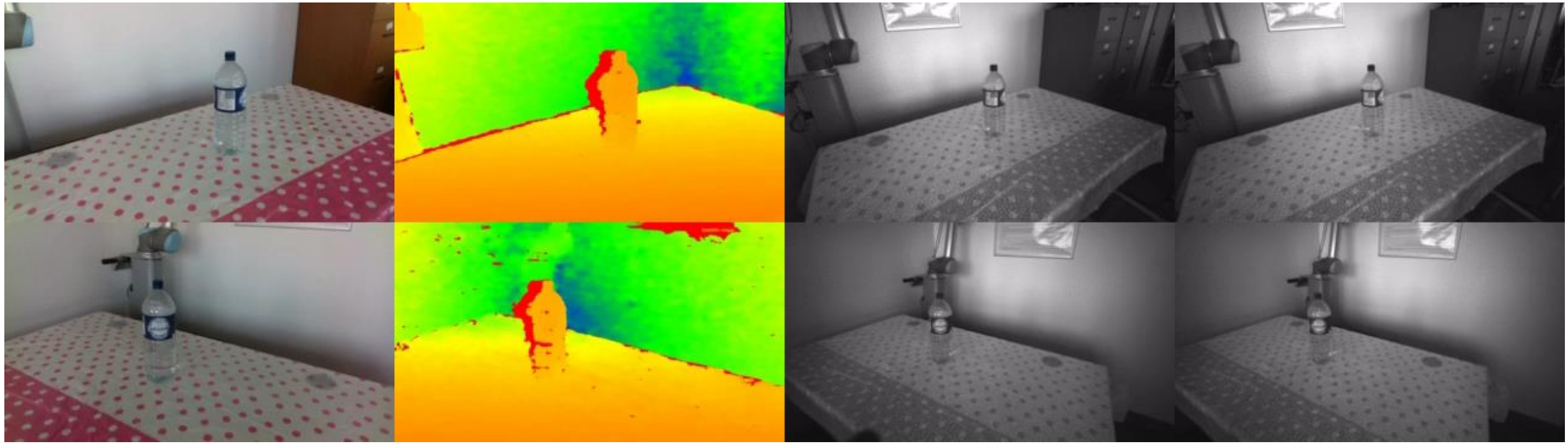


CORSMAL Containers dataset

CORSMAL
Collaborative object recognition, shared manipulation and learning

CORSMAL Containers dataset

- Data: RGB, depth, stereo infrared
- Containers: cups, drinking glasses, bottles
- Varying physical properties: material, texture, transparency, shape



CORSMAL Containers dataset

Objects: 23 containers for liquids with different transparencies, shapes, materials

2 setups:

