Recognition of Human Grasping Patterns for Intention Prediction in Collaborative Tasks

93283 - Pedro Miguel Loureiro Amaral

Master in Robotics and Intelligent Systems

Advisor: Filipe Silva

Co-advisor: Vítor Santos

11/12/2023

DETI-UA

OUTLINE

01 PROBLEM INTRODUCTION

02 HUMAN-ROBOT COLLABORATION SYSTEM

03 RECOGNITION OF HUMAN GRASPED OBJECTS

04 CONCLUSIONS

01

PROBLEM INTRODUCTION

INTRODUCTION

- Human-Robot Collaboration is a research topic increasingly important in the industry to enhance productivity, efficiency, and safety.
- This dissertation aims at the development of an anticipatory system to enhance human-robot collaboration in industrial settings.
- This work was developed in the context of the AUGMANITY mobilizing project.

AUGMANITY

AUGMENTED HUMANITY

OBJECTIVES

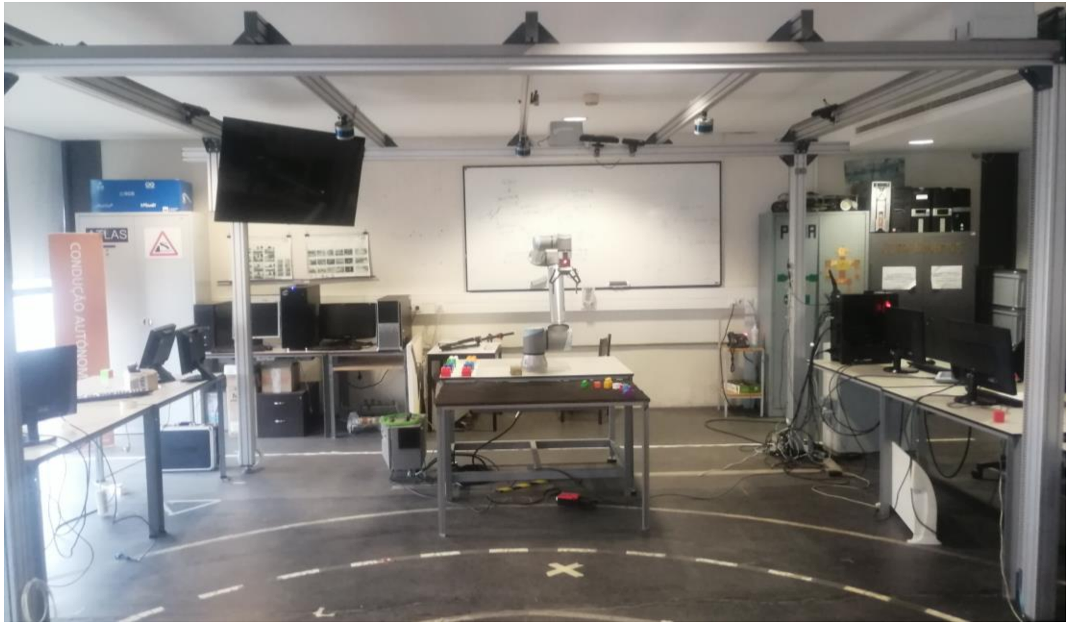
- Review important concepts in HRC and, in this context, study the current research direction of action anticipation and object recognition.
- Develop an infrastructure in ROS to support a practical implementation of action anticipation in the context of HRC.
- Develop a learning framework capable of recognizing the objects being grasped by the human operator to infer his/her intention or needs.

