Markerless Perspective Taking for Humanoid Robots in Unconstrained Environments

Tobias Fischer and Yiannis Demiris
{t.fischer, y.demiris}@imperial.ac.uk
Personal Robotics Lab, Imperial College London
imperial.ac.uk/**PersonalRobotics**

ICRA2016 IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION

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

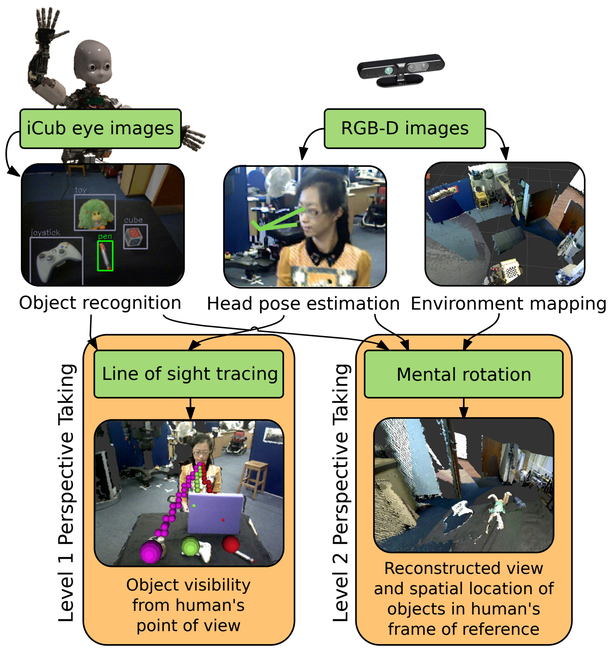

Figure 1: Overall flow of the proposed method

- 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]

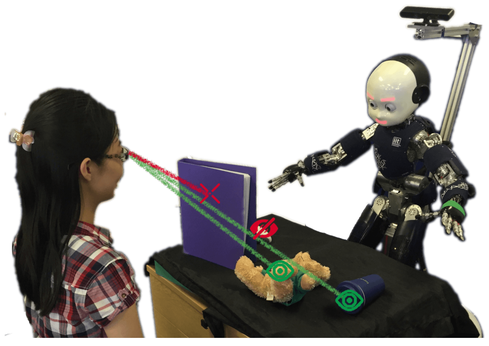

Figure 2: Typical setup

- 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

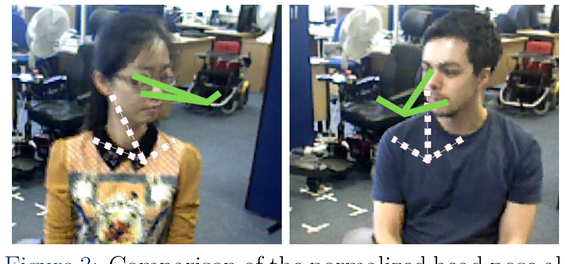

Figure 3: Comparison of the normalized head pose algorithm (green) and original method (white, dotted)

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**

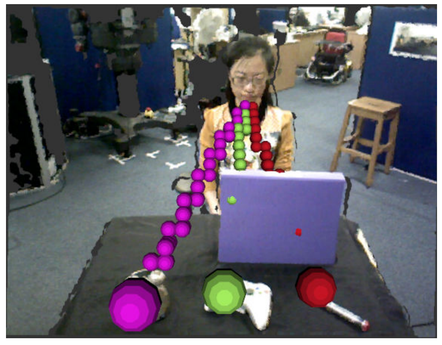

Figure 4: Line of sight tracing

- **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 robots perspective

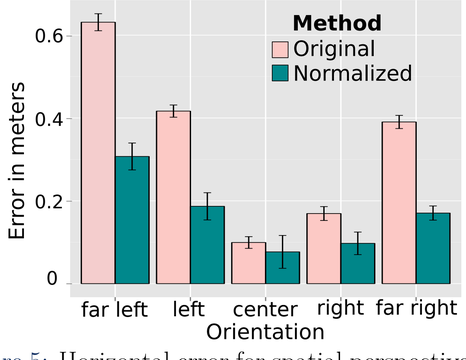

Figure 5: Horizontal error for spatial perspective taking using original / proposed head pose estimation

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

[1] M. Johnson and Y. Demiris, "Perceptual Perspective Taking and Action Recognition," *Int. J. Adv. Robot. Syst.*, 2005.

[2] A. K. Pandey, M. Ali, and R. Alami, "Towards a Task-Aware Proactive Sociable Robot Based on Multi-state Perspective-Taking," *Int. J. Soc. Robot.*, 2013.

[3] G. Pasquale *et al.*, "Teaching iCub to recognize objects using deep Convolutional Neural Networks," in *Proc. Work. Mach. Learning Interactive Syst.*, 2015.

[4] G. Fanelli *et al.*, "Real Time Head Pose Estimation from Consumer Depth Cameras," in *Annu. Symp. German Association Pattern Recognition*, 2011.

[5] P. Michelon and J. M. Zacks, "Two kinds of visual perspective taking," *Perception & Psychophysics*, 2006.