- office with natural light
- studio-like room with no windows

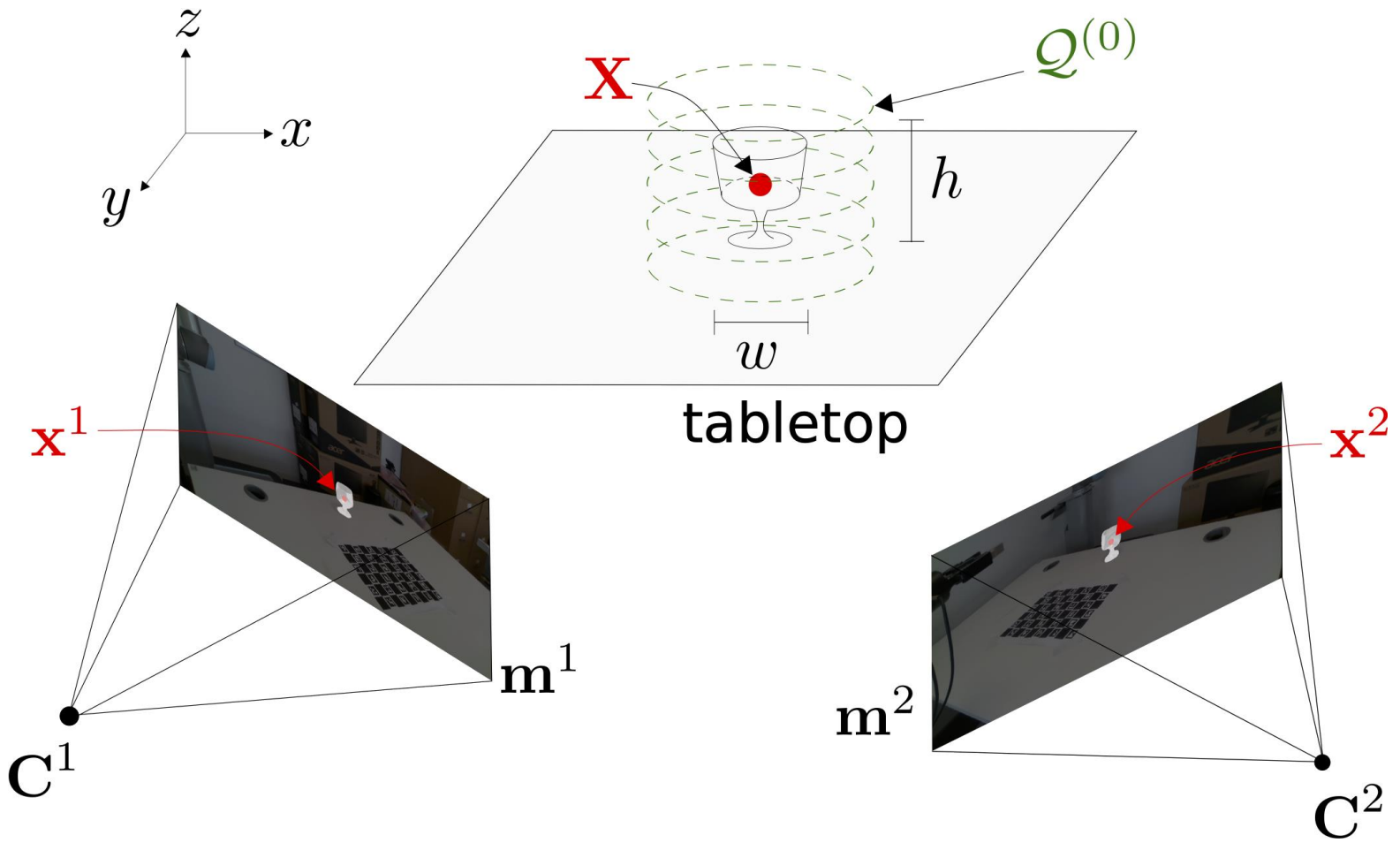
Configurations: (23) objects x (3) background x (3) illuminations = 207

