RUBI: A Robotic Platform for Real-time Social Interaction

Bret Fortenberry, Joel Chenu, Dan Eaton, Javier R. Movellan

Institute for Neural Computation
University of California San Diego
{bret, joel, deaton, movellan}@mplab.ucsd.edu

Abstract

The majority of our waking hours are spent engaging in social interactions. Some of these interactions occur at the level of long-term strategic planning while others take place at faster time scales, such as in conversations or card games. The ability to perceive subtle gestural, postural, and facial cues, in addition to verbal language, in real-time is a critical component of social interaction. An understanding of the underlying perceptual primitives that support this kind of real-time social cognition is key to understanding social development.

This paper presents a humanoid robot designed for research on real-time social interaction between robots and humans. We discuss many aspects of the current system including motor control, face tracking, and speech recognition and how it is utilized for human to robot social interaction. We also describe plans to increase the system's capabilities and communication skills. Last, we describe planned research for the study of real-time social interactions.

Figure 1. The current appearance of RUBI was chosen by Kai Movellan, the 2 year old child on the picture. Kai refused to interact with the early versions of RUBI and proved to be a wonderful critic for the design team.

1. Introduction

Robots present an ideal opportunity to study the development of social interaction in infants and children [2]. It is possible to create robots that exhibit precisely controlled contingency structures. By observing how infants interact with these robots we gain an opportunity to understand how infants identify the operating characteristics of the social agents with whom they interact. In this paper we present progress on the development of a social interaction robot, 'RUBI', designed to communicate and interact with children and to serve as a platform on experiments for social interaction and social development with children.

2 System Architecture

2.1 Robot Structure

RUBI is a three foot tall, pleasantly plump robot with a head and two arms (See Figure 1). It stands on four non-motorized rubber wheels for moving it easily from place to place. RUBI is self-contained; all of its components (computer, microphone, speakers, etc.) are inside its body structure. The external connections consist of a single power

cable and a wireless Ethernet card. The computer (specified below) is on a sliding vibration-reducing rack. The head consists of two cameras 13.5 inches apart representing the eyes, and a disk-shaped face with a small button nose and permanent smile. An omni-directional camera is on a rod that extends above the head for a clear view of the world. The body is a wooden night-stand that is spacious enough to hold all of the RUBI's components, but short enough to keep RUBI non-threatening to children.

Pilot experiments with 2-4 year old children helped shape RUBI's physical design. Children were frightened of the original design, but several iterations of experiments helped make it more child-friendly. Some of the most effective changes included adding clothing to cover mechanical parts, giving RUBI a "smile", and making it shorter.

2.2 Motor System

RUBI's head is based on the early Robovie design from Hiroshi Ishiguro's group [5, 4]. It has 9 degrees of freedom: Three degrees of freedom (pan, tilt and roll) implemented by stepper motors that are driven by a Galil DMC-1832 PCI motor controller. The neck can move 54 degrees up and 30 degrees down from center position and 54 left and right of center The maximum pulse rate of the motor drivers is 144,000 degrees/sec, but the motors for RUBI's neck slip at high rates (over 480 degrees/sec). However, we have manually set RUBI's maximum speed to 60 degrees per second; faster motor control is possible but has an unnatural appearance. The remaining six degrees of freedom are in the eyes, both of which have pan, tilt and zoom motors. The eye cameras are SONY EVI-G20 PTZ (Pan-Tilt-Zoom) cameras. Their horizontal range is ± 30 degrees and their vertical range is ± 15 degrees. Maximum speed on both horizontal and vertical axes is 150 degrees/sec. They are controlled via a 9600 bit/sec VISCA-protocol serial connection.

2.3 Vision System

RUBI's vision sensors consist of two SONY EVI-G20 color cameras which are the "eyes". The third input is a low-resolution stationary omni-directional camera acting as RUBI's peripheral vision. All three use a component out and are routed through a quad video splitter that combines the images into one 640x480 image. The single image is then captured via a BT848 video capture card at 30 Hz.

RUBI's two eyes handle the main face tracking tasks. The commands given to the motor controller are determined by a convolutional HMM architecture that combines both a color-based and a Frontal Face-Detector to produce a distribution of possible locations and scales for faces on the image plane [6, 3].

