Multimodal people engagement with iCub

Salvatore M. Anzalone, Serena Ivaldi, Olivier Sigaud, and Mohamed Chetouani

Institut des Systemes Intelligents et de Robotique
CNRS UMR 7222 & Universite Pierre et Marie Curie, Paris, France

Abstract. In this paper we present an engagement system for the iCub robot that is able to arouse in human partners a sense of "co-presence" during human-robot interaction. This sensation is naturally triggered by simple reflexes of the robot, that speaks to the partners and gazes the current "active partner" (e.g. the talking partner) during interaction tasks. The active partner is perceived through a multimodal approach: a commercial rgb-d sensor is used to recognize the presence of humans in the environment, using both 3d information and sound source localization, while iCub's cameras are used to perceive his face.

Keywords: engagement, attention, personal robots

1 Introduction

A sense of attention and engagement naturally emerges in humans during social interactions. This sensation of "co-presence" depends on different behaviors that show cognitive capabilities [1]. In particular, establishing and maintaining eye contact between people involved in interactions is one of the key factors causing the "social engagement" [3] [4]. Thanks to this behavior we can communicate in a non-verbal way a variety of social information about attention and involvement in a conversation or a context. Harnessing people perception is thus fundamental for enhancing human-robot interaction. For example when robot and humans cooperate to perform a task or share a plan, if the robot expresses its engagement to the human partner by fixation, the latter is naturally prone to consider that the robot is following its commands. Social robots should be provided with this ability to improve the sense of social connection with humans [5] [6]. In this paper we present a simple engagement system for the humanoid robot iCub [7], which will be foundational to more evolved HRI tasks.

2 Multimodal active partner tracking

The proposed system implements a multimodal approach to people detection and is able to let the robot identify and track the "active" partner (Figure 1). A Microsoft Kinect is used to perceive and reconstruct the environment to retrieve the possible presence of humans. At the same time the 4 microphones array embedded in the Kinect is used to localize sound sources. It is assumed that a sound coming from the same direction of a human corresponds to his voice or to some noise that the partner intentionally produces to catch the robot's attention. Hence, when a sound is perceived, the system fuses the
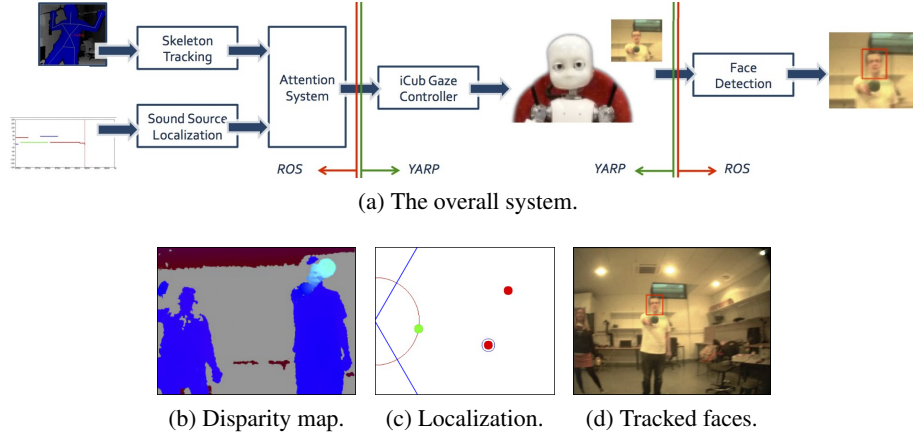
(a) The overall system.

(b) Disparity map.   (c) Localization.   (d) Tracked faces.

Fig. 1: A sketch of the system.

source localization information with the people location information and, accordingly, marks the "active partner", i.e. the current person talking or producing sounds. This information is exploited by a gaze controlling mechanism to move the robot's head and eyes so as to center the person in its field of view. Finally, the iCub camera stream is analyzed to detect human faces, to fine tune the person localization. Notably, software modules have been implemented in different middlewares, namely YARP and ROS, which required the implementation of a bidirectional YARP-ROS bridge to comply with the different communication protocols.

3   Experimental results

The proposed engagement system has been used in several developmental robotics scenarios, where two or more caregivers interact with the robot to teach new objects. As suggested in Figure 1, the system is able to detect and track the partners in the environment, to mark the active partner and to detect his face. As one of the partners speaks, the produced sound is detected and matched with the person's location, then the robot gazes toward him as the "active caregiver". The engagement system has also been used in real-life scenarios in which the robot joins a conversation between people. When people talk alternatively, as during a conversation, the robot gaze at the "active" one, emulating a third "listening" to the conversation. The video showing some of the experiments can be found at: http://macsi.isir.upmc.fr. From the point of view of the human partners, a sensation of "social cohesion", or "togetherness", grows up from the gazing of the robot: humans have the feeling of being perceived by the robot while they interact with it as in a real relation between human student and human teacher. In particular, the pointing behaviour is able to generate a feeling of social connection between the robot and the active partner: the human caregiver feels the attention of the robot as feedback to his teaching attempts. Remarkably, when partners were asked a feedback about the experiments, they admitted that the simple gaze of the robot was enough to give the im-

Fig. 2: iCub focuses on the "active partner", alternatively on the left and on the right.

pression that the robot was effectively listening to their conversation: co-presence has been felt by the partners as an increase of the mutual engagement during interaction.

4   Conclusions and future works

In this paper a multimodal people engagement system for the humanoid iCub has been presented. Experiments performed show that by the use of gazing and faces engagement it is possible to establish a direct connection during interactions between humans and robots and to arouse in the human partners a sense of co-presence of the robot in the environment. Results encourage us to improve the mutual engagement adding new perception-action modalities, and further investigate the role of engagement in teaching scenarios, evaluating its effect in a qualitative way as well as in a quantitative way.

Acknowledgment

References

1. N. Durlach and M. Slater. Presence in shared virtual environments and virtual togetherness. *Presence: Teleoperators & Virtual Environments*, 2000.
2. C. Breazeal. Toward sociable robots. *Robotics and Autonomous Systems*, 2003.
3. C.L. Breazeal. *Designing sociable robots*. The MIT Press, 2004.
4. S. Al Moubayed, M. Baklouti, M. Chetouani, T. Dutoit, A. Mahdhaoui, J.C. Martin, S. Ondas, C. Pelachaud, J. Urbain, and M. Yilmaz. Generating robot/agent backchannels during a storytelling experiment. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009.
5. C. Rich, B. Ponsler, A. Holroyd, and C.L. Sidner. Recognizing engagement in human-robot interaction. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*. IEEE, 2010.
6. L. Natale, F. Nori, G. Metta, M. Fumagalli, S. Ivaldi, U. Pattacini, M. Randazzo, A. Schmitz, and G. G. Sandini. *Intrinsically motivated learning in natural and artificial systems*, chapter The iCub platform: a tool for studying intrinsically motivated learning. Springer-Verlag, 2012.