Selecting and Commanding Individual Robots in a 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

We present a novel real-time computer vision-based system for facilitating interactions between a single human and a multi-robot system: a user first selects an individual robot from a group of robots, by simply looking at it, and then commands the selected robot with a motion-based gesture. Robots estimate which robot the user is looking at by performing a distributed leader election based on the "score" of the detected frontal face.

1 Introduction

Selecting and commanding individual robots in a multi-robot system can be a challenge: interactions typically occur over a conventional human-computer interface (e.g. [19]), or specialized remote control (e.g. [26]). Humans, however, can easily select and command one another in large groups using only eye contact and gestures. Can similar non-verbal communication channels be used for human-robot interactions?

In this work, we describe a novel human-robot interface designed to use face engagement as a method for selecting a particular robot from a group of robots. Face detection is performed by each robot; the resulting score of the detected face is used in a distributed leader election algorithm to guarantee a single robot is selected. Once selected, motion-based gestures are used to assign tasks to the robot. In our demonstration, robots are commanded to drive to one of two predefined locations. An example of a typical interaction is shown in Fig 1; a video demonstration, from which these images originate, can be seen in [4].

The system presented in this paper, which uses face engagement to select a particular robot from a multi-robot system, is the first of its kind.

2 Background

Before two or more people can enter into a focused interaction, they must somehow mutually signal their cognitive focus and readiness. Eye contact and eye gaze play an important role in initiating and regulating communication between people [14]. Throughout this work, we will use the term *face engagement*, as coined by Goffman, to describe the process in which people use eye contact, gaze and facial gestures to interact with or engage each other [9].

The role of eye contact plays such an important role in the development of humans that the ability to detect eye contact is present at birth [8]. We therefore believe that face engagement could be an effective non-verbal communication channel for human-robot interactions.

2.1 Gaze as an input device

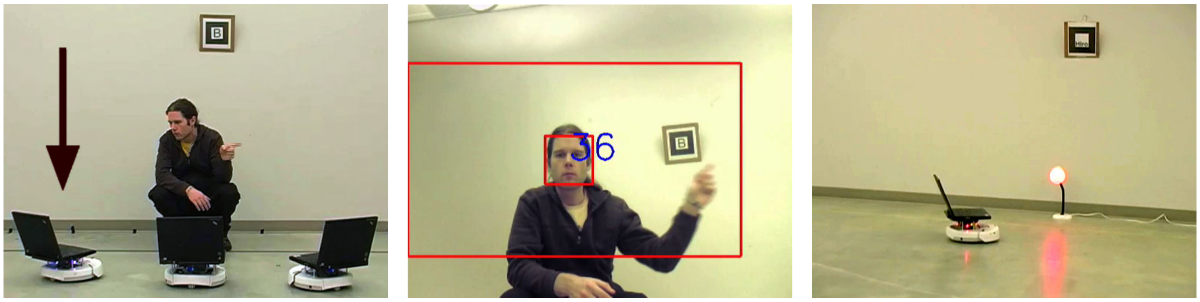
There is a large literature on gaze tracking techniques; Morimoto and Mimica provide an in-depth survey [22]. Applications of gaze trackers can be found in fields ranging from psychology to marketing to computing science; many interesting examples are given in the survey provided by Duchowski [5].

In addition to tracking participants' gazes for subsequent analysis in usability studies, the human-computer interaction (HCI) community has studied using eye gaze tracking devices as hands-free real-time input devices (e.g. [11, 30]).

2.2 Gaze and interactive robots

Researchers argue that anthropomorphizing robots, and therefore exploiting human familiarity, will lead to more natural human-robot interactions; however *too* much anthropomorphization may lead to unrealistic prior expectations [6].

In an experiment by Mutlu et al., gaze is used to regulate conversations between Robovie, a humanoid robot with



(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 selecting and commanding an individual robot from a group of robots.

two controllable eyes[1], and two human participants. Their study showed that a) participants who made eye contact with Robovie liked the robot significantly more than those who were never acknowledged by Robovie's gaze, and b) gaze was an effective tool for yielding speaking turns and reinforcing conversation roles [23].