02

HUMAN-ROBOT COLLABORATION SYSTEM

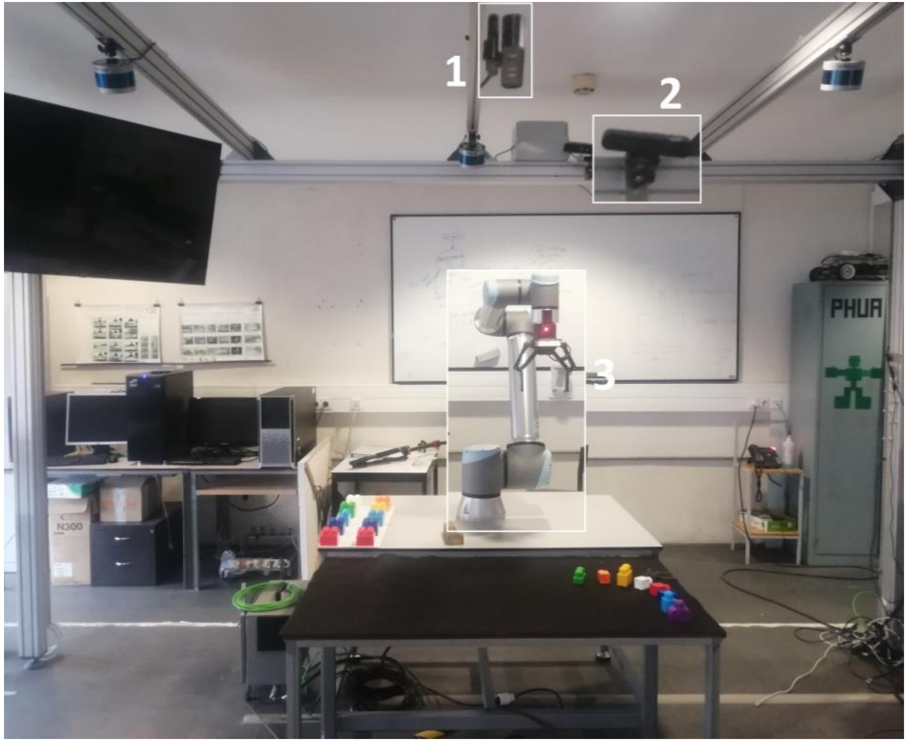
EXPERIMENTAL SETUP

- The experimental setup is part of the collaborative cell (LARCC) at the Laboratory for Automation and Robotics (LAR-DEM-UA).



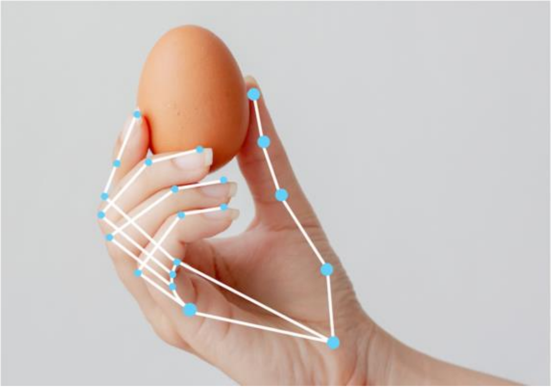
EXPERIMENTAL SETUP

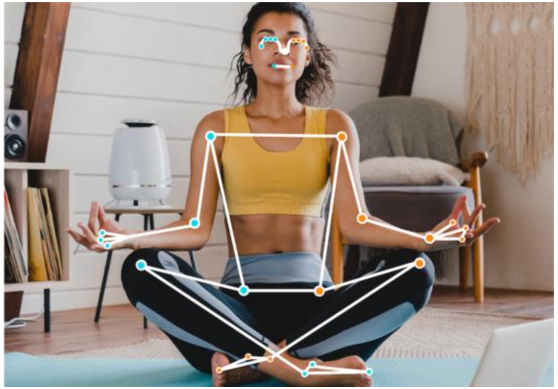
1. Orbbec Astra Pro recording the objects in the table
2. Orbbec Astra Pro recording the human operator
3. UR10e cobot with 6-DOF equipped with a Robotiq 2F-140 gripper



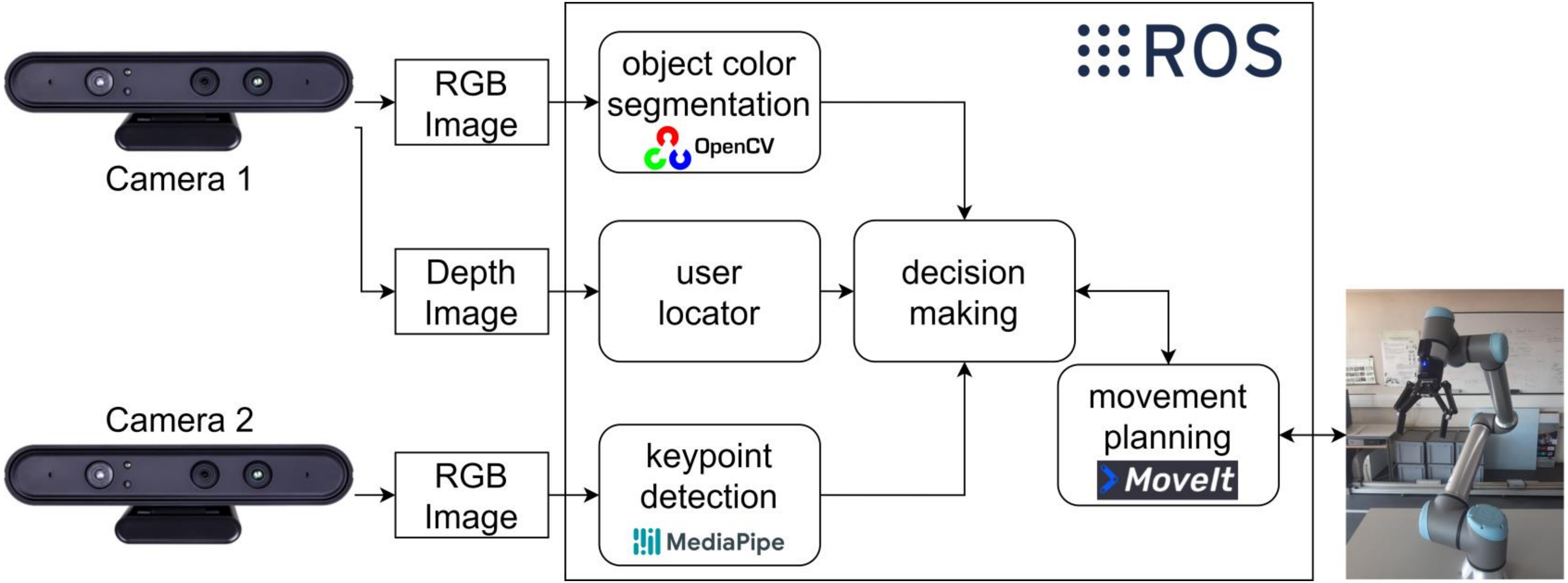
SOFTWARE

- **ROS** – integration and communication among the various components
- **MoveIt** – planning and execution of robot movements
- **OpenCV** – image processing
- **MediaPipe** – keypoint detection