Images: 1656 (414 RGB + 414 depth + 828 IR)

Calibrated cameras

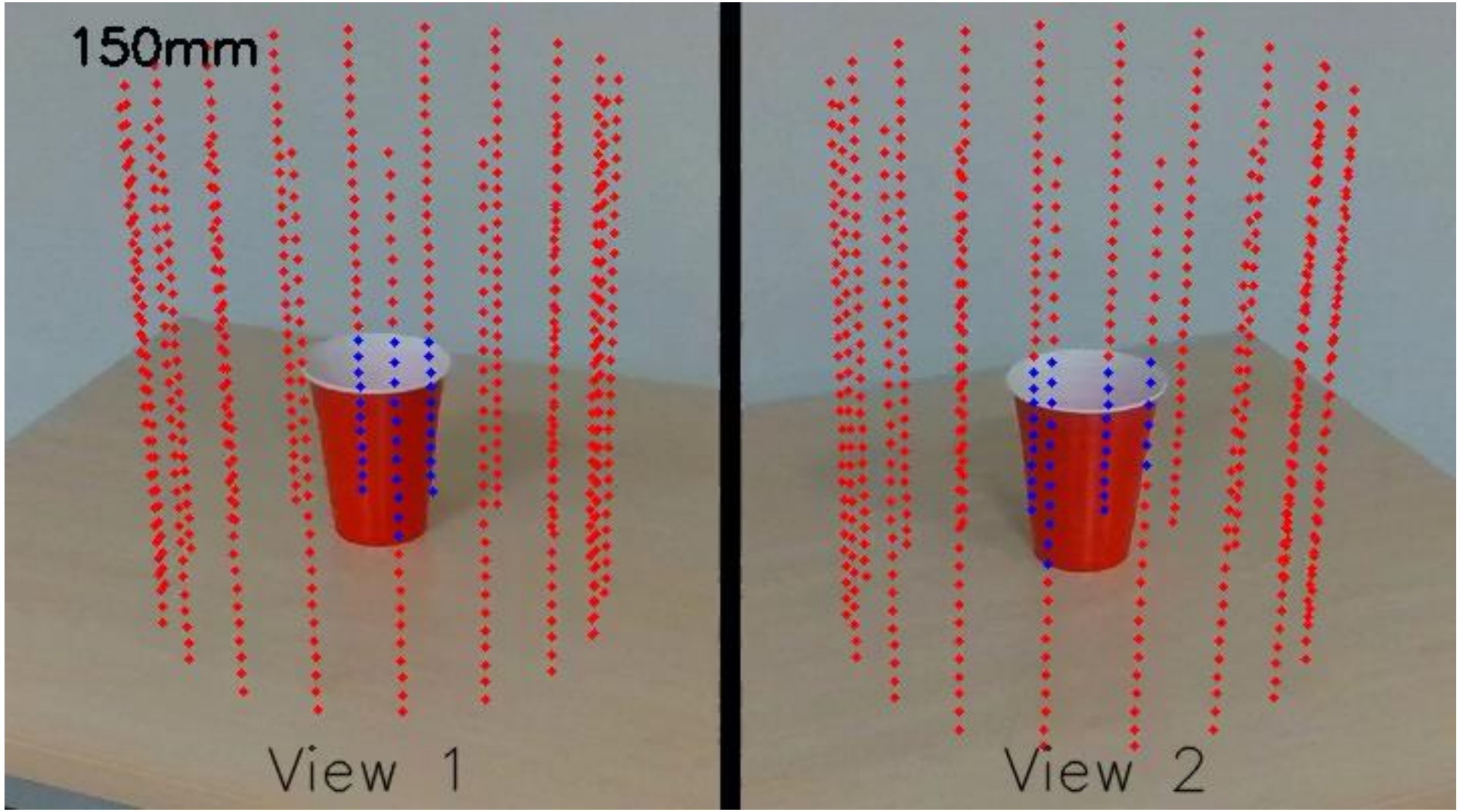
Annotation of the dimensions (width and height) of each object