2.3.1 Frontal Face-Detector

The face-detector system is explained in [3]. It can detect over 98% of faces with minimal false-alarms in difficult background conditions running in real time at 30 frames per second. Its main limitation is that it only detects upright frontal faces. Because of this, the frontal face-detector is complemented by a color-based detector that handles non-frontal views. Source code for the face detector is available at *http://kolmogorov.sourceforge.net* as part of the Kolmogorov project.

2.3.2 Color Tracker

The color-based system utilizes a convolutional HMM architecture desribed in [6]. The system uses standard HMM equations to update a probability distribution of 100000 states representing the possible location and scale of faces on the image plane. The color model uses two 1000 bin hue histograms, one for faces and the other for backgrounds. The initial face histogram model is defined a follows

$$p(c) \propto \exp(-\frac{d^2(h, 17.95)}{12.2^2}), \quad (1)$$

where $\mu = 17.95$, $\sigma = 12.2$ and d is the angular distance in degrees, between h and μ. The initial background histogram model is uniform. Each time the face detector finds a face, the color models for faces and background are updated with a weighted average of the current and past histogram. The most probable location and scale is chosen for display and for use by the head controller.

2.3.3 Peripheral Vision

RUBI's peripheral vision is handled by an omni-directional camera. Since the motors have limited range of motion the omni-directional camera vision software uses only a ± 55 degree field of view. The omni-directional camera uses a motion detection and a non-adaptive color model to search for people. The color model is based on statistics of hue of many example faces.

2.3.4 Head Control: Explore Mode

The goal of RUBI's head controller is to maximize the expected number of face detections. To do so it switches back and forth between two operation modes: (1) Explore mode, and (2) Face tracking mode.

While in explore mode RUBI orients towards motion detected in its peripheral vision (with the omni-directional camera). RUBI's head movement is controlled by a stochastic difference equation that favors a combination of continuous trajectories and areas with high motion energy as detected by the peripheral vision system. Once a face is de-

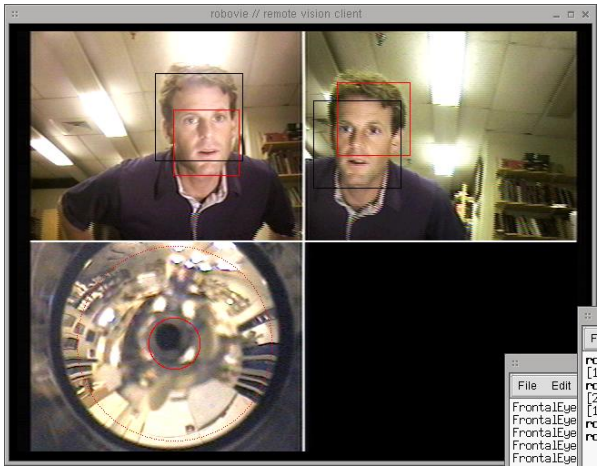

Figure 2. Here are the views that RUBI sees while tracking a face. Top are views from left and right eyes, red box is response from color tracker, black box is response from face-detector. Bottom left is the view of the omni-directional camera.

tected by the face-detector in either eye, RUBI's head controller switches to Face Tracking Mode.

2.3.5 Head Control: Face Tracking Mode

While in face tracking mode, RUBI actively moves its head by means of a Kalman controller whose goal is to minimize the distance from the most-likely probable position of the face and the center of the image plane, average across the two camera views (See Figure 2).

RUBI stays in face tracking mode if a running average $\bar{X}_t$ of the number of faces found by the face detector is above a given threshold. The value of the threshold parameter is dynamically set via reinforcement learning with the goal of optimizing the expected number of faces found by the face detector.

2.4 Auditory and Speech System

RUBI sound sensor is a VoiceTracker 8 microphone array with adaptive beam-forming. Speech detection and speech recognition is handled by the SONIC speech recognition engine from the Center for Spoken Language Research at the University of Colorado Boulder. We are currently training a new noise model to reduce the speech detector's sensitivity to robot motor noise. The detected speech is used to trigger contingent speech-like responses. These speech-like responses consist of baby vocalizations of varying length and pitches. The length of speech spoken to RUBI modifies the length of the response such that longer speech gives longer responses. By changing the characteristics of these responses we can test contingency parameters when interacting with infants in social experiments.

