Robot Affiliation Perception for Social Interaction

Angelique Taylor
University of California San Diego
La Jolla, CA 92093, USA
amt062@eng.ucsd.edu

Laurel D. Riek
University of California San Diego
La Jolla, CA 92093, USA
lriek@eng.ucsd.edu

Abstract
The goal of our research is to design algorithms that allow robots to use group dynamics and nonverbal behavioral cues to estimate affiliative states of groups. It is important that robots build effective partnerships with people as robots become more integrated into human occupied spaces. To achieve this, robots must understand how humans use social and group dynamics to interact with one another. To address this challenge, we adapt principles of attraction from the literature to build computational models for robots. We will discuss our theoretical framework, affiliation system design, and three planned experiments which we will deploy on a mobile robot in real-time to enable it to interact with groups of people in social environments.

Introduction
People naturally gain a sense of whether they “fit-in” with groups based on face-to-face interaction which is known as affiliation, which is a sense of belonging to a group of people. However, robots do not have this ability. In order for robots to be accepted and functional around people, robots need to understand how people achieve affiliation amongst each other.

The human-robot interaction (HRI) literature suggests that people show more positive reactions to in-group robots than they do to out-group robots [3]. However, in social settings, robots are inherently out-group members when they enter social spaces with people. This is because people naturally view robots as being different from themselves. Therefore, to interact with people, robots need a means to affiliate and build rapport with people so they are viewed as in-group members.

Theoretical Framework

To address this challenge, building algorithmic models of group dynamics can inform a robot's understanding of group affiliation [6]. Affiliation is based upon principles of attraction [2]. When people attend social gatherings, they typically scan a room searching for familiar people. However, when people are not familiar within a crowd, the **proximity principle (PP)** states that people tend to join groups that are close by [2].

Once an individual starts interacting with a group, they gain a sense of whether they belong in the group. If they gain a sense of membership in the group, people tend to continue this interaction for a longer period of time; otherwise, they tend to leave this group and explore others. This is known as the **elaboration principle (EP)**; hence, groups are dynamic systems which grow in complexity over time [2].

Groups tend to retain membership based on connections, which may be personal or professional. For instance, people working in the same industry, from common backgrounds, or having similar interests and opinions tend to retain connections in groups. This is known as the **similarity principle (SP)** [2].

However, some people are interested in building mutually beneficial connections. For example, dominant people tend to connect more with people that are submissive (dominant-submissive) than they do with people that are dominant and vise-versa. This is known as the **complementary principle (CP)**; hence, people are attracted to those who possess characteristics that complement their own [2].

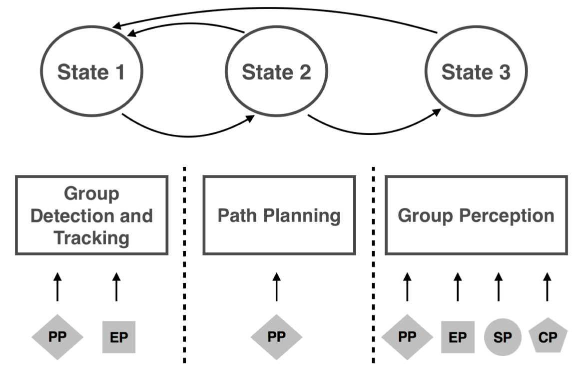

Figure 1: Affiliation System. State 1 is designed based on the PP and EP; therefore, it detects and tracks groups of people. State 2 plans a path trajectory from the robot's current position to a goal position in a group, which is designed based on the PP. State 3 allows the robot to perceive the affiliative state of groups using all four principles of attraction.

Robots can potentially leverage the principles of attraction to engage with groups of people. We argue that sensing social signals is the best way for robots to do this. Social signals are the expression of one's attitude toward a social situation, which encompasses nonverbal behavioral cues such as body posture, gestures, and proxemics [6]. Work in HRI suggests that social signals provide an understanding of a single person's attitude about a robot [5]; thus, they can also help understand the affiliative states of groups and engage accordingly.

In our work, we plan to operationalize the aforementioned principles. The overarching goal of our work is to address the following research question: *how can mobile robots use the principles of attraction to autonomously navigate a so-*

Figure 2: We use the Fetch in our work which stands at human height and is equipped with an RGB-D sensor, back-drivable 7-DOF arm, and a 2D laser scanner.

cial scene and estimate affiliative states of groups? There are several steps required to enable us to explore this question. Fig. 1 shows how we plan to design our system, which we will explain in the next section.

Planned Experiments

In our work, we will employ a Fetch robot, see Fig. 2. We plan to collect about 2 hours of RGB, depth, skeletal joint trajectory, and audio data. The Fetch robot will wander around observing groups of people as they socialize during an on-campus social event. Then, we plan to conduct three major experiments.

