An Active Vision System for a Social Robot

Cynthia Breazeal, Paul Fitzpatrick & Brian Scassellati

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: If a robot is intended to interact with people, it needs an active vision system that can serve both a perceptual and communicative function.

Motivation: Much of human social behavior is influenced by the directional, foveated nature of our visual perception. We have developed characteristic eye and head movements to serve the functional requirements of acquiring information through vision, and we intuitively "read" those movements in others. A humanoid robot whose vision system follows the human organization can enter into this implicit protocol. This introduces an extra dimension to the control problem, where we must design for the social significance that people will assign to the robot's behavior – such as making and breaking eye contact during conversational turn-taking.

Previous Work: Models of attention are fundamental to this work. Attention has been an area of interest both to computational researchers and psychophysics researchers. In the computational fields, other researchers have focused on developing systems that assign saliency based purely on perceptual properties such as color, motion, and orientation [1]. Psychophysicists have studied the "pop-out" effects of raw sensory saliency cues [2] and the effects of subject motivation and goals on attention [3]. Wolfe's *Guided Search v2.0* system was proposed as a model for human visual search behavior [4], and this is the model on which we have based our work.

Approach: We have implemented a version of Wolfe's model for visual search behavior on the social robot Kismet [5]. Incoming video signals are filtered by some basic feature detectors that locate regions of possible interest. An attention system factors in the robot's motivational state to pick out behaviorally relevant regions. A closely integrated tracking system introduces persistence, so that the robot's behavior has time to organize around the output of the attention system. The robot's gaze and head orientation are usually driven by the tracker, unless explicitly inhibited from doing so by the behavior system. Gaze direction is an important cue for reading the robot's locus of attention and "intent". Control of gaze is blended with expressive movements used to reflect the robot's "emotional" state (which reflects its appraisal of its current situation in terms of its current motivation). Extra visual measurements, such as distance and presence of eyes, are made for the target of the robot's gaze which are computationally expensive to make across the entire image. In general the visual system follows a broad division into pre– and post–attentive processing analogous to human vision.

We have implemented a number of behaviors that recruit natural social cues to simplify visual processing through the regulation of interpersonal distance and the manner of object presentation.

Difficulty: The chief difficulty of this work is that the vision system must be simultaneously broad enough to support all the requirements of the robot's behavior, sufficiently integrated to keep that behavior coherent, and fast enough to run in real-time. Also, motor control of eye movements has to be done very carefully, or they can be extremely disturbing – for example, poorly coordinated disjunctive movements can easily give the impression that the eyes are moving independently.

Impact: This work, in conjunction with similar efforts in the auditory domain, provides the perceptual substrate for building "social creatures" [6].

Future Work: The range of the robot's interaction with humans is limited by many factors, including its visual perception. As new percepts are added, new depths of behavior are possible. One example is face recognition, which provides the means for the robot to engage in extended relationships with particular people. Object-centered percepts such as color, shape, and material give the robot the means to respond differentially to particular objects, and greatly extend the types of behavior in which the robot could engage.

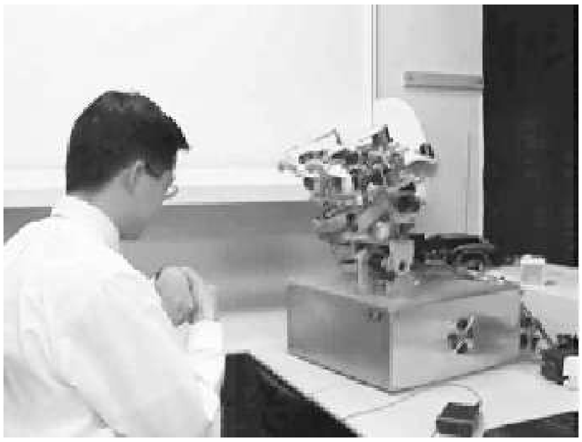

Figure 1: Vision in a social context.

Research Support: Parts of this research are funded by DARPA/ITO under contract number DABT 63-99-1-0012 and parts have been funded by ONR under contract number N00014-95-1-0600, "A Trainable Modular Vision System."

References:

[1] L. Itti, C. Koch, and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE PAMI, 20(11):1254-1259, 1998.

[2] A. Triesman. Features and Objects in Visual Processing. *Scientific American*, 1986, **255**, pp. 114B–125.

[3] P. Niedenthal and S. Kityama. The Heart's Eye: Emotional influences in Perception and Attention. *Academic Press*, 1994.

[4] J. Wolfe. Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1(2):202-238, 1994.

[5] C. Breazeal and B. Scassellati. A context-dependent attention system for a social robot. Proceedings of IJCAI-99, pp. 1146-1151, 1999.

[6] C. Breazeal Learning social behaviors during human-robot play AI Lab Abstract 2000.