ARCHITECTURE



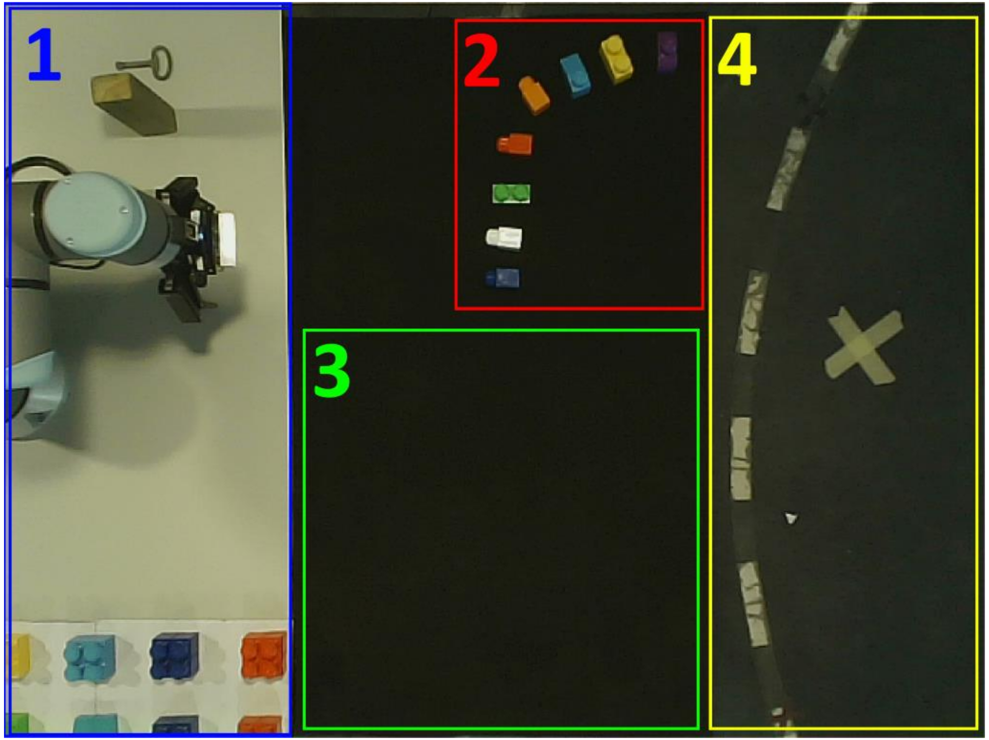
COLLABORATIVE TASK: WORKSPACE

1. Robot area where the big blocks are stored

2. Small blocks that can be hovered by the user to interact with the robot

3. Free space where the robot places the blocks

4. User space



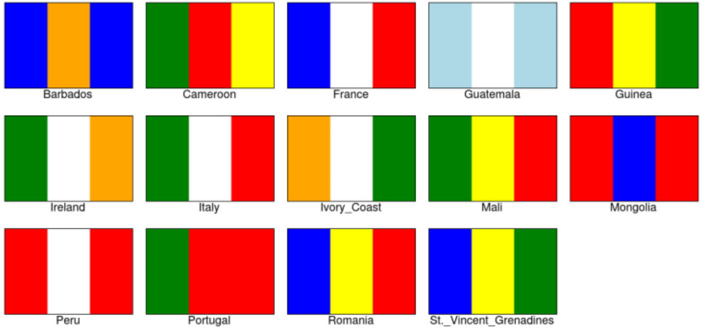
COLLABORATIVE TASK: BUILD A FLAG

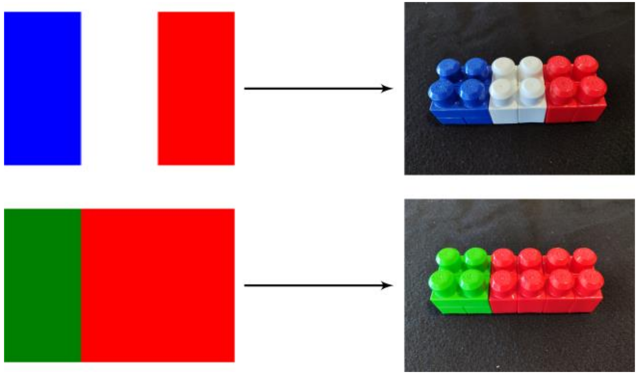
1. The experiment involves the shared task of building a striped flag using Lego bricks (14 pre-selected flags).

