Selecting and Commanding Individual Robots in a Vision-Based Multi-Robot System

Alex Couture-Beil, Richard T. Vaughan, and Greg Mori

Autonomy Lab / Vision & Media Lab
Simon Fraser University
Burnaby, B.C., Canada
{*asc17, vaughan, mori*}*@sfu.ca*

Abstract—This video presents a computer vision based system for interaction between a single human and multiple robots. Face contact and motion-based gestures are used as two different non-verbal communication channels; a user first selects a particular robot by simply looking at it, then assigns it a task by waving his or her hand.

Keywords-human-robot interaction; multi-robot system; task assignment; gesture-controlled robot; face detection; distributed system

I. INTRODUCTION

Eye contact and eye gaze plays an important role in initiating and regulating communication between people [1]; it can be used to indicate cognitive focus, and to disambiguate who is being addressed. Goffman uses the term *face engagement* to describe the process in which people use eye contact, gaze and facial gestures to interact with or engage each other [2]. In this video, we present a system which demonstrates the use of: 1) face engagement as a method for selecting a particular robot from a multi-robot system, and 2) motion-based gestures for commanding individually selected robots. An example of a typical interaction is shown in Fig 1.

II. THE ROBOTS

Our system is comprised of three iRobot Creates, each fitted with a gumstix single-board computer and wireless network card. The computer is plugged into a custom motherboard with programmable multi-colour LEDs for user feedback, and infrared range sensors for obstacle avoidance. The gumstix computer runs the robot controller, while a laptop processes data from its built-in video camera.

The robots are located in a 7x10m room with a single human operator. Fiducial markers [3] are located on three of the walls to help robot navigation and to indicate two different target zones: a red zone and a green zone, as well as a third zone where robots wait for instructions from the user.

Robots first approach and wait at the instruction zone for a user to appear and assign tasks. Once a human has been found, the robots pivot to look at the user "head-on". Faces are detected with the Viola-Jones method [4]; the OpenCV [5] implementation is used. Since the face detector is trained on frontal faces, faces looking directly at the robot will receive a higher score than faces looking at different robots.
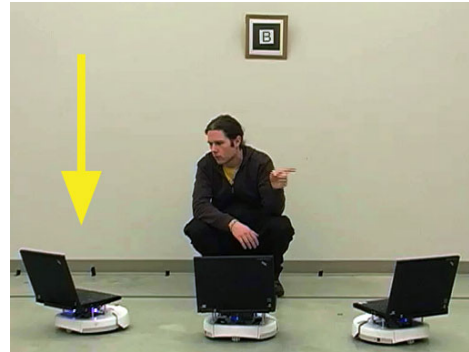
(a) A user selects an individual robot by looking at it, and assigns it a task by waving his hand.
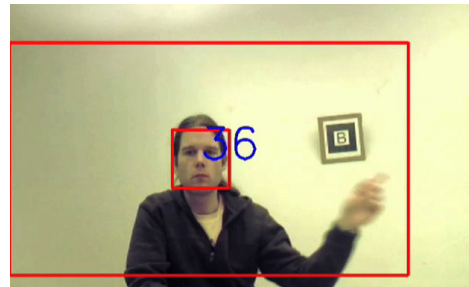
(b) A user-centric region is identified; a learned classifier uses optical flow from the region to discriminate between gestures
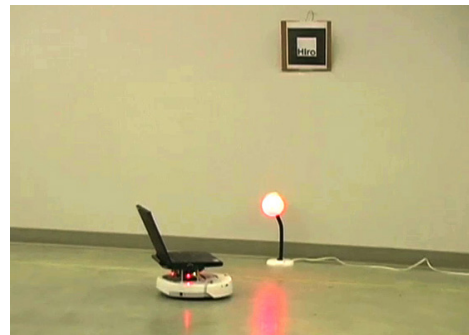
(c) Robots travel to one of two zones as commanded by the user; colours are only used to illustrate different zones – robots use fiducial markers for localization