Experiment 1

In the first experiment, we plan to design State 1 of the affiliation system as shown in Fig. 1. For this state, we will use depth data to design an algorithm that detects and tracks groups of people, which encompasses the PP and EP. However, there are many challenges in designing this algorithm which include object occlusion, clutter, varying lighting illuminations, and limited field of view. In addition, the current state-of-the-art for detecting groups from an ego-centric perspective assumes that groups of people are walking together in outdoor environments – that is, pedestrian-like scenarios [4]; however, this is not always the case in social spaces. Therefore, we plan to design a particle filter based algorithm that takes free-standing and in-motion groups into account and clusters groups of people so that our algorithm is robust to *social* and *pedestrian-like* environments. We plan to evaluate our algorithm using metrics such as multiple object tracking accuracy (MOTA), false positive rate, false negative rate, and mismatches and compare our results to the current state-of-the-art [4].

Experiment 2

In Experiment 2, we use the PP and group detection results from State 1 to plan a path trajectory to the closest group of people. The current state-of-the-art for navigation in robotics typically employs variants of SLAM algorithms to estimate the state of the robot [1]. These algorithms depend on landmarks to map the robot's position to its environment; however, currently existing datasets are collected in empty rooms or hallways with no people present. This is unsuitable for our work as we employ our robots in human-occupied spaces.

We plan to design an algorithm that is inspired by SemanticSLAM algorithm [1], which uses particle filters to estimate states of the robot and cellular phone sensors as landmarks. However, we plan to use humans as landmarks because this will allow us to build path trajectories for the robot to autonomously navigate around people.

There are many challenges in approaching this problem, which include dealing with uncertainty of landmarks, data association as landmarks enter and leave the field-of-view, and pose estimation. We plan to treat the group detections from State 1 as landmarks. This will alleviate common issues with SLAM algorithms such as data association and loop closure detection as we are interested in what is observed locally from the robot. In addition, we plan to iteratively map all possible path trajectories to the robot's goal position so that the robot can quickly update its path trajectory when people change their position. Then, we will evaluate our algorithm using a common metric in the literature, the root mean squared error of landmark positions [1].

Experiment 3
In the third experiment, we will design an algorithm that allows the robot to perceive the affiliative state of a group using all principles of attraction. Therefore, we use features such as height and skin tone/color inspired from the SP. This becomes very cumbersome as people move throughout the environment, which causes high occlusion. Additionally, to identify features such as skin tone/color, we must use RGB data which is very heavyweight on a robot with low memory capacity.

To address these challenges, we plan to design memory efficient algorithms using lightweight features on RGB data. This will allow us to temporarily store this data in memory in a meaningful way without preserving the RGB data itself.

Moreover, there is lack of evidence in the literature about whether social signals can inform our understanding about the CP [7]. Therefore, we plan to explore how we can leverage social signals to identify complementary patterns in movement to better understand the affiliative state of groups. With this data, we can detect body posture and gestures. Additionally, because the literature suggests that 90% of body gestures are associated with speech, we plan to use a multimodal approach to classifying the affective states of groups [7]. Then, we will evaluate our method using crowdsourcing techniques where we ask annotators to rate the level of affiliation in the group. Finally, we will use inter-rater reliability to determine the fidelity of the annotations.

Discussion and Future Work
This is the first step toward building algorithms that understand group dynamics in unpredictable human-spaces. This will lay the foundation for future roboticists that are interested in designing algorithms for robots that interact with people.

REFERENCES

1. H. Abdelnasser, R. Mohamed, A. Elgohary, M. F. Alzantot, H. Wang, S. Sen, R. R. Choudhury, and M. Youssef. 2016. SemanticSLAM: Using environment landmarks for unsupervised indoor localization. *IEEE Trans. Mobile Computing* (2016).
2. D. R. Forsyth. 2009. *Group dynamics.* Cengage Learning.
3. M. Häring, D. Kuchenbrandt, and E. André. 2014. Would you like to play with me?: how robots' group membership and task features influence human-robot interaction.. In *ACM/IEEE HRI*.
4. M. Luber and K. O. Arras. 2013. Multi-Hypothesis Social Grouping and Tracking for Mobile Robots.. In *Robotics: Science and Systems*.
5. R. Mead and M. J. Matarić. 2015. Proxemics and performance: Subjective human evaluations of autonomous sociable robot distance and social signal understanding.. In *IEEE Intell. Robot. and Sys.*
6. A. Taylor and L. D. Riek. 2016. Robot Perception of Human Groups in the Real World: State of the Art.. In *AAAI-HRI*.
7. A. Vinciarelli, M. Pantic, and H. Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image Vision Comput.* (2009).