Robotics and Autonomous Systems 83 (2016) 188–202

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

# Service robots: System design for tracking people through data fusion and initiating interaction with the human group by inferring social situations

Shih-Huan Tseng [a,*], Yen Chao [a], Ching Lin [a], Li-Chen Fu [a,b]

[a] Department of Computer Science & Information Engineering, National Taiwan University, Taipei, Taiwan
[b] Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

## HIGHLIGHTS

- We developed a new system to allow a robot to determine whether it should approach a human group and interact with them by inferring what the current social situation is.
- The proposed system tracks the positions of people through the fusion of different depth-related data within a long range and complex environment.
- Social situations concerning the robot and the target group of people are inferred first and then the robot decides whether it should initiate interactions or not in the human–robot social domain.

## ARTICLE INFO

*Article history:*
Received 16 October 2015
Accepted 11 May 2016
Available online 19 May 2016

*Keywords:*
Cognitive human–robot interaction
Initiate interaction
Decision networks
Data fusion
Social situations

## ABSTRACT

In a situation where a robot initiates conversation with a group of people, questions such as "where is the people group?" and "whether the robot should approach them?" should be addressed. This paper develops a new system that enables a robot to determine whether or not it should approach the aforementioned human group and interact with them after identifying what the current social situation is. The system is mainly to fuse depth-related data to track the positions of a group of people, extract social cues of those people by using depth-related data and a decision network (DN) model, and the main challenge lies in understanding the social cues of the group and the current underlying social situation concerning the relation between the robot and the group. The social cues are based on *Proxemics* and *F-formations*, whereas the social situations are categorized as *individual-to-individual*, *individual-to-robot*, *robot-to-individual*, *group-to-robot*, *robot-to-group*, *confidential discussion* and *group discussion*. Our system proceeds as follows : once a group of people are detected and the social cues of that target group of people are extracted, the corresponding social situation is appropriately inferred, and in turn the robot decides whether it should initiate conversation with the group based on rules to be specified later. The conducted experimental results demonstrate the properness of the system design and the efficacy of the proposed method in recognizing the social cues among individuals of the group as well as the nature of the social situations concerning the group and the robot.



## 1. Introduction

Human–robot interaction (HRI) is an important issue concerning the sociality between robots and humans. A significant aspect of this is the robot's initiation of interactions upon encountering a group of people. If the robot were to be replaced by a human, the above-mentioned aspect is actually a simple matter for most humans, because they are able to achieve sociality almost unconsciously; however, we know it is definitely not a trivial matter for any robot. For instance, consider the social situation as shown in Fig. 1. The robot must first observe the social conditions concerning the encountered groups of humans before deciding whether to approach some human group and to engage with them. To realize this, the robot requires a variety of skills demonstrating three fundamental abilities [1]:

* Corresponding author.
*E-mail addresses:* d96018@csie.ntu.edu.tw (S.-H. Tseng), r03944011@ntu.edu.tw (Y. Chao), r03922121@ntu.edu.tw (C. Lin), lichen@ntu.edu.tw (L.-C. Fu).

http://dx.doi.org/10.1016/j.robot.2016.05.004
0921-8890/

**Fig. 1.** Social environment in which one group of people are looking at robot and the others are talking to each other.

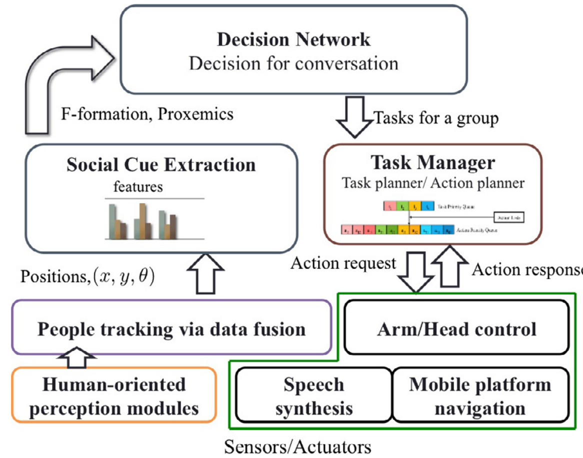


**Fig. 2.** The three layers of the system architecture.

- *human-oriented perception*: skills in detection and tracking of humans as well as recognition of face, speech, and gestures.
- *user modeling*: ability to understand human behavior and to make appropriate decisions [2].
- *sensitivity to the user*: capability of measuring feedback from users and adapting to the responses received.