2. The human operator selects the first block by interacting with the small blocks.

3. The robot completes the sequence trying to anticipate the human intention.

4. The operator can reject a block using the small violet block.





PROBABILITY BASED

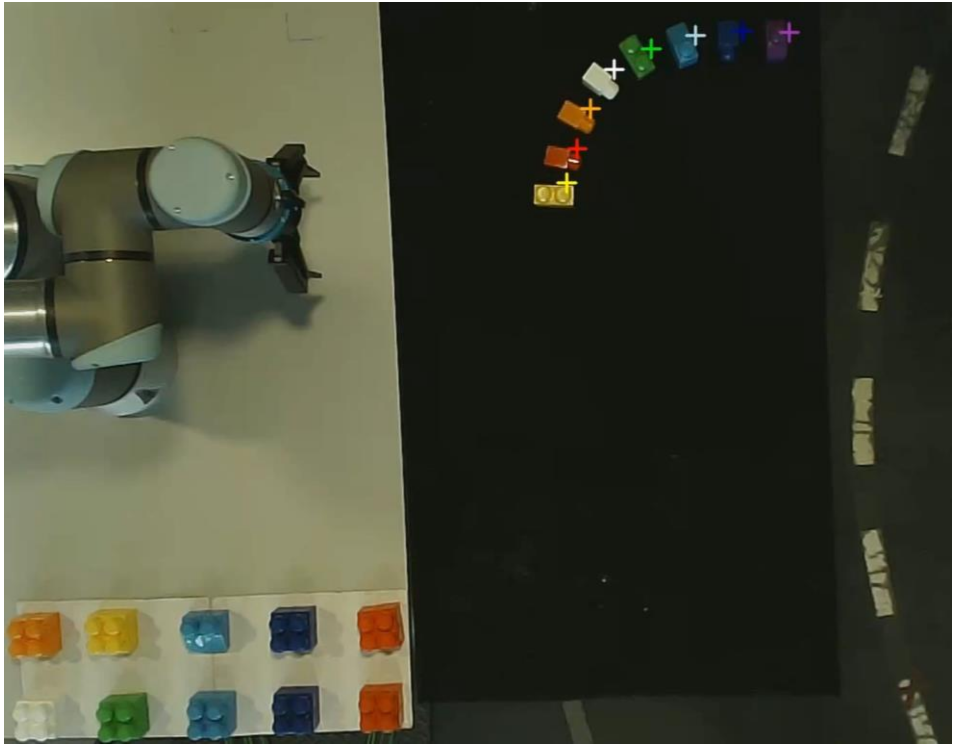
- First block based on a small block interaction
- Next blocks based on a probability table

P(yellow | dark blue) = 50%

P(white | dark blue) = 25%

P(orange | dark blue) = 25%

P(red | dark blue & white) = 100%

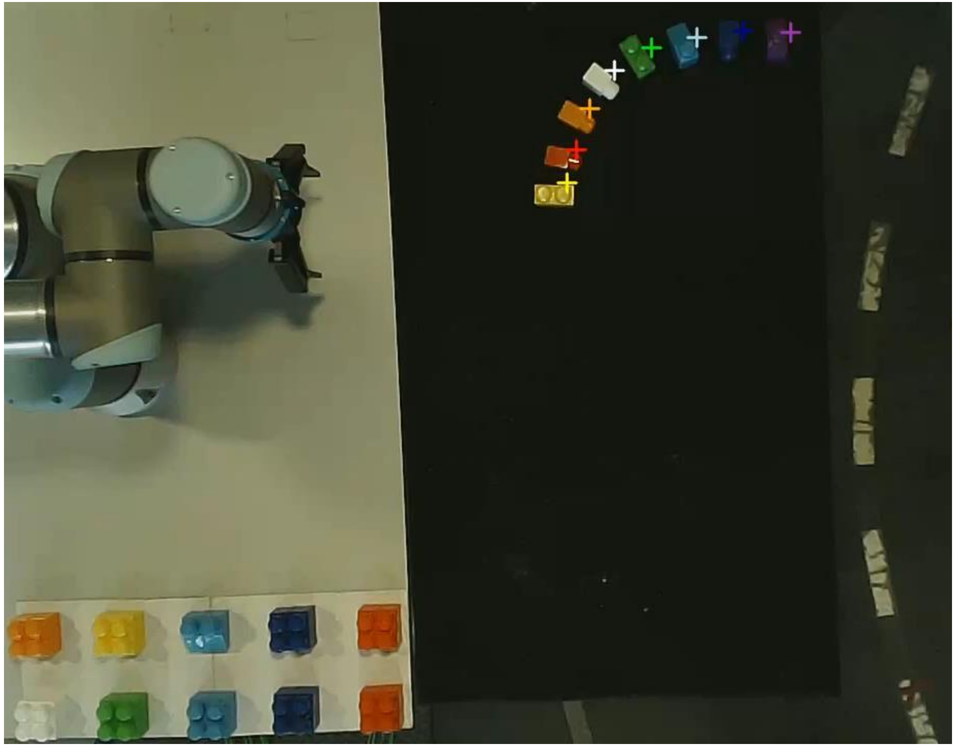


RULE BASED

- First block based on a small block interaction
- Next blocks decided from a set of rules from a configuration file
- The rules are managed by the Experta library

