A Multi-Cue Vision Person Tracking Module
MIT2000-05

Progress Report: July 1, 2001—December 31, 2001

Trevor Darrell and Eric Grimson

Project Overview

This project will develop a multi-cue person tracking sytem that will integrate stereo range processing with other visual processing modalities for robust performace in active environments.

Progress Through December 2001

The goal of this project is to develop robust vision-based modules for tracking and recognizing people. We currently are working on two complementary systems for tracking and recognition, respectively, and plan to integrate them together in the future. Each system has been significantly enhanced in the past six months.

Our tracking system is designed to work in complex indoor environments with highly variable illumination. Using multiple stereo cameras, we detect points on foreground surfaces in the scene and group these together across views and time to form trajectories. The trajectories of different people are disambiguated by examining the distribution of colors, and by their spatial location.

Since this system works by comparing stereo depth estimates of background and new images, it is very sensitive to regions of the scene where stereo depth is hard to obtain (e.g., uniform regions). Recently, we have developed a new algorithm for determing background values using visibility constraints from additional views. The key insight is that constraints on the background depth can be inferred from the empty space observed in other stereo cameras. If a depth value is seen in a second camera, then all points closer to that camera must be empty, and can be considered to be in front of the background surface in the first camera view. A presentation and paper on this is available and will be published in IJCV.

Our recognition system has been designed to combine appearance information across views and across time for recognition. In contrast to the tracking system, our recognition system (currently) presumes presegmented silhouette inputs. Using a visual hull algorithm for 3-D reconstruction, a textured model is created and then rendered into a 2-D image using a virtual camera positioned to observe the frontal view of the face. We have developed algorithms to automatically position the virtual camera based on the trajectory of the person. A gait

recognition algorithm is simultaneously applied to rendered side-views of the person. A presentation and paper on this system is available and will be presented at the CVPR conference.

Research Plan for the Next Six Months

We are planning to bring our system to NTT Atsugi to integrate our person tracking system with the recognition algorithms developed at NTT. MIT researchers will visit to install the system in March 2002, and discuss future collaboration topics.

One issue we will continue to work on is the issue of optimal integration of observations over time in this system. A sequence of face images is available as a user walks through the environment, and potentially each face image can be used for classification. But the naive approaches of multiplying likelihoods fail to deal adequately with outliers, and fail to capture expected variation in the data. Instead, we have been exploring new algorithms to explicitly match distributions of faces, using a "eigenface" density model and a closed-form KL divergence criteria.