Gesture Recognition

Sebastian Thrun (PI), Roseli Romero, Stefan Waldherr, Dimitris Margaritis

Problem: Gestures are an important aspect of human interaction, both interpersonally and in the context of man-machine interfaces. There are many facets to the modeling and recognition of human gesture: gestures can be expressed through hands, faces, or the entire body. Gesture recognition is an important skill for robots that work closely with humans. Gestures help to clarify spoken commands and are a compact means of communicating geometric information. Gesture recognition is especially valuable in applications involving interaction human/robot for several reasons. First, it provides a redundant form of communication between the user and the robot. For example, the user may say ”Stop” at the same time that he is giving a stopping gesture. The robot need only recognize one of two commands, which is crucial in situations where speech may be garbled or drowned out (e.g., in space, underwater, on the battlefield). Second, gestures are an easy way to give geometric information to the robot. Rather than give coordinates to where the robot should move, the user can simply point to a spot on the floor.

Impact: The field of robotics is currently undergoing a change. While in the past, robots where predominately used in factories for purposes such as manufacturing and transportation, a new generation of "service robots" has recently begun to emerge. Service robots cooperate with people, and assist them in their everyday tasks. A landmark service robot is Helpmate Robotics's Helpmate robot, which has already been deployed at numerous hospitals worldwide according to King & Weiman 1990. Helpmate, however, does not interact with people other than by avoiding them. In the near future, similar robots are expected to appear in various branches of entertainment, recreation, health-care, nursing, and others. This upcoming generation of service robots opens up new research opportunities. While the issue of *mobile robot navigation* has been researched quite extensively, considerably little attention has been paid to issues of *human-robot interaction*. However, many service robots will be operated by non-expert users, who might not even be capable of operating a computer keyboard. It is therefore essential that these robots be equipped with "natural" interfaces that make instructing these robots as simple as possible.

State of the Art: While the area of gesture recognition is relatively new, there has been a great deal of activity in the last few years. Some works have used natural language interface for teaching mobile robots to realize determined tasks such as Torrance 1994 and Asoh and colleagues 1997. Other researchers have proposed vision-based interfaces that allow people to instruct mobile robots via arm gestures, as for example, Kortenkamp, Huber & Bonasso 1996 and Firby and his colleagues 1995. Both of these approaches, however, recognize only static pose gestures.

Approach: Our approach extends these works to motion gestures, that is, gestures that are defined through specific temporal patterns of arm movements, such as waving. Motion gestures, which are commonly used for communication among people, provide additional freedom in the design of gestures. In addition, they reduce the chances of accidentally classifying arm poses as gestures that were not intended as such. Thus, they appear better suited for human robot interaction than static pose gestures. More specifically, a vision-based human robot interface to instruct a mobile robot through both pose and motion gestures are being developed. An adaptive dual-color tracking algorithm enables the robot to track and, if required, follow a person around at speeds of up to one foot per second while avoiding collisions with obstacles. This algorithm works in two phases: one that recognizes static arm poses, and one that recognizes gestures (pose and motion).

(a) (b)

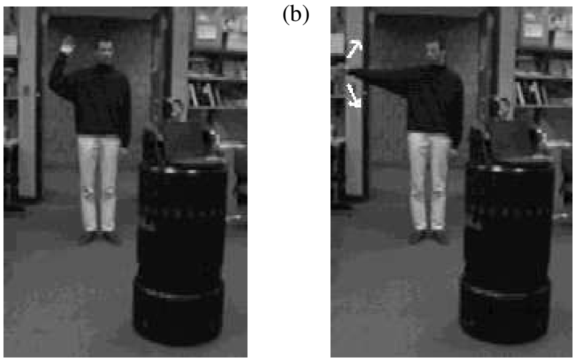

Figure 1: Example gestures: (a) stop gesture and (b) follow gesture. While the stop gesture is a pose gesture, the follow gesture involves motion, as indicated by the arrows.

- **Pose analysis** In the first stage a probability distribution over all poses is computed from a camera image. Our approach integrates two alternative methods for image interpretation: a neural-network based method and a graphical template matcher (called: pose template matcher). Both approaches operate on a color-filtered sub-region of the image which contains the person's right side, as determined by the tracking module. They output a probability distribution over arm poses.

- **Temporal template matching.** In the second phase, the results of the first phase are temporally matched to previously recorded examples of gestures (called: *gesture templates*) of individual gestures, using the Viterbi algorithm proposed by Rabiner & Juang 1986. The gesture template matcher can recognize both pose and motion gestures. The result is a stream of probability distribution over the set of all gestures, which is then thresholded and passed on to the robot's high-level controller.

Pose gestures involve a static configuration of a person's arm, such as the "stop" gesture shown in Figure 1a, whereas motion gestures are defined through specific motion patterns of an arm, such as the "follow me" gesture shown in Figure 1b. More details about our approach can be found in [Waldherr1998].

Future Work: Our current approach is not able deal with multi-colored shirts, or to follow people who do not face the robot. We believe, however, that the robustness can be increased by considering other cues, such as shape and texture, when tracking people. Further, our approach currently lacks a method for teaching robots new gestures. This is not really a limitation of the basic gesture-based interface, as it is a limitation of the robot's finite state machine that controls its operation. Future work will include providing the robot with the ability to learn new gestures, and to associate those with specific actions and/or locations, seeking to explore further the practical utility of gesture-based interface in the context of mobile service robotics.

Selected References:

[Waldherr1998] Waldherr, S.; Thrun, S.; Romero, R.; and Margaritis, D. 1998. In *Proceedings of AAAI-98*, (to appear).