Duo: A Wearable System for Learning about Everyday Objects and Actions

Charles C. Kemp
Massachusetts Institute of Technology
Computer Science and Artificial Intelligence Laboratory
Cambridge, MA 02139, USA
cckemp@csail.mit.edu

Abstract

Duo is a wearable system designed to learn about everyday actions and the objects to which they are applied. Duo uses first person video and kinematic sensing, with a head-mounted camera and four orientation sensors, mounted to the head, torso and dominant arm of the wearer. We describe methods for segmenting the wearer's arm, the objects with which the wearer interacts, and the actions the wearer applies to these objects. Both real-time online methods and off-line annotation tools are used to efficiently generate training data and help the system learn.

1. Motivation

Machines that better understand the everyday activities of people, would be better able to anticipate our desires and meet our everyday needs. Wearable systems have a privileged perspective on human activity that closely resembles that of the wearer, and hence offers advantages for learning about human activity. Duo both passively observes and actively influences the behavior of the wearer in order to learn about the appearance of everyday manipulable objects and the ways in which they are used.

2. The Platform

Figures 1, 2, and 3 show Duo and its sensory input.

3. Segmentation Methods

Both kinematic and visual segmentation are important for Duo. Good segmentations can greatly simplify learning within and between the two domains by meaningfully abstracting away from the vast number of pixels and orientations sensed over time. With Duo, we have explored several methods of producing high-quality segmentations.

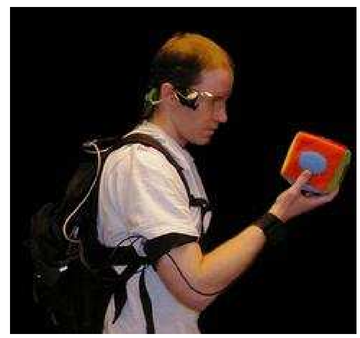

Figure 1. A picture of Duo.

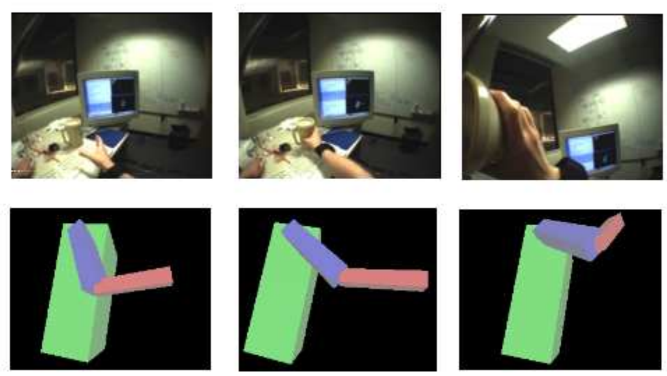

Figure 2. Three captured frames of registered kinematic and visual data.

Our first system made use of real-time cooperation with the wearer to greatly simplify segmenting important objects in the environment. Upon detection of distinct reaching events, Duo would ask the wearer to "Show me the object". [1] If the wearer responded by holding the object up

[1] For an analogous method of wearer assisted kinematic segmentation, Duo could ask the wearer to "Do that again." after an action of interest.

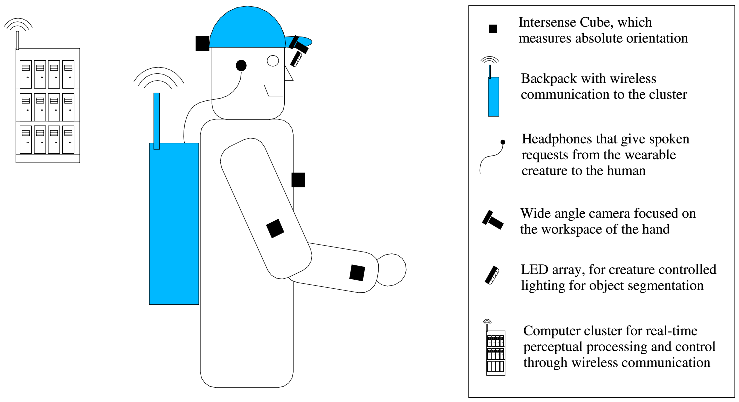

Figure 3. The main components of Duo.

Figure 5. The results of using Duo's new segmentation algorithm on the wearer's arm.

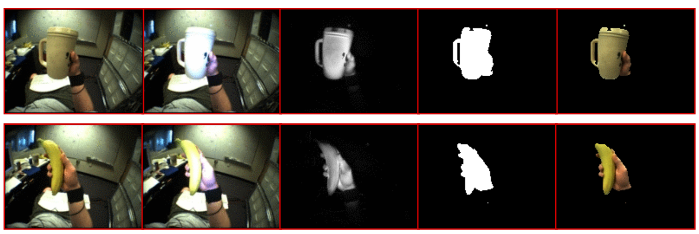

Figure 4. This figure shows the steps of two active segmentations. From left to right, subtracting the first column from the second column, which was illuminated by Duo's LEDs, results in the third column, which is thresholded to form a mask, in the fourth column, which is applied to the first column to produce the high-quality segmentations shown in the last column.

for inspection, Duo would recognize this action and synchronously flash white LEDs mounted around the camera (see figure 4). To improve the quality of the segmentation, Duo would monitor the head motion of the wearer and request that the wearer, "Keep your head still.", if the motion was significant enough to distort the segmentation. The kinematic detection of reaching used an approximate kinematic model and a matched filter. The matched filter was essentially a large scale derivative (Harr wavelet) that looked for strong motion away from the torso over a significant length of time, immediately followed by a strong motion towards the body over a significant length of time.

Our most recent system combines a new visual segmentation algorithm with off-line browsing and annotation tools. The annotation tools allow a teacher to efficiently browse through prerecorded activities and annotate them in both the kinematic and visual modalities. The visual segmentation algorithm, which is used by Duo in real-time to detect and recognize objects, also facilitates annotation by allowing the user to select significant parts of the image with a single click of the mouse. The four frames in figure 5 show segmentations of the wearer's dominant arm during everyday activities. The only prior knowledge used in these segmentations was the approximate location of the arm, which was used to generate points at which to initialize the segmentations. We are currently training the system to use the kinematic model to improve these guesses and hence segment the arm more efficiently and in circumstances with ambiguous visual information. The new visual segmentation algorithm we've developed appears to be well suited to wearable computing. The algorithm performs an efficient form of curve evolution using histogram-based appearance models and a probabilistic graph suitable for the Viterbi algorithm. For kinematic segmentation, we are using multi-scale derivatives over changes in the kinematic configuration of the body.

4. Conclusions

We are currently training Duo to recognize a variety of manipulable objects and artifacts in the environment that have previously been segmented and sparsely labeled. Similarly, we are training Duo to recognize more elaborate kinematic actions that have been segmented and sparsely labeled. Finally, we are training the system to associate these salient visual and kinematic segments, in essence using them to label one another.