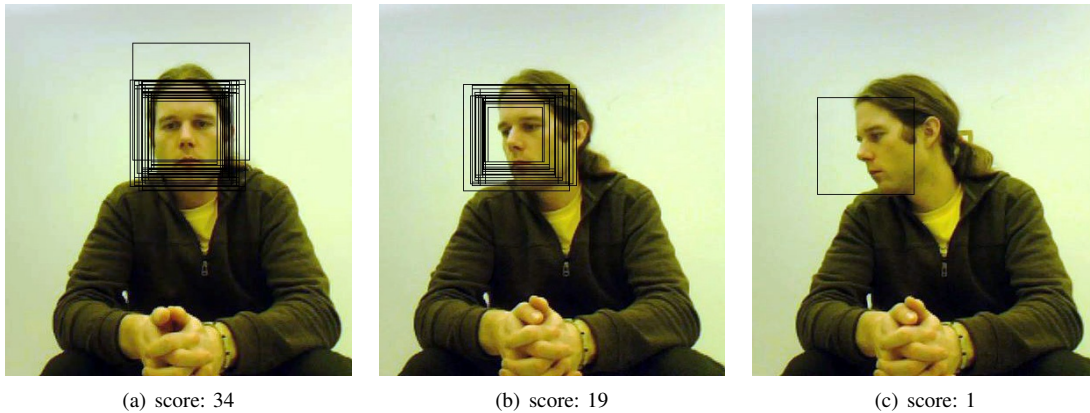Figure 1. An example of a selecting and commanding an individual robot from a group of robots.

(a) score: 34   (b) score: 19   (c) score: 1

Figure 2. The number of candidate rectangles, detected by the OpenCV Haar classifier cascade, is used to "score" how likely the face is a frontal face. A user looking directly at the robot, seen in (a), receives the highest score.

Various views of a user as captured by each of the three robots are displayed in Fig 2.

The robots run a distributed leader election algorithm [6] over the wireless network to designate a single robot as the user-selected robot as determined by the highest detected face score. The elected robot begins to glow to indicate it is currently selected and is ready for a command.

In our demonstration, a motion-based hand gesture is detected by a learned classifier [7] and used to direct the selected robot to autonomously drive to one of the two target zones. On reaching the zone, the robot then returns to the human for further instructions. An overview of the vision-related processing is shown in Fig 3.

We suggest this approach is a very natural and simple way to designate and command robots, as illustrated in the video.

ACKNOWLEDGMENT

REFERENCES

[1] A. Kendon, "Some functions of gaze-direction in social interaction," *Acta psychologica*, vol. 26, pp. 22–63, 1967.

[2] E. Goffman, *Behavior in Public Places: Notes on the Social Organization of Gatherings*. Free Press, September 1966.

[3] H. Kato and M. Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," in *IWAR '99: Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*. Washington, DC, USA: IEEE Computer Society, 1999, p. 85.

[4] P. Viola and M. Jones, "Robust real-time face detection," *Int. Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[5] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*. Cambridge, MA: O'Reilly, 2008.
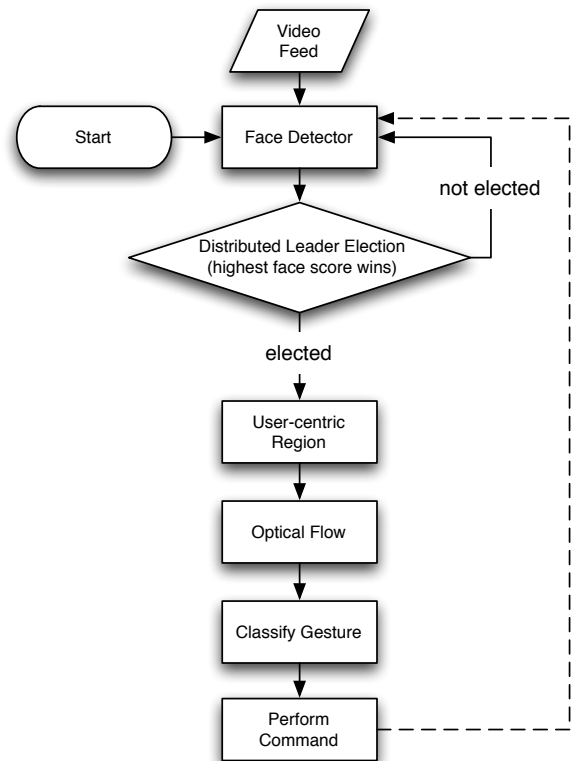
Figure 3. Flowchart of the leader election and motion-based gesture recognition process

[6] E. J. H. Chang and R. Roberts, "An improved algorithm for decentralized extrema-finding in circular configurations of processes," *Communications of the ACM*, vol. 22, no. 5, pp. 281–283, may 1979.

[7] M. Bayazit, A. Couture-Beil, and G. Mori, "Real-time motion-based gesture recognition using the gpu," in *Proceedings of the IAPR Conference on Machine Vision Applications*, May 2009, pp. 9–12.