Besides yielding speaking roles and regulating conversation, gaze can also be used to establish joint attention between a speaker and addressee. The experiments of computation linguists Staudte and Crocker showed that people's cognitive response times increased when a robot used both gaze and speech to refer to objects presented on a table; however, when the robot was programmed to gaze incorrectly (at an irrelavant object), response times were significantly slower than when only speech was used [31]. Similar work by Mutlu et al. showed that participants were capable of picking up nonverbal leakage, that is, seemingly unintentional cues containing information, in a guessing game between a single human and Robovie [24].

Kuno et al. present a museum tour-guide that only responds when directly looked at [16]. Rather than truly performing gaze detection, their "eye-contact" detector, or perhaps what should be referred to as a "face engagement" detector relies on the detection of frontal faces. A telephoto lens is used to capture a high quality image; the robot then estimates if the user is looking at it by detecting if the nostrils are centered between the eyes. A similar method is used by Yonezawa et al. which detects the positions of a user's pupils before responding to voice commands [35].

Literature on eye gaze or face engagement aware human-robot interfaces is limited. While some of the robots discussed here only respond when looked at, it is not absolutely clear how precise their gaze tracking system is or how well it would fare in multi robot situations.

[1]Robovie was developed by Ishiguro et al. at ATR [10]

2.3 Robot selection and task delegation

There is little work on human-robot interfaces for multi-robot systems. Examples can be broken up into two general cases:

2.3.1 World-embodied interactions

World-embodied interactions occur directly between the human and robot, through either physical or sensor-mediated interfaces. These interfaces allow the user to walk freely among the robots, and does not require any form of robot localization. Examples include work by Payton that uses an omnidirectional IR LED to broadcast messages to all robots, and a narrow, directional IR LED to select and command individual robots [26], work by Naghsh et al. present a similar system designed for firefighters, but does not discuss selecting individual robots [25], and work by Zhao et al. which proposes the user interacts with the environment by leaving fiducial-based "notes" (for example, "vacuum the floor" or "mop the floor") for the robots at work site locations [36].

2.3.2 Traditional human-computer interfaces

Rather than interacting directly with robots, a traditional human-computer interface is used to represent the spatial configuration of the robots and allow the user to remotely interact with the robots. Examples of human-robot interactions which occur through a traditional interface include work by McLurkin et al. that presents a overhead-view of the swarm in a traditional point and click GUI named "SwarmCraft" [19], and work by Kato that displays an overhead live video feed of the system on an interactive multi-touch computer table, which users can control the robots' paths by drawing a vector field over top of the world [13].

2.4 Gesture-based robot interaction

There is a vast computer vision literature on the gesture recognition domain: Mitra and Acharya [21] provide a survey. Several gesture-based robot interfaces exist; we do not attempt to provide an exhaustive survey, but rather mention some interesting examples. Systems may use static gestures – where the user holds a certain pose or configuration – or dynamic gestures – where the user performs a combination of actions.

Waldheer at al. use both static and motion-based gestures to control a trash-collecting robot [33]. Loper et al. demonstrate a indoor/outdoor person-following robot that uses an active depth sensing camera to recognize static gestures [17]. Earlier work by Kortenkamp et al. presents a mobile robot that uses an active vision system to recognize static gestures by building a skeleton model of the human operator; a vector of the human's arm is used to direct the robot to a particular point [15]. Perzanowski et al. present a multimodal speech and gesture-based interface; an active vision system is used to interpret pointing gestures as directional vectors, and to measure distance between the user's two hands [28]. In a subsequent paper, Perzanowski et al. discuss the idea of using gaze for directing an utterance at a particular robot; however instead they choose to use a unique name to verbally select a robot [27].

All gesture-based systems discussed so far are designed to work with a single robot, with exception of the work of Perzanowski et al.; however, there are no examples of gesture-based interfaces designed for multi-robot systems which rely solely on non-verbal communication. In this paper, we present such a system: a user first selects an individual robot with face engagement, then uses motion-based gestures to command it. Our novel system allows a user to interact with multiple robots in a shared environment by only using visual cues.

