

A Multi-Cue Vision Person Tracking Module MIT2000-05

Progress Report: July 1, 2002—December 31, 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 December 2002

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

Articulated tracking. The previous system tracked the location of multiple users coarsely in the environment, but was insensitive to fine details. We developed algorithms for fine-scale tracking of face pose and articulated joint configuration, and integrated them into the multiple person tracking system.

Face/gait recognition. The previous MIT-NTT system recognized 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 developed algorithms for view independent gait recognition which can work from segmented silhouette images.

Research Plan for the Next Six Months

We are extending the articulated tracking system to be integrated with multimodal speech recognition, so that face pose and arm gestures can be used for referring to objects as a part of conversational discourse. In collaboration with the LCS SLS group, we are extending the implementation of the multimodal galaxy server to include pointing references using arm gesture as well as face pose. We hope to add a module which can perform simple recognition of objects being held in the users hand, so that a user could ask a system "What is this?" or tell it "This is the NTTMIT report document".

We also plan to extend the face and gait recognition system, to employ prior models to constrain body shape estimation for improved results in noisy conditions. We have initial results using a PCA-based approach to regularize the image-based shape model, and are exploring non-parametric and nearest neighbor approaches. Our results so far show that the quality of recovered shapes can be dramatically improved using a class-specific regularization step.