Group Perception Methods to Support Human-Robot Teaming

Angelique Taylor and Laurel D. Riek
Computer Science and Engineering, Univ. of California San Diego
{amt062, lriek}@eng.ucsd.edu

I. Introduction

The field of robotics is growing at a vast pace with robot deployments in everyday environments such as hospitals, schools, and malls. On average, 70% of people in these environments are in groups [1]: they walk, work, and interact in groups[1]. Therefore, robots need a high-level understanding of groups, which reveals many exciting technical and socio-technical challenges to enable them to fluently assist and interact with groups. Yet much prior work in human-robot interaction (HRI) focuses on dyadic interaction. This prevents robots from understanding how teams work together as well as how to work alongside them. Thus, the goal of this paper is to design perception (computer vision) methods to enable robots to work seamlessly in a group. Here, we discuss our work to date as well as our future work on a group detection and tracking system, group motion forecasting, and a robotic system to enable robots to navigation and interact among groups.

As a first step to achieving this goal, we conducted an in-depth analysis of state-of-the-art of group perception methods within the computer vision and social signal processing (SSP) fields. This analysis identified open challenges that need addressing in order for robots to effectively collaborate and work with groups in real-world settings [2]. For example, most group perception methods employ fixed, overhead cameras (i.e., an exo-centric / third-person perspective) to sense groups of people, rendering them impractical for mobile robots. Instead, perception methods based on an ego-centric (i.e., first-person) perspective are more suitable for mobile robots, to enable them to enter any environment and accomplish their goals without external sensing requirements.

Next, we proposed a theoretical framework that enables robots to perceive groups of people and their level of affiliation (a sense of belonging), inspired by teamwork models from the social sciences [3]. This framework reflects four principles which robots can leverage to behave appropriately in groups, including: 1) the *proximity principle* which states that people tend to join groups in close proximity to them, 2) the *elaboration principle* which states that groups are dynamic systems that grow in complexity, 3) the *similarity principle* which states that people tend to stay in groups longer when they share common goals and interests, 4) the *complementarity principle* which states that people tend to stay in groups longer when they have mutually beneficial characteristics [4]. Computationally, this framework is reflected in three stages: group perception, group detection and tracking, and path planning among groups.

Following this, we designed, implemented, and evaluated a new ego-centric, unsupervised group detection system, the Robot Group Estimation Model (RoboGEM) [5], [6]. Historically, prior group perception work has several limitations in that it tends to: (1) focus on exo-centric perspective approaches, (2) use data captured in well-controlled environments and therefore cannot support real-world situations, and (3) use supervised learning methods which may potentially fail when robots encounter new situations. In contrast, RoboGEM is unsupervised and works well on ego-centric, real-world data, where both pedestrians and the robot are in motion at the same time. RoboGEM outperforms the current top-performing method by 10% in terms of accuracy, and 50% in terms of recall, and it can be used in real-world environments to enable robots to work in teams [6].

In subsequent work, we expanded the scope of RoboGEM by adding a new tracking algorithm and improving its group detection method (RoboGEM 2.0). RoboGEM 2.0 performs well in crowded environments using a tracking-by-detection approach and, to the best of our knowledge, is the first tracking method to leverage deep learning. This provides a more robust affinity representation than prior work (which employs hand-crafted feature representations), and can be used for group data association. RoboGEM 2.0 is based on the intuition that pedestrians are most likely in groups when they have similar trajectories, ground plane coordinates, and proximities. It also includes new methods for group tracking that employ Convolutional Neural Network (CNN) feature maps for group data association, and Kalman filters to track group states over time. This work achieved state-of-the-art performance, and will be submitted for publication this summer.

II. Method

One pressing challenge with deploying robots in real-world environments is ensuring the safety of pedestrians around them. Acquiring a high-level understanding of social dynamics enables robots to effectively predict future trajectories of pedestrians using group motion inferences. This will enable robots to navigate in heavily populated environments and can potentially address the "freezing the robot" problem - that

[1]Groups are defined as two or more people in close proximity to each other with a common motion goal.