Social Tag: Finding the person with the pink hat

Marek Michalowski and **Didac Busquets** and
Carl DiSalvo and **Laura Hiatt** and
Nicholas Melchior and **Reid Simmons**
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Selma Sabanovic
Rensselaer Polytechnic Institute
110 8th Street
Troy, NY 12180

Abstract

At the AAAI 2005 Robot Exhibition, the robot GRACE played a game of "social tag" that involved human-robot social interaction, navigation, and interface design. The task was for GRACE to locate and rendezvous with one of our team members, who was wearing a pink hat. The robot found the target not primarily through the modalities of sight or sound, but rather through social interactions with strangers in the environment.

Introduction

At the AAAI 2005 Robot Exhibition, the robot GRACE (Graduate Robot Attending a ConferencE) played a game of "social tag" in which the task was to locate and rendezvous with a team member who was wearing a pink hat. In this game, our purpose was not simply to create an object localization task (such as a scavenger hunt); rather, it was our intention to create a robot that could enlist the help of humans through frequent social interactions. We designed the game of "social tag" so that the robot's primary source of information about the whereabouts of its target came not through the modalities of sight or sound, but rather through social interactions with strangers in the environment. The task explores issues in human-robot interaction that involve shared space, intuitive interface design, and the negotiation of an environment filled with dynamic, untrained humans.

In this paper, we present our motivation for developing the game of social tag and describe the system we developed to perform this task. Our goal was to build a robot that was social and interactive enough that it was equally capable of requesting and accepting assistance (i.e., either the robot or a human could initiate an interaction that advanced the robot's progress). We present results that suggest that we were successful in this respect.

Background

GRACE

GRACE grew out of a multi-institution collaboration to design a robot capable of performing the AAAI Robot Challenge, which involves autonomously registering for the conference, navigating through the conference area, interacting with people, and delivering a talk. GRACE performed most of these tasks at AAAI 2002 (Simmons *et al.* 2003). The previously developed socially interactive components of GRACE are additionally described in (Bruce, Nourbakhsh, & Simmons 2002) and (Gockley *et al.* 2004). This line of work has been continued in research on designing robots for long-term interaction by incorporating a changing story line in the interactive repertoire of Valerie, a robot receptionist (Gockley *et al.* 2005).

Figure 1: The robot GRACE.

GRACE (Fig. 1) is a RWI B21 mobile robot. For this task, she also included a SICK laser scanner, a Canon VC-C4 PTZ camera, an LCD monitor with an animated face, and an ELO 1224 LCD touch screen. The robot has two computers on board: one for controlling mobility, sensing with the laser, avoiding obstacles, and handling the touch screen interface, and the other for vision, control of the face and voice, and general task control. The software architecture follows the design used in previous years for our entries in the Robot Challenge. Independent processes communicate via IPC message passing (http://www.cs.cmu.edu/∼IPC). No remote communications were used, except for startup and shutdown.

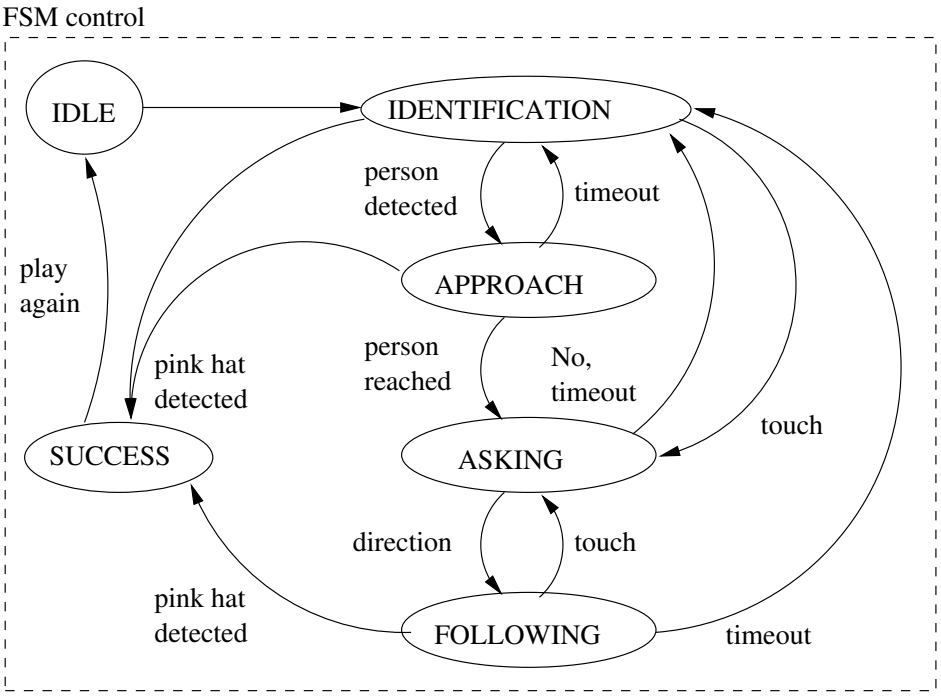

Figure 2: The finite state machine that comprises GRACE's control task for Social Tag.

"Social Tag"

The task of social tag – "finding the person with the pink hat" – was conceived as having a number of important benefits. The task was relatively simple and could be executed reliably, which allowed us to focus on the questions of human-robot interaction. Formulating the task as a fun and commonly-understood game increased the chances that people would be willing to interact with the robot. While these properties were important for designing and observing human-robot interaction, the interaction must serve a purpose; this was to locate the pink hat. The generality of object localization was another important property of our selected task, as it involves search, planning, navigation, etc., and can be applied to a wide range of useful robotic tasks.