3 Robot selection

Before assigning a task to an individual robot, the human operator must first somehow designate a particular robot of interest as the selected robot he or she will be addressing. We will refer to this as the *robot selection problem*: how does a user interact with a particular robot within a group of robots without accidentally selecting or issuing commands to multiple robots.

The difficulty of the robot selection problem depends on the particular human-robot interface of the system. For example, interfaces that have physical buttons or touch screens located on each robot are immune to the problem since there is no disambiguation when the user issues a command to the robot. In this case a private communication channel exists between the human and each robot since the user must physically approach and touch each robot. However, systems that do not have a private communication channel for each robot and rely on broadcasting commands through a shared medium are susceptible to the robot selection problem. These media include audio, infrared, radio and vision. Most systems assign a unique name or identifier which can be used to specify which robot the message is intended for. The Internet, for example, uses IP addresses to deliver a message over a shared medium to a particular computer; however, while IP addresses are easy for computer-to-computer communication, long unique identifiers are not appropriate for HRI as suggested by Fig 2. Assigning names to each robot (akin to the idea of hostnames) would provide a more usable interface; however users would still have to learn each robot's name.

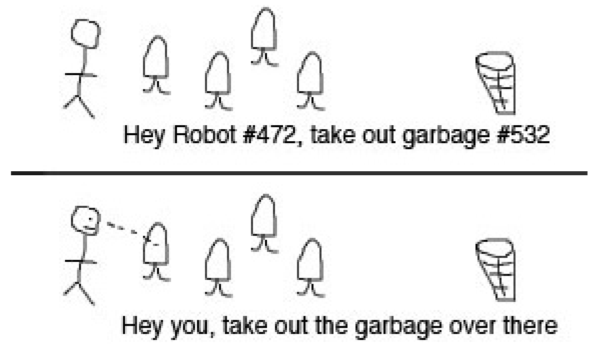
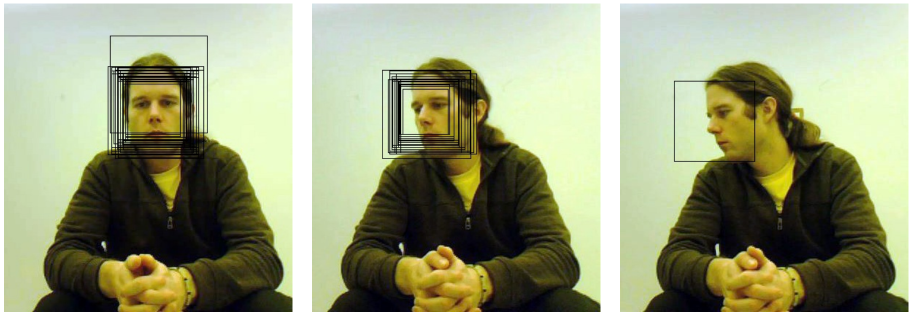


Figure 2. We suggest using face engagement is much more natural than a unique identifier

Our approach to the robot selection problem is focused on maintaining face to face communication between a human and an individual robot. Face engagement serves as the means for designating *that* robot (the one the user is looking at) as the selected robot. However, since face engagement occurs in a shared communication channel between the user and all robots within line of sight, the robots must collectively agree upon a single robot designation to ensure only one robot will ever respond to the user at any given time.

The human operator should be able to select and interact with a robot at different distances; however, implementing eye gaze on a mobile robot for use at larger distances can be a costly endeavour since the use of a telephoto lens or high resolution camera must be used to capture a high quality image of the human's eyes [34]. Our system, on the other hand, use face detection rather than estimating eye gaze; this allows us to use smaller (and cheaper) cameras without zooming capabilities. Our system assumes only a single human will be interacting with the system at any given time; however, this single human will be simultaneously visible to multiple robots. Our system is designed to work at distances varying from 1 to 4 meters. The challenging aspect of our proposed solution to the robot selection problem is disam-



(a) front - score: 34 (b) 45 degree - score: 19 (c) side - score: 1

Figure 3. Candidate rectangles detected by the OpenCV Haar classifier cascade for frontal faces. The number of candidate rectangles are used to indicate how likely the face is a frontal face.

biguating which robot is currently being looked at through means of a distributed leader election algorithm based on the score of the detected face.

