Chapter 59
Computer Vision for Learning to Interact Socially with Humans

Renato Ramos da Silva
Institute of Mathematics and Computer Science, University of Sao Paulo, Brazil

Roseli Aparecida Francelin Romero
Institute of Mathematics and Computer Science, University of Sao Paulo, Brazil

ABSTRACT

Computer vision is essential to develop a social robotic system capable to interact with humans. It is responsible to extract and represent the information around the robot. Furthermore, a learning mechanism, to select correctly an action to be executed in the environment, pro-active mechanism, to engage in an interaction, and voice mechanism, are indispensable to develop a social robot. All these mechanisms together provide a robot emulate some human behavior, like shared attention. Then, this chapter presents a robotic architecture that is composed with such mechanisms to make possible interactions between a robotic head with a caregiver, through of the shared attention learning with identification of some objects.

INTRODUCTION

Robots have been in human's minds for centuries, but robot technology was primarily developed in the mid and late 20th century (Goodrich and Schultz, 2007). At this first time, this technology was designed for a scientific proposes to solve, improve or increases industrial processes. Nowadays, their presence at home and general society become ever more common (Argall and Billard, 2010). Then, a study of robot behavior with a human is necessary to design, understanding, and evaluation of robotic system, that involve humans and robots interacting through communication. For this, multidisciplinary fields of human-robot interaction (HRI) emerge in the mid 1990s and early years of 2000 (Goodrich and Schultz, 2007).

Fields that help scientists to construct social or sociable robots have contributions from linguistics, psychology, philosophy, engineering, mathematics, cognitive and computer science (Brezeal, 2002). From a computer view, several

DOI: 10.4018/978-1-4666-3994-2.ch059

aspects have great influence on all development of the robots, such as autonomy, information exchange, learning, etc (Goodrich and Schultz, 2007). In order to create a computational system that supplies a minimal requirement of social robot, some mechanisms are considered as fundamental and crucial: computer vision, learning, pro-active and voice mechanism. Because this, we will call them as basic elements. Moreover, others mechanisms, as emotion expression, are examples of others important features that help a robot to be more social and make the interaction easily and pleasant.

The basic elements are responsible to process external information, analyze it and performance actions that express acceptable human behavior.

We can note that in the beginning years of the child's life, some important behaviors and cognitive skills are learned and it is essential for her survival and develops a more complex behavior. One example of this is the shared attention ability. It is defined as redirecting attention to match another's focus of attention, based on the other's behavior (Deák and Triesch, 2006). It is important for a person transfer his/her intentions about the environment to another who are interacting.

The shared attention ability has high importance for people during the interaction and make necessary to develop a system that emulate such mechanism in robots. In spite of this, we present a simplest robotic architecture to provide a real robot head the capability of learning the shared attention ability by interacting with a caregiver. Furthermore, this robotic architecture was designed according to biological plausible reinforcement learning mechanism (Deák et al., 2001; Triesch et al., 2006; Triesch et al., 2007; Kim et al., 2008). In our case, a relational representation of data was used, because it is considered the most economical representation. All this facts have been made our architecture different than others and the results have been showed the capacity of learning shared attention ability. Therefore, the results show that the architecture is a potential tool to control sociable robots during interactions in a social environment.

The remaining of this chapter is organized as follows. First, a background section presents a brief overview of systems developed to provide a robot learn shared attention. After, we present our perspective on the issues, controversies, problems, etc., as they relate to theme and arguments supporting our position. Then, we discuss solutions and recommendations in dealing with the issues, controversies, or problems presented in the preceding section. In future research, we discuss future and emerging trends. Provide insight about the future of the book's theme from the perspective of the chapter focus. Finally, a discussion of the overall coverage of the chapter and concluding remarks.

BACKGROUND

Computer vision researches aim to extract information from images using many ways, such as video sequences, views from multiple cameras, or multi-dimensional data. It helps to solve some task, or "understand" the scene in either a broad or limited sense. Applications range from industrial machine vision systems which operate in a production line, medicine helping in disease medical diagnostic to artificial intelligence, helping to the computers or robots to comprehend better the world around them.

In this direction, the first step of all robots is to perceive the environment and to encode in a useful representation. This process is particularly complex, because of large number details of the environment, to filter what information is important and encode it without lose information. Breazeal classified this task as a hard problem to attempt for matching human performance using actual technology, both hardware (cameras, laser, etc) and software (Breazeal & Scassellati, 1999).

Robots have more difficulties than others to percept the environment. This difference is about