Toward the Detection of the Human Intention to Interact with a Service Robot

G. Abbate[1], A. Giusti[1], V. Schmuck[2], O. Celiktutan[2] and A. Paolillo[1]

[1]IDSIA, USI-SUPSI, Lugano, Switzerland. Email: name.surname@idsia.ch
[2]Department of Engineering, King's College London, UK. Email: name.surname@kcl.ac.uk

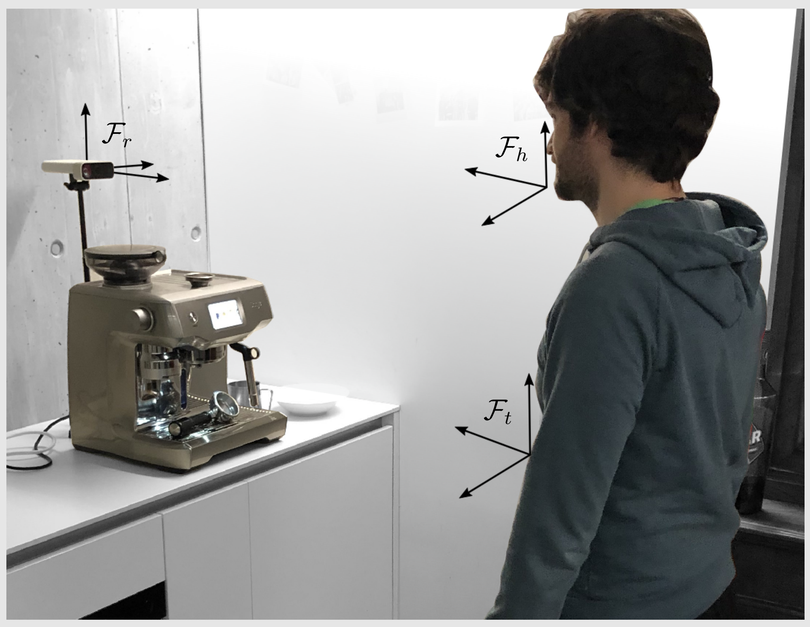
Socially Acceptable Service Robots

- *Social robots* help us in services like reception, home assistance, personal care
- Effective service is obtained by interpreting *social cues* in body motion, to anticipate the *intention to interact* of the user
- A *social service robot* should be able to:
  - keep track of nearby people
  - predict when a person intends to interact
  - react accordingly

Our Approach

- We track people using an off-the-shelf *body tracking* technology
- We focus on detecting the *human's intention to interact*
- We train a *binary classifier* to predict whether a person will interact by looking at their tracked motion, before they actually interact
- We consider as *features* the user's planar pose, linear velocity and head orientation

Dataset Collection

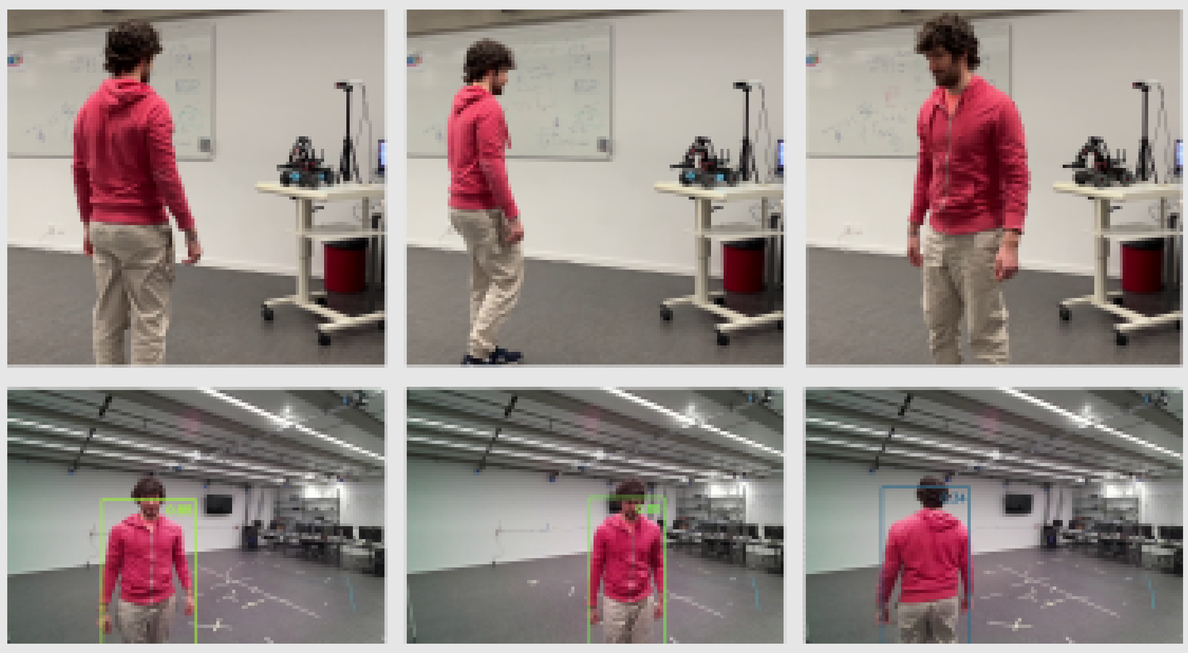


- We collect a *real-world dataset* of people interacting with a coffee machine
- An *RGB-D sensor* is placed on top of the coffee machine to simulate the point of view of a robot
- The motion of people entering the proximity of the coffee machine area is *tracked* and *recorded*
- The dataset contains *3422 unique sequences* of tracked users (more than 12 hours of data)
- If a user is within 1 m of the coffee machine for more than 5 seconds, they are labeled as *interacting*
- The motion of people walking nearby the machine is used to *predict* their *intention to interact*

Preliminary Results

- The users' distance is a very strong cue of whether they will interact or not
- The classifier, when evaluated against all testing frames pooled together, shows high performance, i.e. Area Under the ROC Curve (AUROC) > 0.9
- Splitting the samples into seven distance bins, we observe that richer user motion (e.g. using body motion *and* head orientation) yields better results
- With rich sensory information the prediction at short distances is more difficult (AUROC ≈ 0.65) than at long distances (AUROC ≈ 0.8)
- It is difficult to understand whether someone close to the machine is there to interact or to do something else
- The intention to interact of people approaching from afar is well predictable from their body language and head orientation

Qualitative Results



- When the user wants to interact, the robot rotates toward them and turns its LEDs on (first two snapshots)
- If the user does not interact, it does not move and keeps LEDs off

Future Work and Open Challenges

- Measure the level of *social acceptance* via qualitative HRI questionnaires
- Investigate other tools from the *HRI community*, such as proxemics notions, social spaces and non-verbal communication modalities
- New and extended *dataset collection campaign* in public environments
- Deployment with *human-sized robots*, to also analyze how the appearance of the robot can influence the interaction
- More *complex scenarios*, where the robot may have to interact with several users at the same time

ICRA LONDON · 2023

SERMAS

Funded by the European Union