green -> red

green & red -> red

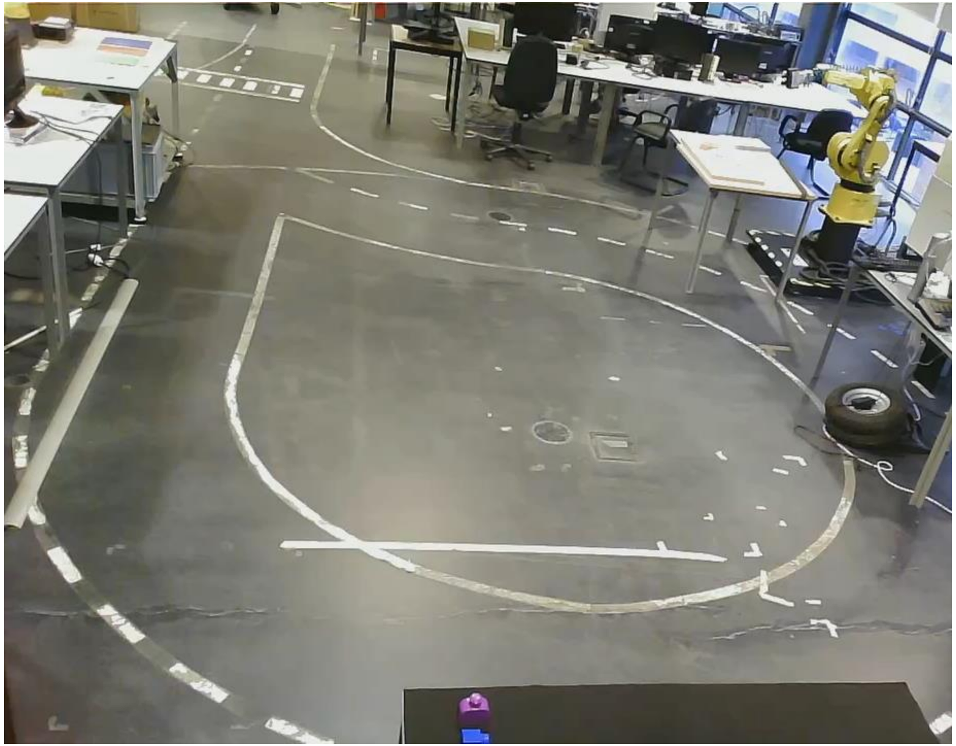


RULE BASED + MEDIAPIPE

- Rules are still used to decide the following blocks.
- The first block is now picked up by the user.
- MediaPipe is used to detect the right-hand (the zone of interest) and the object is detected with color segmentation.