LoDE: Localisalition and object Dimension Estimator

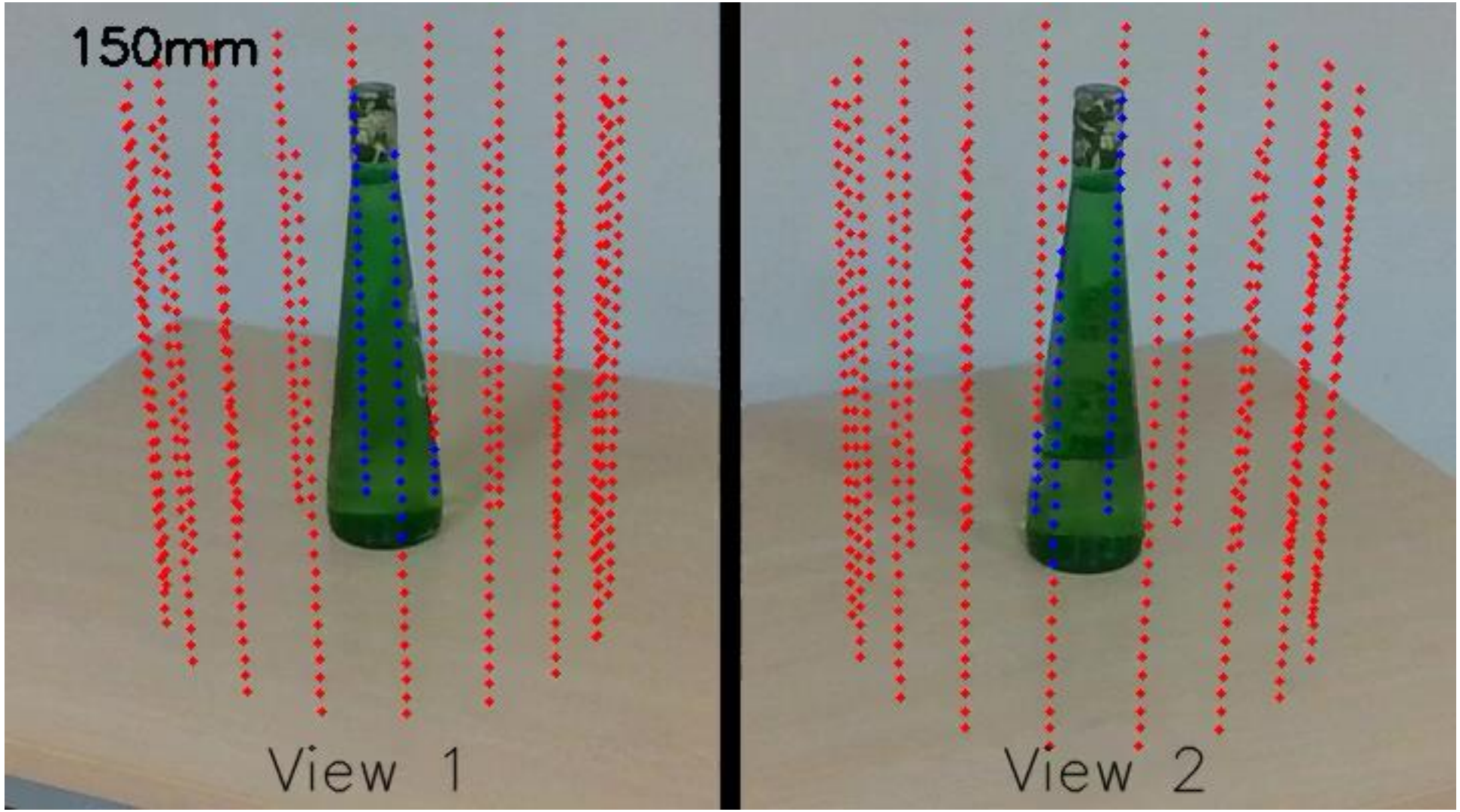


Multi-view shape estimation of transparent containers
Xompero, Sanchez-Matilla, Modas, Frossard, Cavallaro
IEEE ICASSP, May 2020

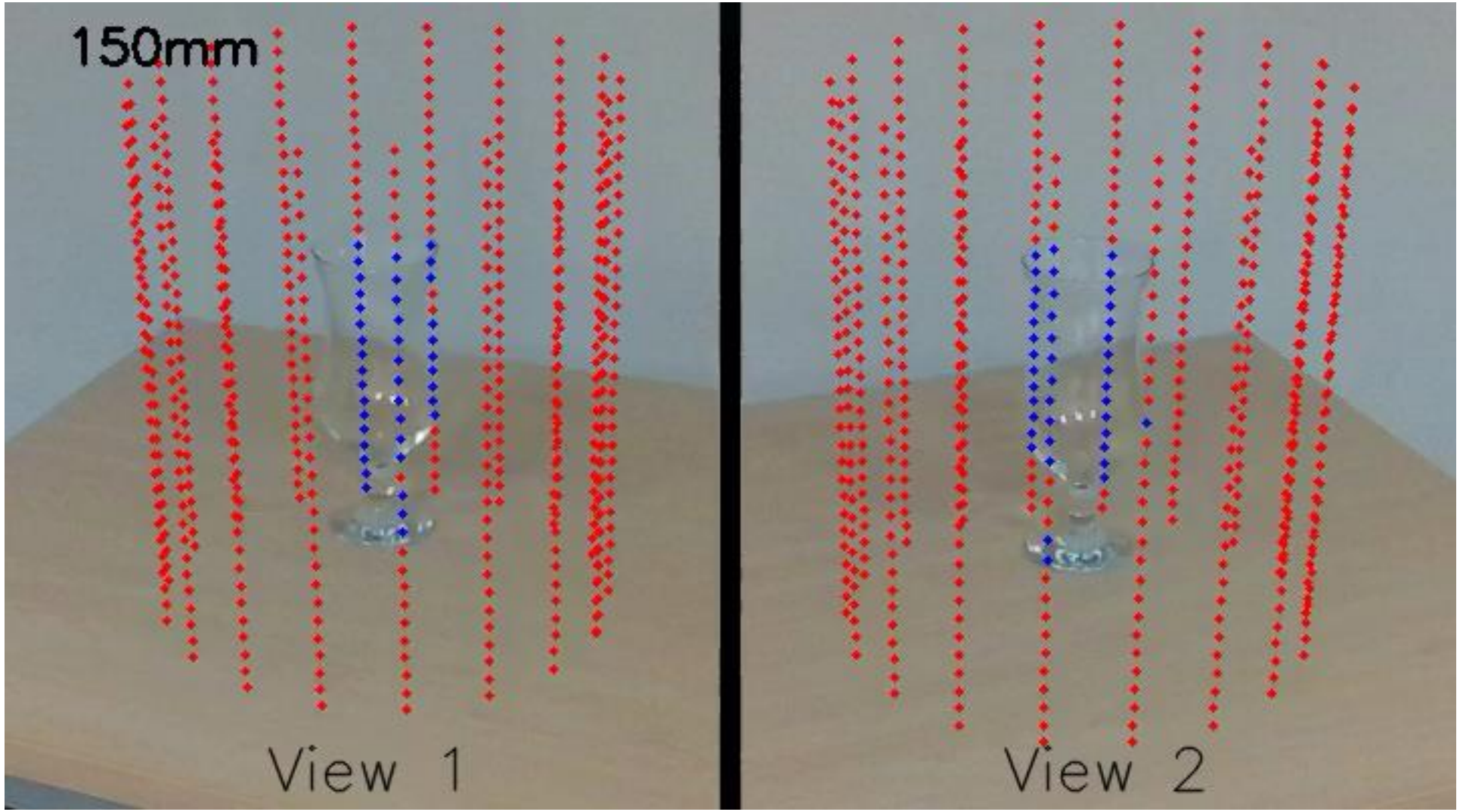
Iterative multi-view 3D-2D shape fitting



