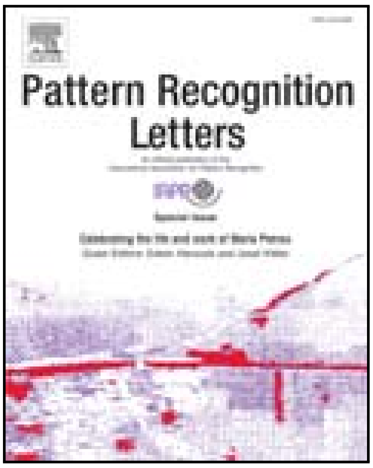

COOPERATIVE AND SOCIAL ROBOTS UNDERSTANDING HUMAN ACTIVITIES AND INTENTIONS

Jorge M. Miranda Dias , Antonio Bandera , George Azzopardi , Rebeca Marfil

PII: S0167-8655(18)30294-0
DOI: 10.1016/j.patrec.2018.07.007
Reference: PATREC 7233

To appear in: *Pattern Recognition Letters*

Editorial

COOPERATIVE AND SOCIAL ROBOTS:

UNDERSTANDING HUMAN ACTIVITIES AND INTENTIONS

1. Motivation

A key challenge of autonomous robots is to assist people in their daily-life activities or cooperate with them working together as a team. The coming generations of cooperative robots will take care of elderly people at home or help doctors in rehabilitation sessions for disabled people. In other kind of applications as, for example, rescue operations or agriculture, robots will be members of teams formed by humans and/or other robots. Collaborative tasks between humans and robotic manipulators will also improve the performance of industrial environments.

Much effort has been done in the development of social interactive robots, as entities that exhibit social characteristics in order to establish a natural and intuitive interaction with humans. Human-robot collaboration goes a step beyond. And while interaction implies action or influence on someone or something, collaboration will suppose working with others to achieve a common goal. To collaborate with a person in an effective way, a robot must be able to predict the result of an ongoing human action in order to know why its partner is doing that and therefore, determine what it should do next. Furthermore, it must be able to estimate human intentions and needs in order to react according to them; avoiding acting in conflict with what the human is trying to do.

This special issue provides a collection of approaches that cover from the behavior understanding of human actions to the analysis of scenarios where robots and humans unfold their activities in a shared environment. Within this road, we could appreciate as the understanding of actions, which imply to recognize an ongoing process, will be giving way to an understanding of intentions, and its inherent predictive component. Thus, if the first papers on the special issue deal with the problem of recognizing the activities of a person and their interactions with the rest of the environment, when the robot adopts an active role in the interaction, it is needed to consider joint attention or engagement. The closest relationship between people and robots implies the addition of new factors such as proactivity or adaptation to the user, concepts that need to complement offline training with online learning.

2. Special issue overview

Social robots need to understand human activities, dynamics, and the intentions behind their behaviors. Most of the time, this implies the modeling of the whole scene. The recognition of the activities and intentions of a person are inferred from the perception of the individual, but also from their interactions with the rest of the environment (i.e., objects and/or people). Within the general framework of interaction of robots with their human counterparts, the work by **Adriana Tapus et al** reviews recent approaches for recognizing human activities, but also for perceiving social signals emanated from a person or a group of people during an interaction. The paper summarizes some of the recent advances for modeling the social setting where the robot is involved and for extracting the relevant information during the interaction. Recurrent neural networks (CNN and LTSM) represent promising techniques for the detection and classification tasks in the interaction of a social robot. These techniques typically required a vast number of labeled training patterns, but this is not a problem due to the availability of large labeled datasets and trained networks. These approaches have shown impressive results in the recognition of human activity in the field of computer vision. In fact, the work by **Guangle Yao et al** provides a more detailed review of the use of Convolution Neural Networks (CNN) for action recognition in the video domain. The authors provide a comprehensive review according to the core issue of the extension from the image to the video domain, which is the temporal information exploitation. In the paper by **Alessia Saggese et al**, skeleton poses are analyzed for recognizing human actions. The approach is based on novel trainable feature extractors, which can learn the representation of prototype skeleton examples and can be employed to recognize skeleton poses of interest. The proposed feature extractors are combined with a technique for classification of pose sequences based on string kernels. But the perception of human actions can be also acquired using non-visual sensors. Advances in sensor technology have permitted the rapid development of small size and low cost wearable devices, with sophisticated functions for monitoring human movements. The work by **Uriel Martinez-Hernandez and Abbas A. Dehghani-Sanij** identifies and tracks activity states and transition phases using acceleration measurements, acquired from a wearable sensor attached to the thigh of participants. A probabilistic approach, composed of a Bayesian formulation and a sequential analysis method, is presented for identification and tracking of sit-to-stand and stand-to-sit activities.