green -> red

green & red -> red

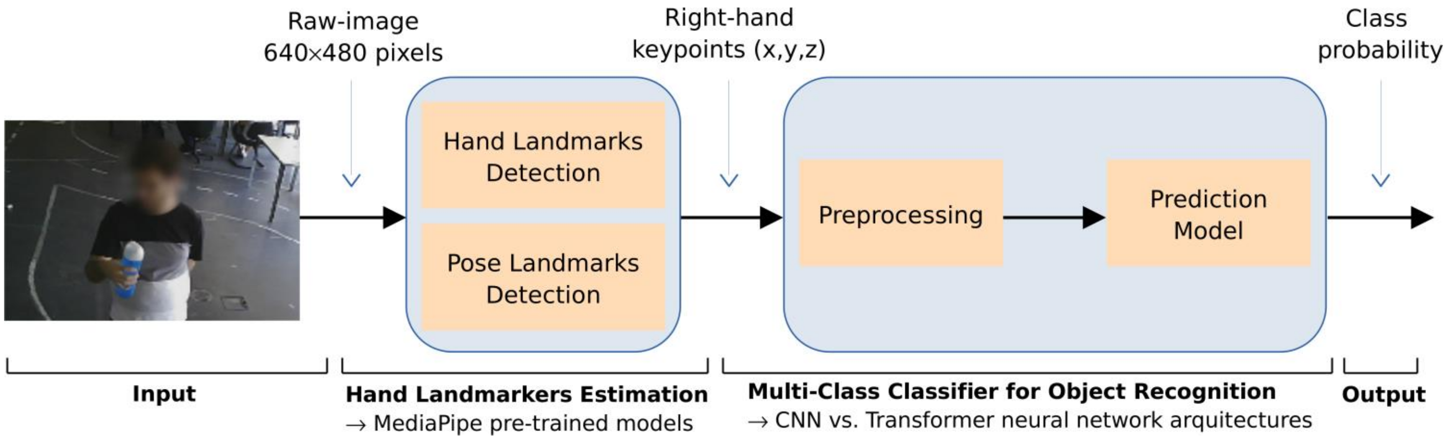


03

RECOGNITION OF HUMAN GRASPED OBJECTS

APPROACH

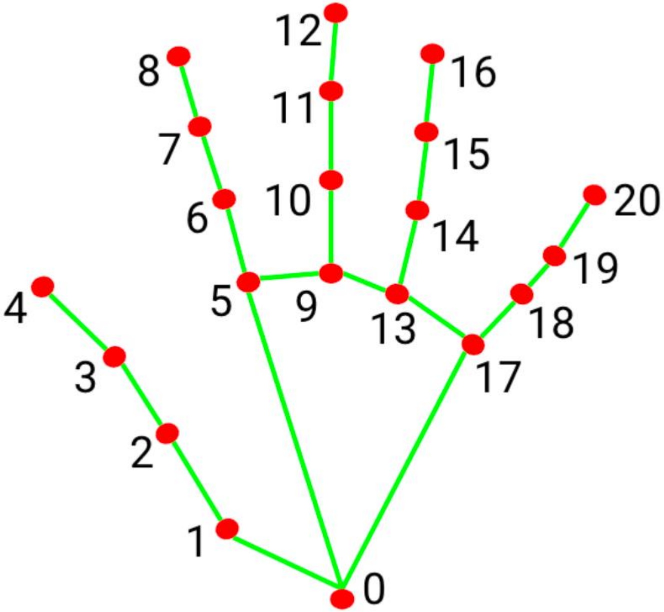
- This work proposes a learning-based framework to enable an assistive robot to recognize the object grasped by the human operator.
- This framework combines MediaPipe and a deep muti-class classifier.



MEDIAPIPE EVALUATION

- The MediaPipe models either extract all 21 right-hand (x, y, z) keypoints or none.
- Previous studies in the literature show trust in its tracking accuracy.
- Experiences with and without occlusions were made showing acceptable results.

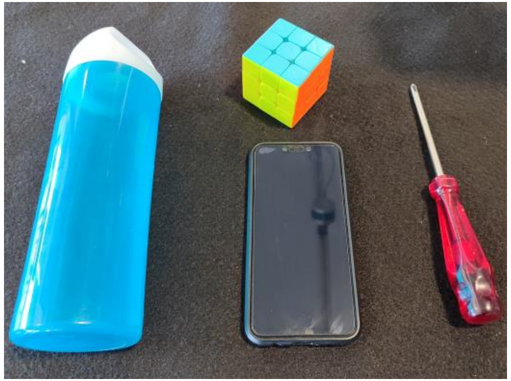
	Hand Open	Hand Closed
No Movement	100%	99.6%
In Movement	93.2%	98.4%

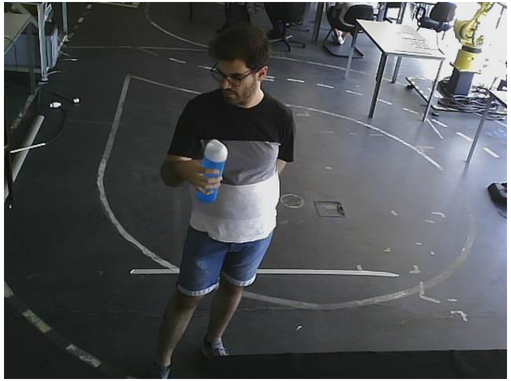


DATASET ACQUISITION

- The objects used include a water bottle, a Rubik's cube, a smartphone, and a screwdriver.
- The acquisition involved the participation of three volunteers, who were asked to naturally grab an object and execute small movements.

A tool was developed in ROS to ease the creation of a new dataset.





DATASET ACQUISITION

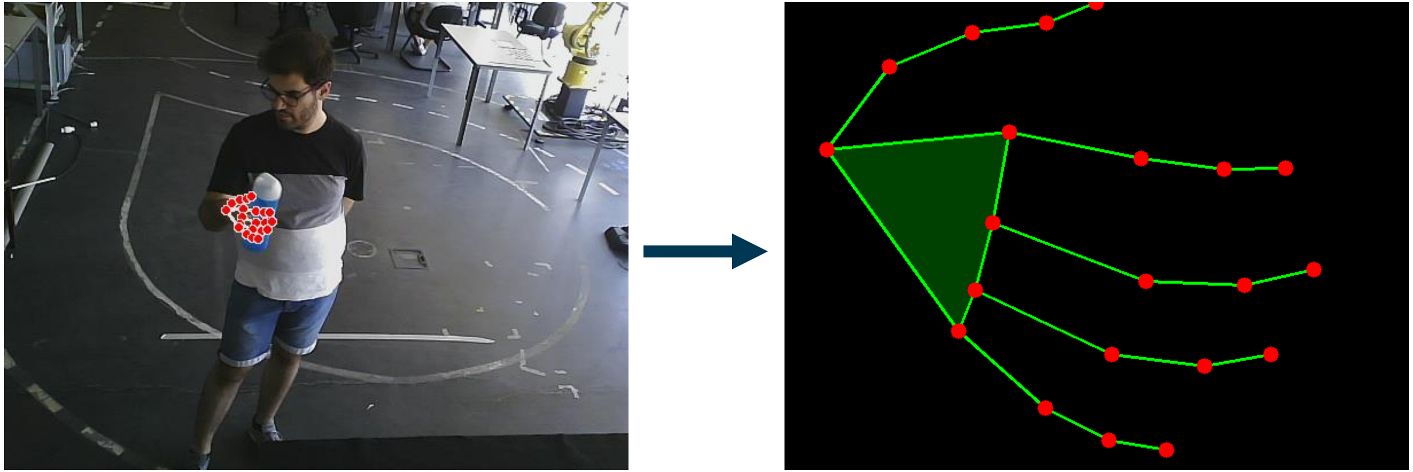
- The dataset is distributed almost equally between the 3 participants and the 4 objects in the acquisition.
- For each participant and object, 4 sessions were recorded to ensure variability within the same human-object interaction.