3.1 Face detection

The first phase of robot selection involves face detection. Each robot is equipped with a Lenovo ThinkPad R61 7744 laptop with an Intel Core 2 Duo 2.2GHz dual-core processor and 2GB of memory; we use the built in 640x480 resolution video-camera to capture images. Given an image such as the one presented in Fig 3(a), we are interested in locating a rectangular region in the image that contains a face. Furthermore, we want to extract a corresponding score indicating how likely is it that a frontal face has been detected.

Faces are detected with the Viola-Jones method [32]. We use an implementation provided by the OpenCV software library [2].

3.2 Face score

The face detector is trained on frontal faces only[2]. Therefore, the best matches occur when the detected face is looking directly at the camera. Since the face detector is insensitive to small changes in scale or position, multiple sub-windows are often clustered around faces. We use the number of neighbouring sub-windows in each cluster as a score to assess the quality of the detected face[3]. The score, however, does not necessarily indicate how frontal the face is. An obscured frontal face, for example, may receive a lower score than a visible and well lit non-frontal face. However, if the *same* face is captured simultaneously by multiple cameras (and thus under the same lighting conditions), then the scores *can* be used to detect the most frontal face. This observation is a novel contribution of this work.

Fig 3 provides an example of three different images of a person looking in three different directions. A frontal face is captured in the first image (Fig 3(a)) which has the highest score; as the person looks away from the camera the score decreases. In the extreme case where only a profile of the face is captured, the face is barely detected.

3.3 Leader election

The second phase of our solution to the robot selection problem is to perform a distributed leader election algorithm; this ensures only a single robot will ever be designated as the selected robot. The election determines which robot is most likely being looked at "head-on" by the user, as estimated by the highest detected face score.

Since the user might be visible to multiple robots, it is crucial that only a single robot ever respond to the user at any given time. This in effect requires some form of mutual exclusion among the robots, which are hereafter referred to as nodes. To solve this problem we use a variation of the ring-based election algorithm first described by Chang and Roberts [3]. Each node is assigned a unique IP address, hereafter referred to as UID, and is located in a virtual ring comprised of all other nodes. Each node creates tuple $(UID, score)$ and forwards it to its neighbour. When a tuple is received at a node it either:

1. recognizes itself as the elected node if the received tuple contains its own UID,
2. passes the unmodified tuple to the next node if the contained score is greater than its own, or
3. replaces the contents of the tuple with its own score and UID if the contained score is less than or equal to its own.

[2] We use the pretrained frontal face Haar classifier supplied with OpenCV

[3] The Viola-Jones classifier score could also be used; however, the OpenCV implementation makes it difficult to retrieve, and using multiple detections is arguably more robust.

Even though our network is totally ordered, and could break any ties by comparing the UIDs, we choose to replace tuples with *equal* scores, therefore resulting in no elected robot. We do this to force the user to move closer to the intended robot, thus selecting the robot the user really wanted rather than arbitrarily breaking the tie.

4 Gesture recognition

Once a robot has been selected by the user (thus winning the election), it can then be commanded by the user with motion-based gestures. Our classifier uses motion cues to discriminate between different gestures. Examples of the set of gestures used to command the robots are shown in Fig 4.

A detailed description of our classifier, with experimental results, is presented in [1]. Our algorithm is summarized:

1. Motion features are first calculated by computing the optical flow for each frame. The optical flow vector field F is then split into four non-negative channels F_{x^+}, F_{x^-}, F_{y^+}, F_{y^-} representing the half-wave rectified horizontal and vertical components of the flow; this process is similar to Efros et al. [7]. These channels are first box-filtered, to reduce sensitivity to small translations, then aggregated over a temporal history of the last k frames, for some k which is large enough to capture all frames from a gesture period[4].

2. Face detection is used to create a normalized, user-centric view; motion features within this user-centric rectangle are cropped and resized to 30×40; all channels are then flattened into a single vector v.

