Interacting with a Pet Robot using Hand Gestures

Milyn C. Moy

MIT Artificial Intelligence Lab / Oracle Corporation
500 Oracle Parkway
Redwood Shores, CA 94065
milyn@ai.mit.edu

Pet robots are autonomous robots capable of exhibiting animal-like behaviors, including emotional ones. As they become more integrated into our lives, we will need a more natural way of communicating with them. Similarly, to perceive our intentions more effectively, they will need to understand human gestures.

This work focuses on the real-time, visual interpretation of 2D dynamic hand gestures in complex environments. Our goal is to enable humans to communicate and interact with Yuppy, a pet robot being developed at the MIT AI Lab. The gesture lexicon consists of a set of 2D gesture classes (primitives) that include linear (vertical, horizontal, and diagonal) as well as circular (clockwise and counterclockwise) gestures.

As the user makes the gesture, images taken by the robot's camera are processed on a frame by frame basis. We assume that the user wears long-sleeved clothes and the robot is static while observing the gesture. Our strategy for interpreting hand gestures consists of: hand segmentation, feature extraction, and gesture recognition. Currently, our algorithms run in a user-specified window of attention which excludes the user's face.

We use both motion and color information to segment the hand from the cluttered background. Motion is detected by taking the image difference of three sequential RGB images. The skin-colored regions are computed by comparing the HLS representation of an input image with an a priori model of skin color in HS space. The results of these two modules are combined and the hand is chosen to be the skin-colored region with the largest area and the greatest number of displaced pixels. The motion of the hand's centroid is tracked in real-time as new image frames are processed.

Our system assumes that once the hand's velocity exceeds a certain threshold, the user has started a gesture. As the hand moves, the horizontal and vertical displacements (dx, dy) of the hand's centroid are stored in a feature vector until the hand pauses for 2-3 seconds.

To recognize a gesture, we analyze the feature vector. For linear gestures, the (dx, dy) displacements cluster around fixed axes in the dx-dy plane: vertical gestures around the dy axis, horizontal gestures around the dx axis, and diagonal gestures around the two bisecting axes (45° with respect to the dx-dy axes). The direction of motion is determined by the side of the axis (positive/negative) on which clustering occurs. For circular gestures, the centroid of these displacements coincides with the origin, and the direction of motion is deduced from the time sequence of (dx, dy) in the feature vector.

Once the gesture is identified, it is queried in a database of user-specified gestures. If found, the command associated with the gesture is issued to the robot; otherwise, the gesture is ignored. Composite gestures can also be recognized by combining these primitives.

We performed an initial experiment where 14 subjects were asked to perform 5 times a sequence of 16 gestures. We achieved above 15 frames per second using a 300 MHz Pentium Pro system. The accuracy rate is over 90% for primitive gestures and slightly above 70% for composite gestures. These results demonstrate the viability of our system for unstructured environments and its robustness to different skin tonalities and varying lighting conditions. Some reasons for which the obtained accuracy was not higher include problems in reliably tracking the hand and detecting the beginning of gestures, distortions caused by the camera's tilt, and errors made by the subjects while gesturing.

Despite these problems, via an off-board color microcamera, any person can easily navigate the robot around the lab using each primitive gesture as a command. New behaviors such as approaching a person and searching for a bone can be implemented on Yuppy using this interface. Future work will address additional competency in tracking the hand's motion, coping with simultaneous motion of both the robot and the human, and supporting simultaneous interaction of multiple people with the robot. We are also interested in gesture learning, the robot's reaction to both what it perceives and how it feels, and the interpretation of humans' emotions implicit in their gestures. See (Moy 1999) for a more detailed description of this work.

References

Moy, M. C. 1999. Gesture-based interaction with a pet robot. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence.*