Dataset	Bottle	Cube	Phone	Screwdriver	Total
User1	828	928	950	957	3663
User2	886	926	939	946	3697
User3	904	907	937	946	3694
Total	2618	2761	2826	2849	11 054

DATASET PREPROCESSING

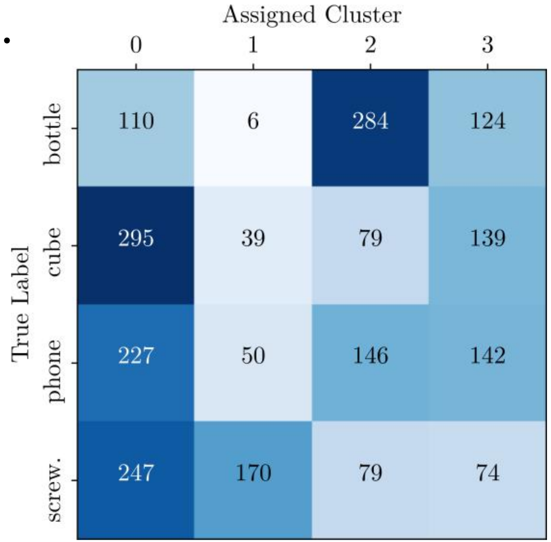
- The keypoints generated by MediaPipe corresponding to the right hand suffer further transformations and normalization.

- This ensures maximum separation between keypoints helping the data be more consistent and invariable to hand position, size or scale.



DATASET STATISTICAL ANALYSIS

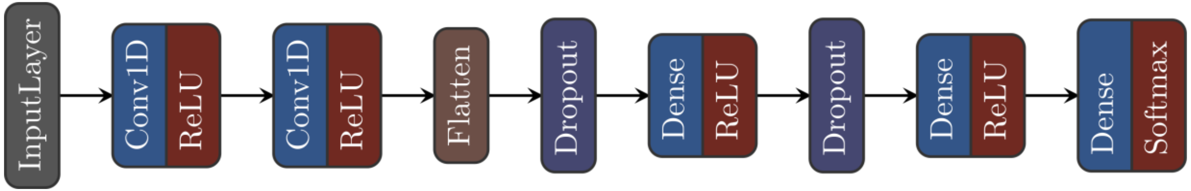
- K-Means clustering was used on the normalized keypoints as a statistical analysis of the complexity of this problem.
  - Clusters 0 and 3 exhibit a mix of all classes.
  - Cluster 0 encompasses 39.8% of all data.
  - Clusters do not align with classes.
- Given the results, deep learning models were used showing a more robust generalization.



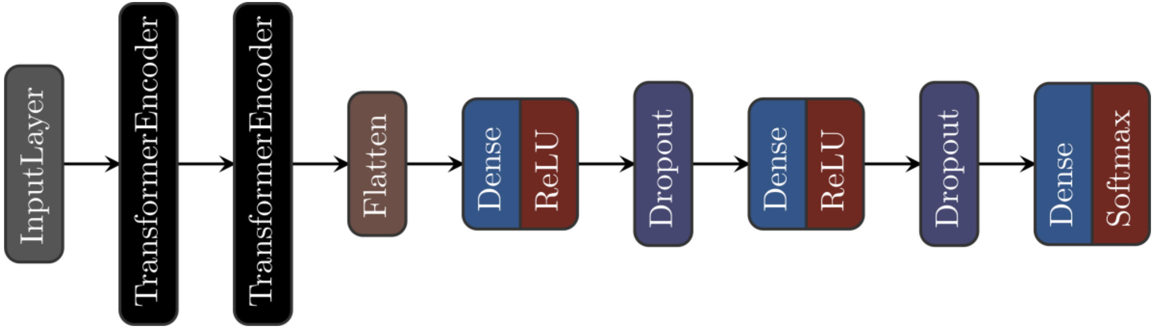
	0	1	2	3
bottle	110	6	284	124
cube	295	39	79	139
phone	227	50	146	142
screw.	247	170	79	74

DEEP ARCHITECTURES

- CNN – 156 644 trainable parameters & 127s training time



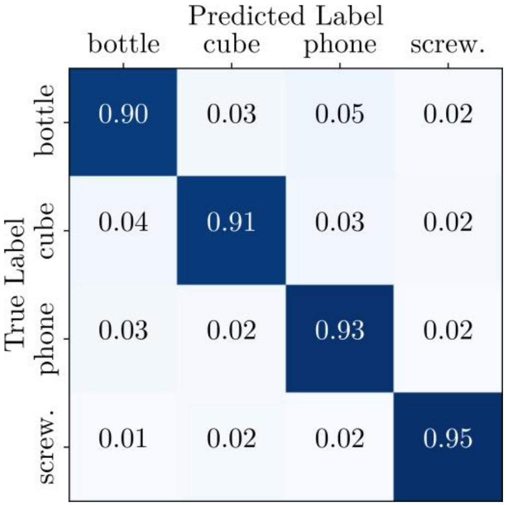
- Transformer – 16 384 trainable parameters & 1351s training time