A robot that interacts or cooperates with humans must not only perceive them, but also to perform actions that, in turn, will have an impact on the interaction. Traditionally, this scenario has been focused on the interaction between a single person with a robot. Thus, the work by **Salvatore M. Anzalone et al** explores the behavior of children with Autism Spectrum Disorder (ASD) during a Joint Attention elicitation task involving a Nao robot, the Aldebaran's small humanoid robot. The proposed protocol takes advantage of a RGB-D sensor to capture the movements of the child, while the robot is employed as powerful tool to induce Joint Attention. The robot is able to engage typical development (TD) children by exchanging simple, multimodal, social signals, taking advantage of its simplified but communicative shape, able to reduce the complexity of the interpersonal interactions. Results from experiments with TD children and children with ASD show the usefulness and the benefits of the presented protocol as well as of the informativeness of the model introduced. The need of capturing an exact model of the human and its movements is addressed in the paper by **Juan P. Bandera et al**. The scenario is the design of a portable system, able to be mounted on a socially assistive robot. This system is used to help in monitoring and evaluating the autonomy level of elderly people in the upcoming silver society. The work presents a new paradigm to describe and evaluate human motion that can be used in these scenarios. The proposal is based on parametric segmentation and evaluation of action primitives. These actions can be combined in different sequences or even evaluated in parallel, providing a modular solution that can easily adapt to the analysis of new behaviors or motion tests. The particular use case of the Get Up & Go test has been used to study the validity of the proposal. Contrary to these two works, the one by **Stéphane Lathuilière et al** focuses on robots that are part of groups and teams. Specifically, the work analyzes the case of the gaze control problem in a multi-person scenario, where the fact of focusing on only one person could lead to omit important information and, therefore, to make wrong decisions. That paper introduces a novel neural network-based reinforcement learning approach for robot gaze control. Their approach enables a robot to learn and adapt its gaze control strategy for human-robot interaction without the use of external sensors or human supervision. The robot learns to focus its attention on groups of people from its own audio-visual experiences, and independently of the number of people in the environment, their position and physical appearance. Using recurrent neural networks and Q-learning, it is able to find an optimal action-selection policy. To avoid the need of interacting with people for hours, they pre-train the system on a synthetic environment that simulates sound sources and moving participants. In any case, although robots and humans share the same environment, there does not exist a close interaction among them in these three works, as the goal of the robot is to drive or perceive the situation. The robot's actions are bounded to command the user with the sequence of movements that s/he must perform.

The coexistence of robots and humans is taken into account in the three last papers of this Special Issue. The work by **Pedro Núñez et al** focuses on how building a robot that will be able to navigate in an environment populated by humans, while exhibiting a socially-correct behavior. The proposed human-aware navigation strategy is built upon the use of an adaptive spatial density function, which efficiently cluster groups of people according to their spatial arrangement. Affordance spaces are also used for defining Activity Spaces considering the objects in the scene. In this way, the method allows the robot to reason about the best decisions it can perform for moving, as this proposed function defines regions where navigation is either discouraged or forbidden according to the current and predicted presence of the humans. The idea that robots need to exhibit a proactive behavior for being accepted in human-centered environments is the main motivation of the work by **Jasmin Grosinger et al.** In that paper, the authors propose a computational framework for proactive robot behavior. Their framework is grounded on the notion of Equilibrium Maintenance. Current and future states are continuously evaluated to identify opportunities for acting that steer the system into more desirable states. This process leads the robot to reason about the actions it can perform, considering the state of the environment and a model of users' intentions and preferences. Thus, it is able to proactively generate its own goals and enact them, which will depend on the temporal horizon used in prediction. At the end, a proactive robot should be able to automatically adapt and adjust to the characteristics of the user. As it is analyzed in the work by **Gonçalo S. Martins et al**, this ability is a relevant factor in the acceptance of the robot by the user. In their work, they present αPOMDP, a POMDP-based decision-making mechanism that, making use of a pre-existing model of the user, is able to learn and adapt to the user. The method is based on two main guidelines: (i) it is able to learn all the information it needs to interact properly with the user, and (ii) its actions take into account the impact that they produce on the user.

Acknowledgments

We wish to thank to all the people that enabled the publication of this special issue. First of all, we wish to thank Doctor Gabriella Sanniti di Baja, Editor-in-chief for the special issues of this

journal, for accepting the idea and for her support, patience and motivation. Our gratitude also goes to the Journal Manager and to all the staff from Elsevier, for the impeccable and timely logistical support. Eighteen articles were submitted to this special issue based on an open call for papers, out of which ten papers were accepted for publication. Submissions typically went through two rounds of reviews. We wish to equally thank the authors and the reviewers for all their hard work and contribution for the excellence of this special issue.

Jorge M. Miranda Dias (Guest Editor)
jorge.dias@kustar.ac.ae, jorge@deec.uc.pt
Khalifa University, Abu Dhabi
Universidade do Coimbra, Portugal

Antonio Bandera (Guest Editor)
ajbandera@uma.es
University of Malaga, Spain

George Azzopardi (Guest Editor)
g.azzopardi@rug.nl
University of Groningen, The Netherlands

Rebeca Marfil (Managing Guest Editor)
rebeca@uma.es
University of Malaga, Spain