3. The aforementioned motion features describe the user's entire motion. Given the labelled training data, we have a multi-class classification problem. Using the multi-class boosting algorithm AdaBoost.MH [29], we learn a discriminative classifier that only uses a subset of the motion feature vector v.

A schematic summary of the leader election and motion-based gesture recognition algorithms is provided in Fig 5.

5 The robots

We use three modified iRobot Create robots, pictured in Fig 1(a), which feature six IR range sensors, five colourful RGB LEDs, and a single-board Gumstix computer with

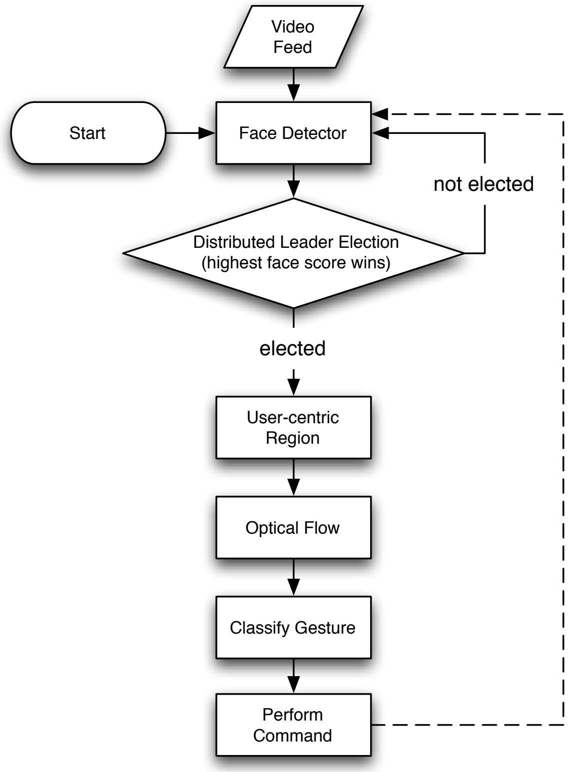


Figure 5. Flowchart of the leader election and motion-based gesture recognition process

an 802.11 wireless network adapter. The modified Creates, hereafter referred to as Chatterbox robots, were designed and built by the Autonomy Lab at SFU[5]. A laptop is mounted to each robot for video capture and processing as discussed in section 3.1.

5.1 Demonstration task

To demonstrate our system, we perform a robot navigation task. Three robots and a human operator are located in a 7x10 meter room clear of static obstacles. Robots navigate around the user and each other using the nearness diagram obstacle avoidance algorithm [20]. The robots: 1) first approach the user, who is located at a predefined location, 2) wait to be selected by the user. Once a robot has been selected by the user, it begins to glow random colours; the selected robot then 3) receive a command, and 4) travel to a predefined zone which corresponds to the issued command.

Robots either travel to the *red* zone or a *green* zone which corresponds to the received gesture: *wave-left* or *wave-right* respectively. Upon reaching the two meter wide circular zone, each robot then return to the user to await a further

[4] In practice, we set k to be the FPS of our capture data, i.e., the number of frames required to capture one second of history.

[5] See `http://autonomy.cs.sfu.ca/robots.html` for more information.

(a) wave-left (b) wave-right (c) idle

Figure 4. Example frames from our robot-command dataset used for training

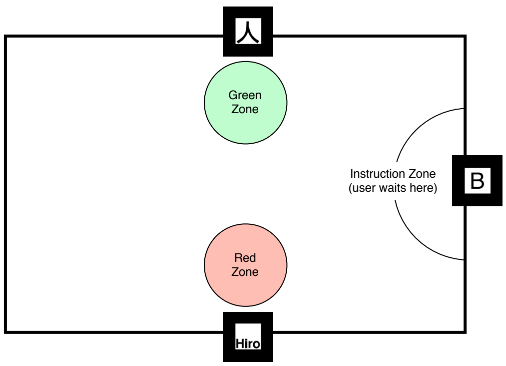


Figure 6. Robots wait at the instruction zone until they are selected and commanded to drive to either the red or green zone. Fiducial markers are placed on the walls near each zone.

command. Three unique ARtoolKit fiducial markers [12] provide global robot localization. An overview of the zones and room layout is shown in Fig 6.