PERFORMANCE EVALUATION

CNN

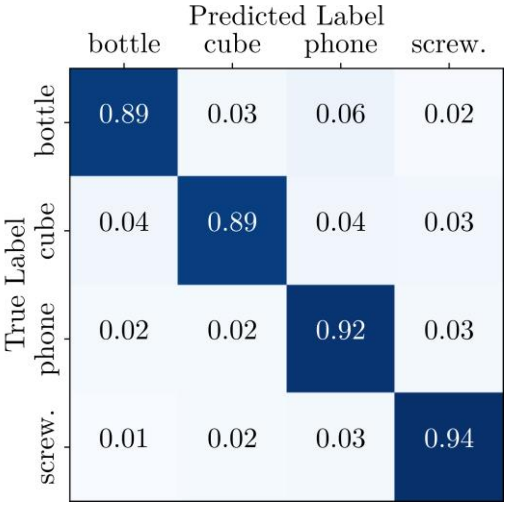
Accuracy	Precision	Recall	F1-Score
0.9240	0.9242	0.9240	0.9240



	bottle	cube	phone	screw.
bottle	0.90	0.03	0.05	0.02
cube	0.04	0.91	0.03	0.02
phone	0.03	0.02	0.93	0.02
screw.	0.01	0.02	0.02	0.95

Transformer

Accuracy	Precision	Recall	F1-Score
0.9109	0.9115	0.9109	0.9109



	bottle	cube	phone	screw.
bottle	0.89	0.03	0.06	0.02
cube	0.04	0.89	0.04	0.03
phone	0.02	0.02	0.92	0.03
screw.	0.01	0.02	0.03	0.94

PRACTICAL CHALLENGES

SESSION BASED

Acess the impact of the sessions on the classifier’s performance.

USER SPECIFIC

Provide insights into the classifier’s performance on data from a single user.

LEAVE ONE USER OUT

Evaluate the performance on data from a user not included in the training data.

SESSION BASED TESTING

- Access the impact of session-specific data on the classifier’s performance.
  - “Full Dataset” – trained and tested on all users and all sessions

Model	Accuracy	Precision	Recall	F1-Score
CNN	0.9210	0.9214	0.9211	0.9211
Transformer	0.9017	0.902	0.9016	0.9017

  - “Session-Based Testing” – tested on a session not used for training

Metric	Model	Session 1	Session 2	Session 3	Session 4
Accuracy	CNN	0.8493	0.8138	0.7844	0.7718
	Transformer	0.8458	0.8027	0.7902	0.7613

USER SPECIFIC TESTING

- Provide insights into the classifier’s performance on data from a single user.

  - “Full User Dataset” – trained and tested on all sessions from one user

Metric	Model	User1	User2	User3
Accuracy	CNN	0.9674	0.9100	0.9163
	Transformer	0.9423	0.8929	0.8730

  - “Session-Based User1 Testing” – using only user1’s data, tested on a session not used for training

Metric	Model	Session 1	Session 2	Session 3	Session 4
Accuracy	CNN	0.9257	0.8364	0.9053	0.7883
	Transformer	0.9078	0.8171	0.8742	0.7636

LEAVE ONE USER OUT

- Evaluate the performance on data from a user not included in the training data.
  - “Leave-One-User-Out Test” – tested on all sessions from one user not used in training

Metric	Model	User1	User2	User3
Accuracy	CNN	0.7969	0.5827	0.5488
	Transformer	0.8006	0.5730	0.5350

04

CONCLUSIONS

CONCLUSIONS

- The collaborative workspace and infrastructure was validated and allowed for preliminary anticipation experiments.
- The approach adopted in this work was to perform action anticipation by recognizing the object being grasped by the user.
- The object recognition framework was tested in different cases and showed promising results for a practical implementation.

FUTURE WORK

- **More Data** - test the current architectures with a bigger dataset.
- **Adaptability Strategies** - explore advanced techniques to further enhance the adaptability and generalization capabilities of the system.
- **Real-Time Integration** – test the integration of the object recognition pipeline into a real-time collaborative application.

CONTRIBUTIONS

- **Article** - P. Amaral, F. Silva, V. Santos, «Recognition of Grasping Patterns using Deep Learning for Human-Robot Collaboration», Sensors, https://www.mdpi.com/1424-8220/23/21/8989.

- **Github Repository** - Pedro Amaral, Recognition of Human Grasping Patterns for Intention Prediction in Collaborative Tasks, https://github.com/pedromiglou/MRSI_Thesis_Action_Anticipation.

- **Kaggle Dataset** - Pedro Amaral, Human Grasping Patterns for Object Recognition, https://www.kaggle.com/datasets/pedromiglou/human-grasping-patterns-for-object-recognition.

THANK YOU

Do you have any questions?