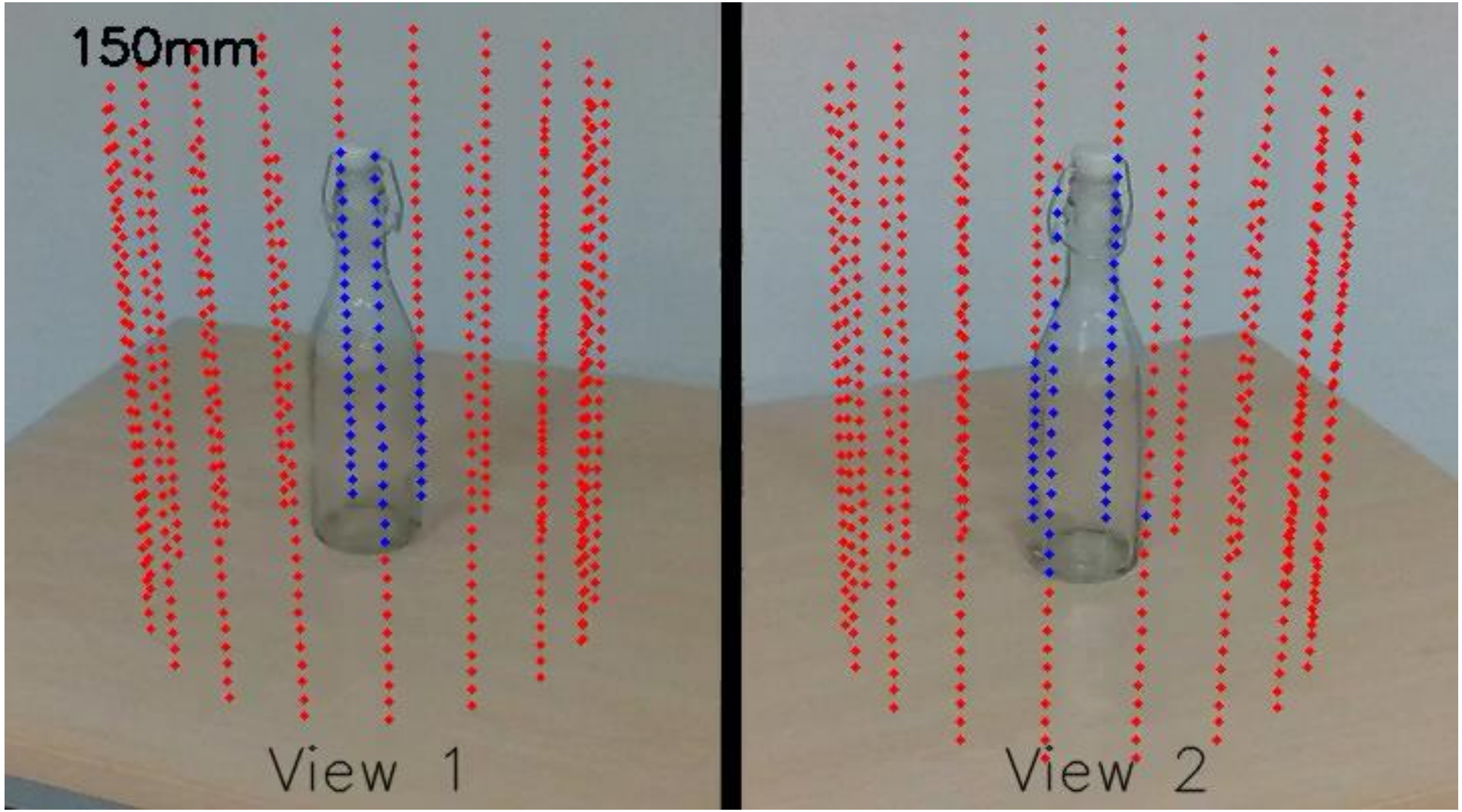
Iterative multi-view 3D-2D shape fitting