6 Discussion

We have tested our system with 7 participants who were not involved with the development of the system. A single participant interacted with the robots at any given time. Each participant was instructed to:

1. first select a robot by looking at it, then, once the robot starts to glow;
2. direct the robot to one of two zones: a green zone, by waving your right hand, or a red zone, by waving your left hand.

Each participant was asked to command two robots to the same zone, and the third to the other zone. Once the robots reached their goal zone, they returned to the user. Participants were then encouraged to assign new tasks to returning robots as they saw fit. Some participants waited for all three robots to return before assigning tasks, where as others decided to immediately assign new tasks as each robot arrived.

Throughout the demonstration, a total number of 77 tasks were assigned to the robots. The next two sections are an informal description and discussion of the performance of the human-robot interface and is split up between robot selection, and task designation. A comprehensive user study will be the subject of future work.

6.1 Robot selection

Our leader election algorithm performed as intended: only a single robot ever responded, by glowing, at a given time. In some cases no robots were elected due to equal face scores resulting in a tie. In these cases, we encouraged participants to reposition themselves to break the tie. Ties occurred when two robots awaiting instructions were located very close to each other; however, once the user approached a particular robot, the angle at which the user had to turn his or her head increased, resulting in a single robot seeing a full frontal face. In some cases, rather than re-engage one of the two side by side robots, users appeared to be discouraged by a tie and simply tried to select a third robot which they did not originally intend to select.

The face detector implementation provided with OpenCV worked well provided the faces were not too far away from the camera. In our demonstration, the user was at most 4 meters away from the camera and was in a well lit environment. Even with the camera pointing upwards towards the overhead lights, the system was still able to detect faces.

To provide visual feedback, the selected robot would glow with randomly alternating colours. Unfortunately the laptop partly covered the LED which made it hard to see for taller participants. We tried using the laptop's screen as a giant "LED" to provide feedback, but users reported that it was not as satisfying as the glowing LEDs.

Initially, two of the seven participants were unsure which robot was selected and issued commands before any robot had started to glow. However, after we encouraged them to move closer to the robot (and into its video-frame), these two participants were able to command the robots. In other cases, the robots were too close to both the wall and the user, ultimately forcing the user's back up against the wall. These localization-related problems forced the participants to squat down in order to be in the robot's field of view.

6.2 Task designation

Participants then assigned tasks to the selected robot with a motion-based hand gesture. We explicitly demonstrated the two gestures: left hand waving, and right hand waving, hereby referred to as wave-left and wave-right respectively. Using hand waving gestures to assign robots location dependent tasks proved to be challenging in three cases:

1. two of our participants, at first, extended their hands to point left or right rather than wave their hands in a continuous motion,
2. the full hand waving motion of participants who were located too close to the robot, was never captured by the video camera, and
3. the one second optical flow history window required for gesture recognition gave the interface a slow feel.

The accuracy of the system was good: 74 out of 77 commands were correctly executed. The 3 errors occurred when a user issued a command to a robot, and then quickly selected a different robot. This resulted in the newly selected robot classifying the previously issued gesture based on the motion features stored in the optical flow history window. This unintended behaviour could be remedied by reinitializing the optical flow history window whenever a robot is *not* elected.

To avoid classification errors, robots used a high classification threshold. Choosing a high threshold gives a high level of precision, which prevents the robots from incorrectly classifying a command resulting in opposite behaviour; however, setting the threshold too high limits our level of recall which became an irritant to some participants. After some exposure to the system participants were able to fine tune their gestures to achieve quicker recognition. Providing some sort of feedback mechanism may have decreased the interface's learning curve.