In typical robot tasks involving detection of a visual target, distinctively colored objects such as pink hats are used to simplify the vision problem. However, in this case, the pink hat was intended to be as much for the benefit of other people as for GRACE herself. The team member was to be a highly visible individual who would be easily recognized and remembered by conference participants, so that it would be easy for them to help a wandering robot. Accordingly, it was not the completion of the goal (finding the pink hat) in which we were most interested, but rather in GRACE's journey and her social interactions along the way.

From a machine perception standpoint, the task of finding a hat in a crowded room is an extremely difficult one. While research in computer vision and speech recognition continues to improve robot perception, GRACE relies on the assistance of humans with fully developed senses of sight and hearing, and the ability to communicate. Indeed, there are cases in which no amount of sensory ability would enable successful task completion; if the hat is in a different room, even a human would need to enlist the help of other humans who are moving through the environment, communicating with each other, and collectively remembering the location of prominent objects. GRACE depends on her own sense of vision for the completion of the task (i.e. recognizing the pink hat), but her primary mode of gathering information is asking people for help in an intuitive and socially acceptable manner.

The task has five main phases:

1. *Identification* of approachable humans.
2. *Approach* toward a human with whom the robot would like to interact.
3. *Asking* for directions to the person with the pink hat.
4. *Following* those directions until a pink hat is found visually or more help is required.
5. Demonstrate *success* when the hat has been found.

As in typical human interactions, these phases may be regarded as a "script" that suggests an appropriate plan of action. However, the robot's human partner in the interaction may interfere or deviate from this script, and the robot should be able to handle such deviations in a context-dependent manner. Specifically, humans may volunteer to offer help when the robot is not actively looking for it, and GRACE can accept these offers during various phases of performing the task; this is a form of mixed-initiative interaction.

Implementation

Interaction and interface design

An LCD touch screen mediated the interaction between GRACE and conference attendees. The touch screen was mounted on the front of GRACE, below the screen that displayed her face, at approximately chest level with an average-sized adult.

The touch screen interaction consisted of a number of full-screen interfaces that both conveyed the state of the robot and prompted certain types of input. The general approach to designing the screens was to make them as bold and simple as possible while still being aesthetically appealing. To this end, we limited the amount of text on the screen and supplemented the screens with spoken information (via a text-to-speech system with lip-synchronization on the animated face). In order that GRACE's verbalizations do not become repetitive or boring, she chooses randomly from a library of appropriate phrases at each point in the task.

Music is another important component of GRACE's repertoire. For example, when GRACE begins looking for the person in the pink hat, she plays an excerpt from the theme music to "The Good, The Bad, and The Ugly." In addition to the music adding to the playful character of GRACE, it also functions as another mode of (auditory) expression to participants, emphasizing the current state of the robot.

The robot was controlled by a simple state machine (Fig. 2), where states correspond with the phases of the task as mentioned above. We now describe these phases and discuss significant implementation details.

Task phases

Identification & sensor fusion When the game is started, GRACE is in the `IDENTIFICATION` state. She is looking for the pink hat or for a person that might help her find it. The touch screen displays the *wandering* image (Fig. 3), which depicts a pink hat, with question marks, and the phrase "Touch Me. I am looking for a pink hat." The *wandering* screen serves three purposes: it informs people about what GRACE is doing; it provides an opportunity to interact with the robot; and it mitigates the limitations of the vision system by not relying on it to find people. Meanwhile, GRACE plays music and periodically says, "Where is the person in the pink hat?" Her animated face frequently changes direction, giving the appearance of actively looking around while wandering.

GRACE is equipped with a laser scanner near human knee-height and a camera near human face-height. The laser scanner clusters short adjacent range readings, labels those that appear to be human beings, and tracks those objects over time using a Kalman filter. The camera locates faces using appearance-based frontal face detectors and tracks them using skin color models. Data from these two sensors are combined to determine more reliably where there are people. This is done by registering the locations of camera-located faces and laser-located obstacles with each other in a robot-centric coordinate frame and labeling a laser obstacle as a person according to whether there is a face located above it.

Figure 3: The *wandering* screen.

A color model was obtained for the pink hat during a calibration procedure in the particular lighting environment of the conference center. If the number of pixels in the camera image that match this color model exceed a certain threshold, the hat is considered found. GRACE enters the `SUCCESS` state and the *Gotcha!* screen is displayed (Fig. 5), asking to be touched to start the game again.

If neither the pink hat nor a person is found after some time, the robot moves randomly and looks again. When a person (face) is detected, the robot enters the `APPROACH` state. However, if at any time someone touches the screen before being detected, GRACE directly enters the `ASKING` state.

Approach When a person is detected, GRACE first says, "I think I've found someone to ask," and begins to move toward the person to initiate an interaction. During this approach, the robot tries to observe societal norms such as speed, direction of approach, greeting (e.g. saying, "Excuse me!"), and personal space. These were designed from common sense understanding, and fine-tuned through testing. Without these behaviors, people who are not accustomed to interacting with robots may not realize that the robot is attempting to engage them, and they may move away from the robot rather than stand in her path.

When the approach is complete, if the person is still there (approximately four feet away), GRACE enters the `ASKING` state.

Asking In the `ASKING` state, GRACE displays a "Can you help me?" screen, with buttons "Yes" and "No." Meanwhile, she says "I am looking for the person in the pink hat. Can you help me?" In this case, the touch screen contains minimal text regarding the question, so as not to potentially burden the participant.

If the person presses "No," GRACE thanks the person, displays a *Thank you* screen, and returns to the `IDENTIFICATION` state.

If the person presses "Yes," GRACE asks the person to point her in the direction of the pink hat. To facilitate this, the *directions* image is displayed on the touch screen (Fig. 4). This image depicts GRACE in the center and eight arrows pointing outwards from the robot along with the line "Which way is the person in the pink hat?" The arrows are foreshortened to provide directional perspective from the