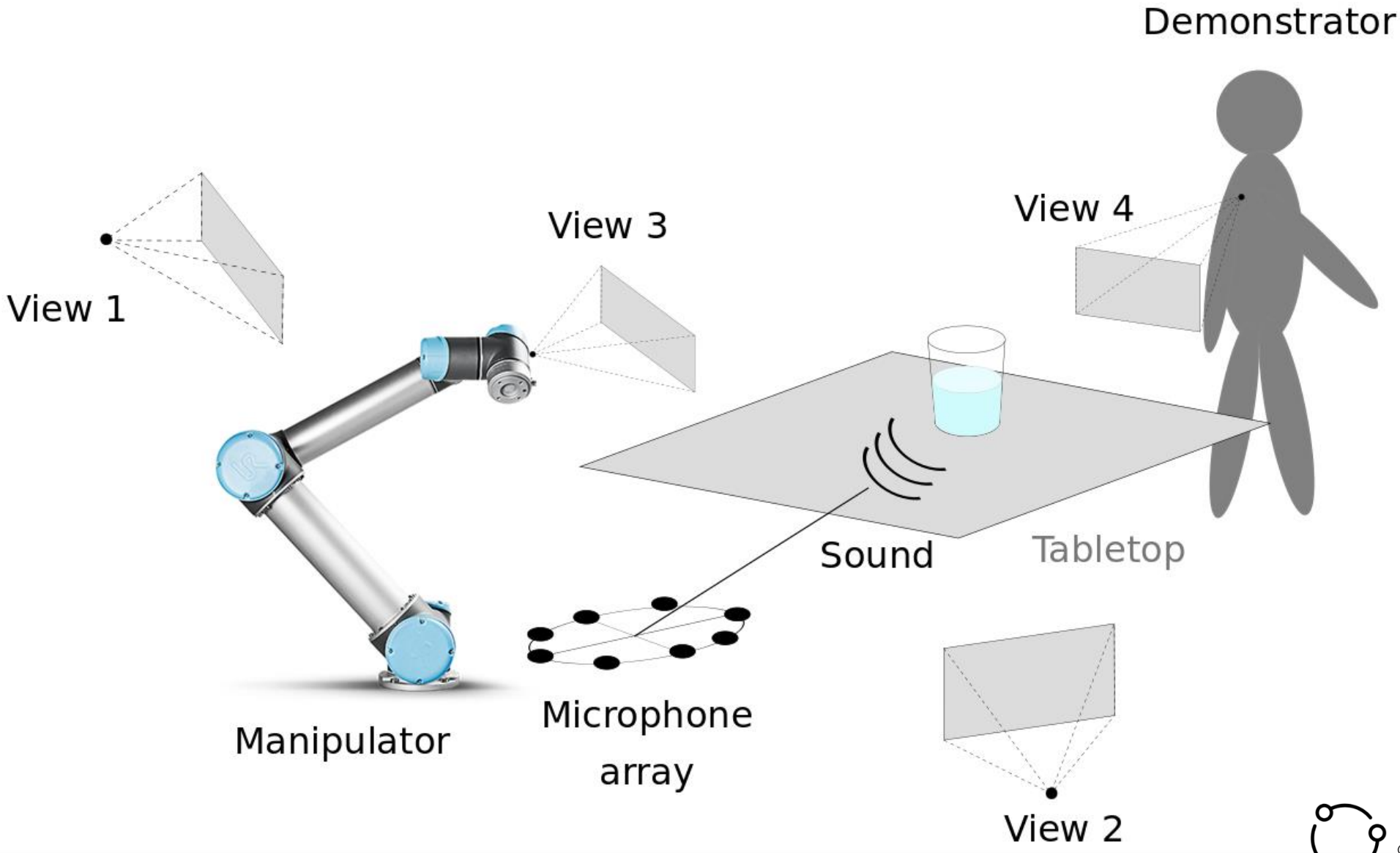
Iterative multi-view 3D-2D shape fitting



