Introduction

Interactive robots benefit from perspective taking mechanisms, allowing to infer what a user can see and to more accurately predict their behavior [1]. For self-referential understanding of actions of others, assuming their visual perspective is essential.

Previous works [1, 2] are constrained to environments equipped with markers and/or motion capture systems.

Our goal: estimate perceived world of humans and the surroundings of the robot, whilst not constraining the environment.

Markerless Perception

Self-exploration phase: build a map of the environment using RGB-D images and visual SLAM

Reason about parts of the environment which are currently out of sight, but visited previously

Object recognition using deep learning on iCub eye images [3].

To reason about the human’s field of view, we improve performance of a state of the art head pose estimator [4].

We propose new method which normalizes the input so it becomes more similar to the training data.

Perspective Taking

Following the literature [5], we distinguish level 1 and level 2 perspective taking

Level 1 perspective taking:

Find whether human can see an object

We propose using line of sight tracing

Level 2 perspective taking:

Estimate what world looks like to human

Mentally rotate environment map in the frame of reference of the human

Left-right judgments using the same spatial reasoning algorithm as from the robot’s perspective.

Conclusions

Improvements in key parts of the perspective taking pipeline allow system to work in markerless setups.

Validation in several experiments (head pose estimation + line of sight + mental transformation) using iCub.

Extend application scenarios of popular head pose estimation algorithm.

Future work

More accurate gaze estimates: take human’s eye movements into account.

Investigate developmental process of perspective taking and relationship to joint attention.

Acknowledgements

This work was supported in part by the EU FP7 project WYSIWYD under Grant 612139.

References