This paper addresses two of the above issues: (1) human-oriented perception, which refers to the skills of a robot used to detect human's anatomical features, such as the legs, torsos, and shoulders; (2) user modeling, being used to help the robots to infer the nature of social situations by interpreting the nonverbal social cues, which often can be even more important than verbal social cues [3–5]. Such information then allows the robot to infer what the current social situation is involving a human group. By employing a utility-based strategy, initiating interactions of the robot is then implemented with a decision network (DN).

Essentially, the proposed system that allows a robot to determine whether it should approach a human group and interact with them by identifying what the current social situation is through the following four steps: (1) to track the positions of people through the fusion of different depth-related data, (2) to extract the social cues of the group such as Proxemics and F-formations, (3) to make a decision with a decision network based on social situations inferred from social cues and (4) to initiate interactions with a group of people through serious actions executed by task manager. Fig. 2 presents an overview of the system architecture which is inspired by the research [6]. The bottom layer includes basic skills, such as perception and actuation; the middle layer includes extraction of social cues as well as a task manager to handle series of actions; the top layer outlines the decision-making process using a decision network.

## 2. Related work

The study of HRI can be roughly classified as investigation of pre-interaction (*i.e.*, initiation) and that of post-interaction (*i.e.*, engagement). Under engagement circumstance, the robot must be able to appropriately recognize the appropriate social behavior of the humans in order to effectively deal with the humans and react accordingly. As illustrated by the works in [7,8], the robot is able to reconfigure its engagement with human subjects by altering both its head and body orientations. On the other hand, researchers in [9–11] investigated the effects of gaze, gesture, and other non-verbal cues on human–robot communication. However, most literatures on HRI mentioned above have focused on humans intend to interact with the robot shortly and thus the robot is unable to explore people with a longer distance [12], about 4–6 m away.

Despite the above, in the initiation context, humans do not necessarily intend to interact with the robot. A number of literatures [13–15] have investigated spatial proximity and demonstrated different behaviors when a robot first approaches to and then encounters humans. Although studies on the approaching behavior of a robot have been made in Kanda's works [13,14], their results will always let the robot approach the humans soon after it detects humans and for sure it will initiate a conversation with them. In other words, no social situation is identified a priori in these approaches. In contrast to those works, this underlying research makes use of the social cues of humans to explore and analyze social situation, whereby the robot determines whether or not to pursue interaction.

In regard to social cues for human interactions, Hall [16] introduced the concept, called Proxemics, which considers personal space and defines four spatial regions for different types of human interactions. Numerous subsequent literatures in robotics have then investigated the relationships between humans and robots using proxemics [17–20]. From a different perspective, Ciolek and Kendon [21] proposed a human communications model, named, F-formation to characterize a number of common spatial relationships among humans when they talk to one another. Based on F-formation, several methods have subsequently been developed for analyzing human interactions [22–24]. However, these approaches only detect whether F-formation is taking place rather than recognize what the current F-formation is. Although F-formation has also been applied to human–robot interaction, the crucial human orientations in the above literatures were extracted from cameras which are installed in the environment, not on the mobile robot [25], or are given in simulation results [26]. In addition, most previous studies have focused either on Proxemics or F-formation, but hardly any work has made efforts to apply both simultaneously to the authors' knowledge.

For literatures in HRI on how to develop and to enable the integration of various types of observation and action for a robot, the partially observable Markov Decision Process (POMDP) as well as the decision network have been popularly employed. POMDP is a probabilistic decision making technique used in high-level robot behavior control [27] and symbolic, scenario-level decisions [28], whereas decision networks provide a theoretical, utility-based strategy to facilitate decision-making [29,30]. Lately, Leslie & Tomas [31] described another integrated strategy for planning, perception, state estimation, and actions in complex arm-manipulation domains of the robots.

On the other hand, the simultaneous use of a laser rangefinder and a camera to carry out the detection or tracking of a person