2.5 Social Movements

RUBI combines motor control with three social behaviors: face tracking, speech detection and response, and external environment contingency. RUBI's motor control system currently has 3 components; neck control (azimuth, elevation, roll), eye control (azimuth, elevation, zoom), and control of external objects. In adding to tracking faces, the eyes and head are used to perform social motor actions such as head nodding and gaze shifting. External objects can also be controlled via wireless Ethernet to allow RUBI to respond in a contingent manner to toys or lights when turned on or off. RUBI's architecture combines these smaller action into groups of behaviors and allows for recording and playback of social interactions. Currently RUBI's behaviors are not coordinated for socially interaction. The planned research will determine the timing of switching behaviors and the contingency of motor control for each behavior to optimize social interaction between robot and humans.

2.6 Computational System

RUBI is powered by two dual-processor 2.8 GHz Intel P4 Xeon PCs with 512 RAM connected via gigabit Ethernet. RUBI's PCs use Red Hat Linux 7.3 with open-source drivers. The first PC currently handles the face-detection and color-detection on both eyes, voice detection, and peripheral vision. In the future the first PC will handle the face-detection, color-tracking, periphery vision, and all eye movements (neck movements are handled by the Galil DMC-1832). The second PC will handle speech-detection/recognition, arm/hand movements, and emotion recognition.

3 Current Performance

Currently RUBI is capable of updating the color tracker for each eye in 15 milliseconds and the face detector in 70 milliseconds. To account for both eyes RUBI spends 15 milliseconds correcting for the error in the left eye and 15 milliseconds correcting the error in the right eye. With a frame rate of 30 Hz RUBI is capable of tracking a face at speed up to a 60 degrees/sec.

Figure 3. RUBI in a pilot study with an infant. Here RUBI (near the camera) and the subject attending to a toy.

4 Planned Research

4.1 Autistic Children Study

RUBI is being developed as part of the MESA project sponsored by the National Association for Autism Research (NAAR). The goal is to investigate the effects of contingency and timing on real-time social interaction in typically developing children and in children with autism. Experiments will be conducted in the near future at UCSD's Autism Research Laboratory. The child will be held by a parent to keep her/him comfortable (see figure 3). There will be electronic toys in the room that RUBI can attend to and actively turn off and on via wireless communication. During the course of the experiment we will alter the timing of three different social components. The first is the timing and duration of RUBI attending to the children. The second is the probability of response and timing and length of responses contingent on the children's speech. The third is the probability of response to the behavior of toys, and timing and duration of this response.

4.2 Infant Study

The infant study will have the same setup and environment as the Autistic study. The only difference between the two studies is that the experiment will be run with 18 month old infants. This will minimize the possibility of the participants having prior knowledge or expectations about robots. The infant study will have two control groups: a human (stranger) and an object to replace RUBI. The object will contribute the same level of human-like features as RUBI. RUBI's timing and contingency factors will be altered to find infant responses that are more consistent with the results of the human control study than to the results of the object control study.

5 Conclusion

The goal for the design of RUBI is to create a robot that can test child/infant responses to variances in robot behaviors. We want to determine general robot movements and behaviors that will combine current Artificial Intelligence programs and to create social behaviors that Breazeal refers to as believable behaviors [1]. For robots to work with autistic children, students or the elderly they will need the participants to be socially engaged. From the research we perform using RUBI we hope to gain a better understanding of how to keep people socially engaged while interacting with robots.

References

[1] C. Breazeal. *Designing Sociable Robots*. MIT Press, Cambridge, MA, 2002.

[2] I. Fasel, G. O. Deak, J. Triesch, and J. R. Movellan. Combining embodied models and empirical research for understanding the development of shared attention. In *Proceedings of the second international conference on development and learning*. ICDL, 2002.

[3] Ian Fasel, Bret Fortenberry, and J. R. Movellan. A generative framework for real-time object detection and classification. *Computer Vision and Image Understanding*, In Press.

[4] H. Ishiguro, T. Ono, M. Imai, and T. Kanda. Tdevelopment of an interactive humanoid robor "robovie" – an interdisciplinary approach. In R. A. Jarvis and A. Zelinsky, editors, *Robotics Research*, pages 179–191. Springer, 2003.

[5] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu. Development and evaluation of an interactive humanoid robot "robovie". In *IEEE International Conference on Robotics and Automation*, pages 1848–1855, 2002.

[6] J. R. Movellan, J. Hershey, and Josh Susskind. Large scale convolutional HMMs for real time video tracking. *Computer Vision and Pattern Recognition*, 2004.