Iterative multi-view 3D-2D shape fitting



CORSMAL Containers Manipulation dataset

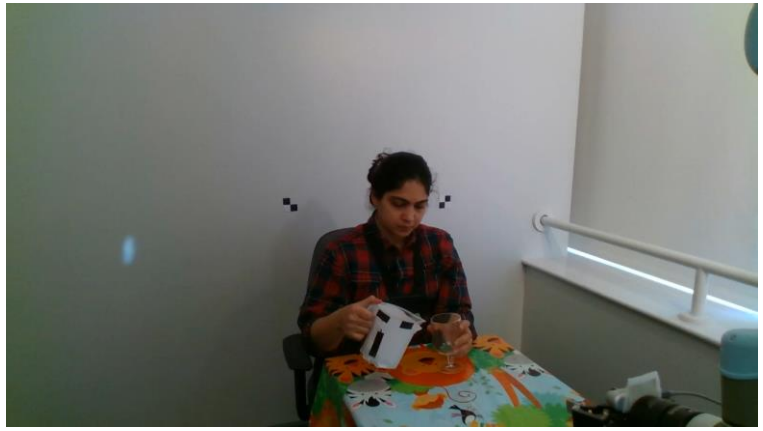


CORSMAL
Collaborative object recognition,
shared manipulation and learning

Object types

1-pound coin as reference size for the objects

CORSMAL Containers Manipulation dataset

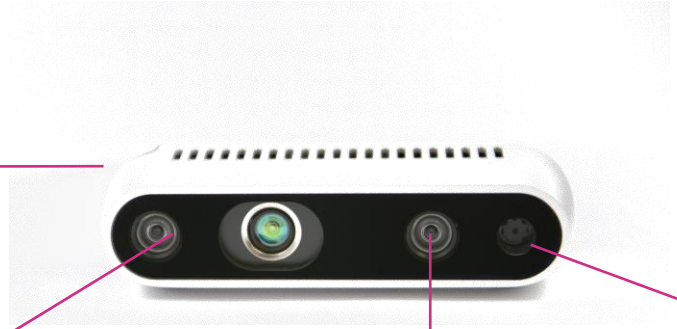


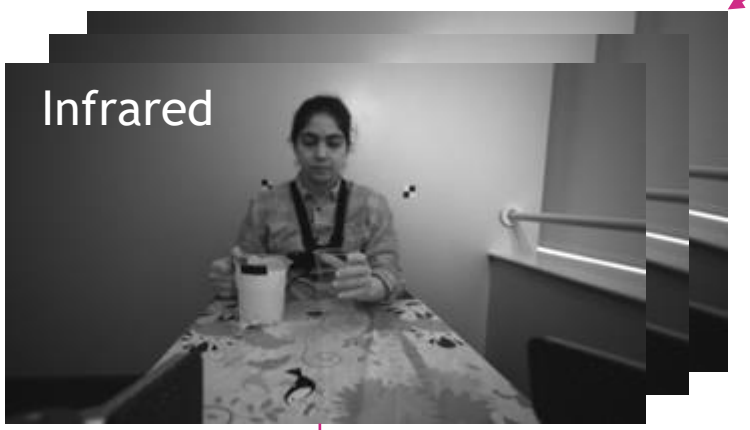
CORSMAL
Collaborative object recognition, shared manipulation and learning

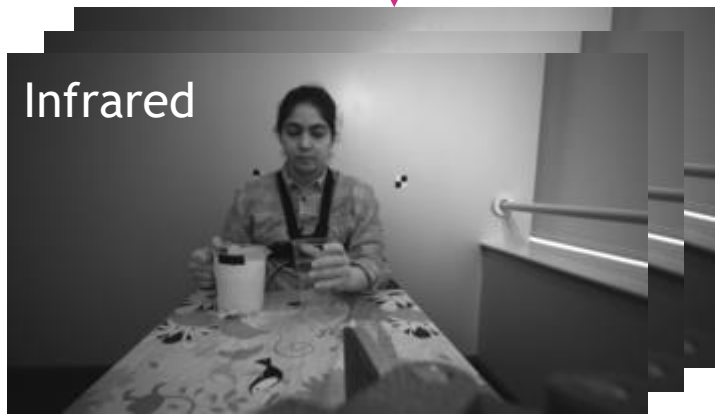
Sensor data

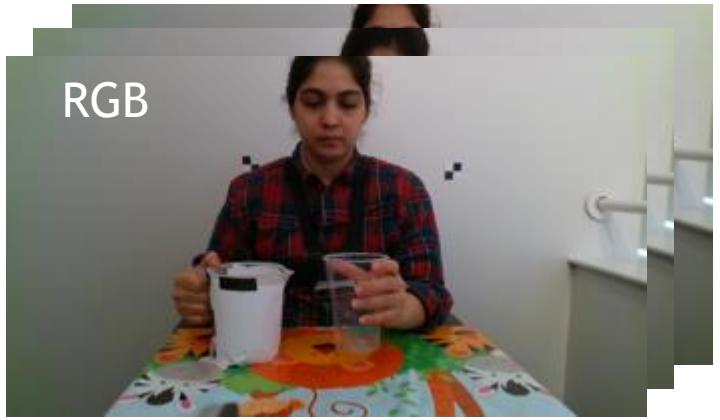
Inertial measurements
accelerometer, gyroscope

