(Not) Interacting with a Robot Photographer

William D. Smart
Cindy M. Grimm
Department of Computer Science
Washington University in St. Louis
St. Louis, MO 63130
United States
{wds,cmg}@cs.wustl.edu

1 Introduction

In this paper, we describe our experiences with a mobile robot photography project. A mobile robot wanders around a space, taking candid pictures of people, much in the manner of a wedding photographer. We discuss our experiences fielding the system over a period of five days at a major conference.

2 Robot Photography

The basic idea of the project is to have the robot navigate in a space populated by humans, and take candid pictures of them, much in the way a wedding photographer does. We initially designed the robot to be unobtrusive so that it could take more candid photos, and not just a series of frontal "mug shots" of people starting at it.

In the current implementation, the robot wanders randomly around the space, using mostly reactive obstacle avoidance. It constantly takes images and looks for faces in them. Once it identifies one or more faces in an image, it calculates a good composition, according to some basic rules of photography. It then pans, tilts and zooms the camera to get the desired framing, and takes the shot.

To prevent the robot from wandering outside of its designated space it periodically quit taking pictures, looked for a known landmark in the scene, and re-centered itself in the space. This happens approximately every 15 to 20 minutes.

3 Observations

We ran the robot photographer system in the Emerging Technologies exhibit at SIGGRAPH 2002. The robot ran for more 40 hours over the period of five days, interacted with over 5000 people, and took approximately 3000 pictures. In this section, we briefly describe some of our observations and thoughts on this sort of robot interaction.

3.1 Bimodal Interactions

One of the goals of the project was for the robot to take candid pictures of people, not just front-on shots. This requires that the subjects of the photograph not be attending to the robot. We were initially concerned that, since the robot is large and bright red, there would be a problem with it being the center of attention. However, this turned out not to be a problem. When a person first entered the space they tended to either deliberately stay away from the robot (reading posters or standing with their friends), or walk up to the robot and try to get its attention (by waving at it, or purposefully standing in front of it). Once the initial reaction wore off, people tended to form small social groups, and ignore the robot except for brief glances when the robot approached them.

From our observations, this task has a bimodal interaction. Either the robot should be directly attending to someone who is trying to get its attention, or no one is paying attention to it and it should try to blend into the background. This implies that the interaction mode that the robot is in should be driven by the humans in the environment. If a human tries to engage the robot (see below), it should interrupt what it is doing and attend to them. Otherwise, it should try to be inconspicuous.

We hypothesize that people tended to ignore the robot partly because it made no attempt to externalize its state, and partly because it moves relatively slowly and smoothly. People are more likely to attend to things in the environment that are more "human-like", that is, those things that give behavioral cues similar to those that people do. For example, making eye contact or a rapid hand movement to get attention. When the robot does not do this, it gets classified as a (moving) piece of furniture, and is largely ignored.

3.2 Externalizing Robot State

We found that when people are actively attending to the robot, they are much more comfortable when they have some idea of what it's doing. The most obvious example of this is when they are posing for a picture. In the fielded system, there was an audible signal when a picture was taken, but this turned out not to be loud enough for most people to hear over the background noise. This led to some confusion about whether or not a picture had been taken. Adding in a more easily identifiable signal would alleviate this problem.

Another problem was knowing when the robot was lining up a shot, and when it was simply navigating through the space. Our zoom time was especially long (up to four seconds) and externally undetectable, unlike a photographer adjusting a camera. This led to confusion about whether the robot was going to take a picture, was stuck, or doing something else. There were multiple suggestions from attendees about adding in feedback about the state of the robot. This was probably the single most common comment — people wanted the robot to say "cheese", or show a "birdie" picture when it was about to take a picture.

Often people stood in front of the robot, expecting to be photographed, but were ignored. This usually happened because the robot was in landmark-seeking mode. However, this was not apparent to the general public.

We deliberately designed the robot to take pictures relatively frequently if it suspected there was a human in the shot, even if it currently had no "good" composition. This alleviated some of the frustration experienced by participants, because the robot almost always took their picture eventually.

How best to communicate the robot's state is still an open question. The robot should communicate state when someone is already attending to it, but if it continually broadcasts its state this could be distracting and prevent candid photographs.

3.3 Expectations and Intelligence

When directly interacting with the robot, several people tried to catch its attention by waving at it. The algorithms that we currently have in place have no way of detecting this, and this often led to some frustration. It seems that when interacting with the robot, people have the expectation that it will react in a "human-like" way to stimuli.

When the robot did react as expected, due to some lucky coincidence, people tended to regard it as being "more intelligent" in some sense. This seems especially true of the camera motion. In the cases where someone waved at the robot, and the pan/tilt unit pointed the camera in their general direction, this seemed to deepen the interaction. It seems reasonable to suppose that, since eye contact is so important in human-human interactions, eye-camera contact will be similarly engaging in human-robot interactions.

In fact, one of the most compelling interactions was the result of our lack of code optimizations. When looking for candidate faces, the camera typically pans over a scene. Since we did not optimize our code, the process of detecting faces is slower than it could be. The result of this is that faces are generally detected after the camera has panned past them. This necessitates panning back to the position from which the image with the faces in it was taken. This produces a "double-take" sort of motion, which is surprising engaging. Again, this is the sort of reaction one expects from a human, and seeing it on a robot seems to imply intelligence.

There were a few robot behaviours that, while sensible to a robot, communicated a distinct lack of intelligence. Due to the random walk used for navigation, the robot would often spend time pointed at a wall, or moving along the wall, not "seeing" a group of people behind it. When the robot was localizing itself it appeared to be staring at the ceiling and unresponsive.

4 Questions and Future Work

This project is still in its early stages, and is very much a work-in-progress. In the short-term, we are interested in addressing a number of questions related to human-robot interaction in the context of the robot photographer. Briefly, these include:

Is the interaction really bimodal? It seems from our initial experiments that there are two distinct interaction modes. Is this really the case, or is there a continuum between the two modes?

What effect does user sophistication have? Most of the people that the robot has interacted with so far are technically sophisticated, and familiar with the limitations of computer interaction. Does the quality of the interaction change with naïve users?

How should the interaction be driven? Our initial suspicion is that humans should drive the interaction. The robot should be inconspicuous until a human tries to directly interact with it. At this point, it should change modes. How do we detect this attempt at engagement? How do we detect when it stops?

Can we use cues to shape the interaction? Can we leverage off of cues that humans use, such as gaze-direction, speed and smoothness of movement, the notion of personal space, and (simulated) emotional state in order to direct the interaction?

We are currently working on new versions of the control, navigation and picture-taking code for this system. We will also be deploying it at a number of different functions in late 2002. We plan to use these deployments to test out some of the interaction ideas that we have outlined above, and to thoroughly record what happens. By the time of the Spring Symposium, we hope to have more experimental evidence supporting our hypothesized interaction modes.