7 Conclusion

In this paper, we presented a computer vision-based human-robot interface for selecting and commanding an individual robot from a multi-robot system. A user first selects a robot with face engagement by simply looking at it. We employed a standard frontal face detector to detect the user's face. The detected-face score of each robot is used in a distributed leader election algorithm to guarantee at most a single robot is selected. Once a robot has been selected by the user, it can then be commanded by using a motion-based gesture. We retrained a previously developed real-time classifier which uses motion-cues to discriminate between gestures corresponding to robot commands.

A demonstration task was described to investigate the feasibility of using face engagement and motion-based gestures for commanding an individual robot in a multi-robot system. Our demonstration showed that our face engagement-based leader-election could be effectively used to select an individual robot, which could then be commanded with motion-based gestures.

7.1 Future work

A proper user-study with a larger number of participants would be the next step for evaluating the system; however, the observations so far suggest some useful improvements: the human-robot interface could first be improved by providing a better feedback mechanism. The use of LEDs works well for quickly determining the current robot state; however, it would be valuable to see how users respond to an anthropomorphized robot with eyes. This could easily be implemented with virtual eyes on the laptop screen.

An extension to this system would be to allow users to first select a subset of the robots. The set of gestures could also be extended to allow a user to point to *any* arbitrary place in the environment, and have the robots drive to that location. This has been done for a single robot system (e.g. [15, 18]); however, a challenging task would be to coordinate multiple robots to cooperatively estimate the vector given the system's ability to simultaneously capture images of the user from multiple angles.

References

[1] M. Bayazit, A. Couture-Beil, and G. Mori. Real-time motion-based gesture recognition using the gpu. In *Proc. of the IAPR conf. on Machine Vision Applications*, pages 9–12, May 2009.

[2] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, 2008.

[3] E. J. H. Chang and R. Roberts. An improved algorithm for decentralized extrema-finding in circular configurations of processes. *Communications of the ACM*, 22(5):281–283, May 1979.

[4] A. Couture-Beil, R. T. Vaughan, and G. Mori. Selecting and commanding individual robots in a vision-based multi-robot system. In *HRI '10: Proc. of the 5th ACM/IEEE intl. conf. on Human robot interaction*, page (to appear), 2010.

[5] A. T. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, and Computers*, 34(4):455–470, 2002.

[6] B. R. Duffy. Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3–4):177–190, 2003.

[7] A. A. Efros, A. C. Berg, G. Mori, and J. Malik. Recognizing action at a distance. In *Proc. 9th Int. Conf. Computer Vision*, volume 2, pages 726–733, 2003.

[8] T. Farroni, G. Csibra, F. Simion, and M. H. Johnson. Eye contact detection in humans from birth. In *Proc. of the National Academy of Sciences of the United States of America*, pages 9602–9605, 1999.

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

[10] H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda, and R. Nakatsu. Robovie: an interactive humanoid robot. *Industrial robot: An intl. journal*, 28(6):498–503, 2001.

[11] R. J. K. Jacob and K. S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, pages 573–603, 2003.

[12] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proc. of the 2nd IEEE and ACM intl. Workshop on Augmented Reality*, page 85. IEEE Computer Society, 1999.

[13] J. Kato, D. Sakamoto, M. Inami, and T. Igarashi. Multi-touch interface for controlling multiple mobile robots. In *Proc. of the 27th intl. conf. extended abstracts on Human factors in computing systems*, pages 3443–3448. ACM, 2009.

[14] A. Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63, 1967.

[15] D. Kortenkamp, E. Huber, R. P. Bonasso, and M. Inc. Recognizing and interpreting gestures on a mobile robot. In *Proc. of the Thirteenth National conf. on Artificial Intelligence*, pages 915–921. AAAI Press/The MIT Press, 1996.

[16] Y. Kuno, M. Kawashima, K. Yamazaki, and A. Yamazaki. Importance of vision in human-robot communication understanding speech using robot vision and demonstrating proper actions to human vision. In *Intelligent Environments*, Advanced Information and Knowledge Processing, pages 183–202. Springer London, 2009.

[17] M. M. Loper, N. P. Koenig, S. H. Chernova, C. V. Jones, and O. C. Jenkins. Mobile human-robot teaming with environmental tolerance. In *HRI '09: Proc. of the 4th ACM/IEEE intl. conf. on Human robot interaction*, pages 157–164. ACM, 2009.