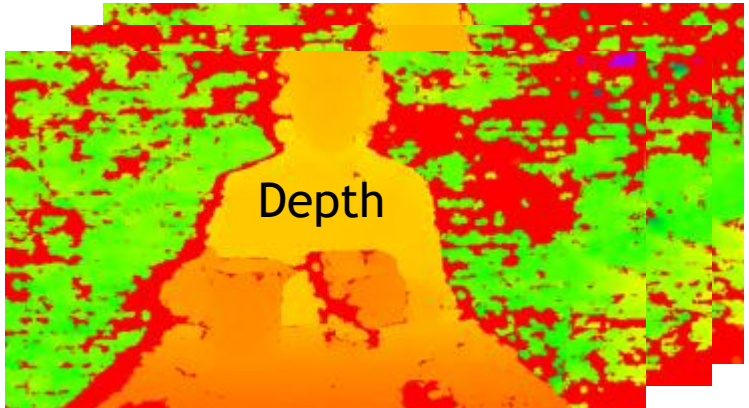
Calibration information
intrinsic, extrinsic parameters







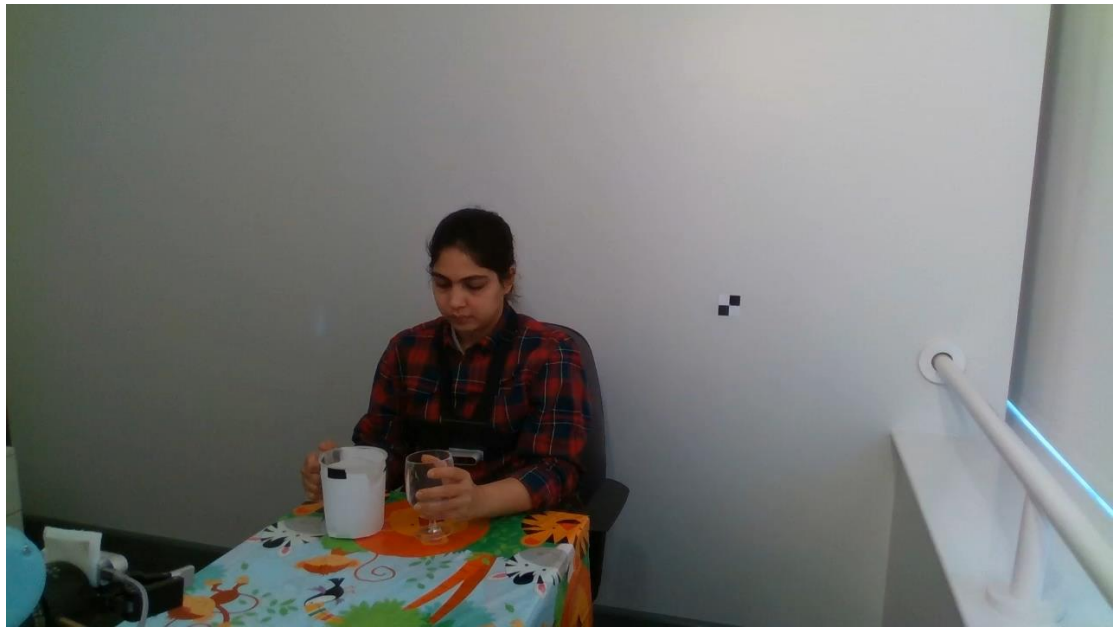




RGB and depth images are spatially aligned

(1280x720 pixels, 30 Hz)



Video (RGB, IR, depth) + audio + inertial data with over 1,000 sequences

Summary

Benchmark for Human-to-Robot Handovers

- Protocol
- Baseline code + specific research modules

Datasets

- CORSMAL Containers dataset
- CORSMAL Containers Manipulation dataset

CORSMAL events at

- Intelligent Sensing Summer School 2019, 2020
- IEEE Int. Conf. on Multimedia and Expo 2020
- Int. Conf. on Pattern Recognition 2020

Partners

Queen Mary University of London

SORBONNE UNIVERSITÉ

EPFL

Sponsors

EPSRC Engineering and Physical Sciences Research Council

Agence Nationale de la Recherche ANR

FNSNF

chist-era

Contact:
a.cavallaro@qmul.ac.uk

CORSMAL
Collaborative object recognition, shared manipulation and learning