Who Should I Approach?
A probabilistic framework to determine the most likely person willing to have close interaction

[Demonstration Abstract]

Shokoofeh Pourmehr, Mani Monajjemi, Jake Bruce, Jens Wawerla, Richard Vaughan
School of Computing Science, Simon Fraser University, Burnaby, BC, Canada
vaughan@sfu.ca

Abstract

We present a system that enables a mobile robot to detect and approach the person who appears most likely to want to interact with it. We present a probabilistic method for integrating information extracted from multiple sensors, including vision, audio and LIDAR. The robot dynamically estimates the relative position of surrounding people and their level of 'engagement', and approaches the most-engaged user to begin close interaction.

Keywords

Human-Robot Interaction, Attention Estimation, Robot Perception, Human-Robot Spatial Interaction, Sensor Fusion

Demonstration

Real-time human detection is understood to be one of the key components of social robots and autonomous vehicles working in human environments. We demonstrate a simple but widely useful functionality in which our robot recognizes and engages the person in a crowd who appears to be most actively seeking its attention.

In this demonstration scenario, a robot is co-located with a group of people, one of whom wants to engage the robot in a close-up face-to-face interaction. The robot starts out too far away from the people to perform this interaction. The people are standing in arbitrary positions and poses. We assume that the person who seeks an interaction will stand facing the robot and try to get it's attention by emitting a sound. The robot's mission is to find and approach the person who wants the robot's attention.

Determining the goal person and estimating her position on a mobile robot can be a difficult task due to sensor uncertainties and detector failures, the cluttered environment and the variety of human appearance. Our robot is equipped with different sensory modalities including laser range finder, RGB camera and multi-array microphone. Integrating different modalities is intended to increase robustness since the shortcomings of one modality are compensated by the strengths of the others.

Several different human-related features are extracted from the multimodal sensor data: (1) Laser range-finder scans are analyzed for extracting the leg-patterns; (2) The direction of sound sources are used as audio cues to detect the person who actively emit sounds to attract the robot's attention; (3) Haar feature-based cascade classifiers to detect human torso in RGB images.

Figure 1: The picture demonstrates a scenario of four people standing in robot's work area. The person on the right is not in the camera's field of view, but is trying to get the robot's attention by blowing a whistle.

Based on these detected sensory cues, multiple sensor-specific evidence grids will be generated to map the probabilistic estimation of human presence in the corresponding sensor's perception range.

After starting the mission, the robot approaches the location that has the highest probability of containing an interested human. To find this location, the multimodal evidence grids are integrated into a likelihood grid. As a result people who are detected by only camera and/or laser scanner are less likely to be interested in interaction. However, the probability is higher for an interactive person since there will be more cues (e.g. sound, torso and leg orientation) indicating human attention. The integrated likelihood grid is used as the input to the robot's navigation system and enables a behaviour that approaches the most attentive person.

We intend this multimodal interaction system to enable a natural and effective means to initiate a one-on-one Human-Robot Interaction starting from a crowd of people at large distances from the robot.