

A Multi-Cue Vision Person Tracking Module MIT2000-05

Progress Report: January 1, 2002—June 30, 2002

Trevor Darrell and Eric Grimson

Project Overview

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

Progress Through June 2002

We have developed a multi-view based person tracking system using robust background estimation techniques. This system can detect the location of multiple people moving in a complex indoor environment with dynamic illumination (such as from video projection). Multiple stereo cameras are used to observe the environment, and 3-D points in the scene that differ from a background pattern are detected, clustered, and classified into individual trajectories.

We developed a real-time version of our robust multi-view background estimation algorithm. The key insight in this algorithm 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.

In March 2002, Dr. David Demirdjian visited NTT and installed a version of this system for collaborative research. An integrated system for tracking the head region of people moving through an office environment was developed. This system used MIT's person tracking technology and NTT's active search technology for fast and efficient recognition.

Research Plan for the Next Six Months

We are extending this system in three main directions in the next six months, and will decide which to pursue with greatest emphasis in consultation with our NTT colleagues.

Audio-visual tracking. We are integrating our visual tracking system with a microphone array that can focus audiovisual streams on multiple locations in the environment.

Articulated tracking. The present system tracks the location of multiple users coarsely in the environment, but is insensitive to fine details. We are developing algorithms for fine-scale tracking of face pose and articulated configuration, and are integrating them into the multiple person tracking system. The pose and articulated tracking algorithm require stereo range information, which can be provided by the foreground masks in the person tracking system.

Face/gait recognition. The current MIT-NTT system recognizes individuals based on face appearance. In many environments, other cues such as body shape or gait dynamics are even more important than face for making quick estimates of identity. Shape and gait cues can be integrated with face recognition for more robust and

accurate recognition. We have been developing algorithms for view independent gait recognition which can work from segmented silhouette images. We are researching ways to integrate this approach within our multiple-person stereo tracking framework.