[18] C. Martin, F.-F. Steege, and H.-M. Gross. Estimation of pointing poses for visual instructing mobile robots under real-world conditions. *Robotics and Autonomous Systems*, 57:(to appear), 2009.

[19] J. McLurkin, J. Smith, J. Frankel, D. Sotkowitz, D. Blau, and B. Schmidt. Speaking swarmish: Human-Robot interface design for large swarms of autonomous mobile robots. In *Proc. of the AAAI Spring Symposium*, 2006.

[20] J. Minguez, A. Member, and L. Montano. Nearness diagram (nd) navigation: Collision avoidance in troublesome scenarios. *IEEE Transactions on Robotics and Automation*, 20:2004, 2004.

[21] S. Mitra and T. Acharya. Gesture recognition: A survey. *IEEE Transactions On Systems, Man, And Cybernetics - Part C: Applications And Reviews*, 37(3):311–324, May 2007.

[22] C. H. Morimoto and M. R. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, 2005.

[23] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *HRI '09: Proc. of the 4th ACM/IEEE intl. conf. on Human robot interaction*, pages 61–68. ACM, 2009.

[24] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In *HRI '09: Proc. of the 4th ACM/IEEE intl. conf. on Human robot interaction*, pages 69–76. ACM, 2009.

[25] A. M. Naghsh, J. Gancet, A. Tanoto, and C. Roast. Analysis and design of human-robot swarm interaction in firefighting. In *IEEE intl. Symposium on Robot and Human Interactive Communication (RO-MAN'08)*, pages 255–260, 2008.

[26] D. Payton. Pheromone robotics. `http://www.swarm-robotics.org/SAB04/presentations/payton-review.pdf`, 2004. Slides from a presentation at the Swarm Robotics Workshop, SAB04. Retrieved September 28, 2009.

[27] D. Perzanowski, A. C. Schultz, W. Adams, M. Bugajska, E. Marsh, G. Trafton, D. Brock, M. Skubic, and M. Abramson. Communicating with teams of cooperative robots. In *Proc. from the 2002 NRL Workshop on Multi-Robot Systems*, pages 16–20. Kluwer, 2002.

[28] D. Perzanowski, A. C. Schultz, W. Adams, E. Marsh, and M. Bugajska. Building a multimodal human-robot interface. *Intelligent Systems, IEEE*, 16(1):16–21, Jan-Feb 2001.

[29] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3):297–336, 1999.

[30] I. Starker and R. A. Bolt. A gaze-responsive self-disclosing display. In *CHI '90: Proc. of the SIGCHI conf. on Human factors in computing systems*, pages 3–10. ACM, 1990.

[31] M. Staudte and M. W. Crocker. Visual attention in spoken human-robot interaction. In *HRI '09: Proc. of the 4th ACM/IEEE intl. conf. on Human robot interaction*, pages 77–84. ACM, 2009.

[32] P. Viola and M. Jones. Robust real-time face detection. *intl. Journal of Computer Vision*, 57(2):137–154, 2004.

[33] S. Waldherr, R. Romero, and S. Thrun. A gesture based interface for human-robot interaction. *Autonomous Robots*, 9:151–173, 2000.

[34] H. Yamazoe, A. Utsumi, T. Yonezawa, and S. Abe. Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions. In *ETRA '08: Proc. of the 2008 symposium on Eye tracking research & applications*, pages 245–250. ACM, 2008.

[35] T. Yonezawa, H. Yamazoe, A. Utsumi, and S. Abe. Evaluating crossmodal awareness of daily-partner robot to user's behaviors with gaze and utterance detection. In *Casemans '09: Proc. of the 3rd ACM intl. Workshop on Context-Awareness for Self-Managing Systems*, pages 1–8. ACM, 2009.

[36] S. Zhao, K. Nakamura, K. Ishii, and T. Igarashi. Magic cards: a paper tag interface for implicit robot control. In *CHI '09: Proc. of the 27th intl. conf. on Human factors in computing systems*, pages 173–182. ACM, 2009.