Social Activity Recognition based on Probabilistic Merging of Skeleton Features with Proximity Priors from RGB-D Data

Claudio Coppola[1], Diego R. Faria[2], Urbano Nunes[2] and Nicola Bellotto[1]

Abstract— **Social activity based on body motion is a key feature for non-verbal and physical behavior defined as function for communicative signal and social interaction between individuals. Social activity recognition is important to study human-human communication and also human-robot interaction. Based on that, this research has threefold goals: (1) recognition of social behavior (e.g. human-human interaction) using a probabilistic approach that merges spatio-temporal features from individual bodies and social features from the relationship between two individuals; (2) learn priors based on physical proximity between individuals during an interaction using proxemics theory to feed a probabilistic ensemble of classifiers; and (3) provide a public dataset with RGB-D data of social daily activities including risk situations useful to test approaches for assisted living, since this type of dataset is still missing. Results show that using a modified dynamic Bayesian mixture model designed to merge features with different semantics and also with proximity priors, the proposed framework can correctly recognize social activities in different situations, e.g. using data from one or two individuals.**

I. Introduction

In recent years, there has been a growing interest in recognizing social behavior. The main focus from the psychology side is on understanding how people's thoughts, feelings, and behaviors are influenced by the actual, imagined, or implied presence of others and also the way humans are influenced by ethics, attitudes, culture, etc. From the robotics side, roboticists try to use this knowledge to model and design robots with capabilities not only to recognize human behavior, but also to interact with humans in different contexts to serve as assistants. Ambient Assisted Living (AAL) is becoming a central focus for robotics research since there is a drastic increase of aging population. Robots could be used to improve the quality of life for those people by assisting them in their daily life or detecting anomalous situations. In this context, human activity recognition plays a central role in identifying potential problems to apply corrective strategies as soon as possible. In particular, a robot that is able to analyze the daily social interaction between humans, can also detect dangerous situations such as identification of social problems, aggression, etc. Due to the aforementioned reasons, big effort has been made for creation of datasets with RGB-D data [1], [2], [3] and development of approaches for recognition of Activities of Daily Living (ADL) [4], [5]. In [6], a simple way to apply qualitative trajectory calculus to model 3D movements of the tracked human body using hidden Markov models (HMMs) is presented. Faria *et al*. [7], [8] have proposed a probabilistic ensemble of classifiers called Dynamic Bayesian Mixture Model (DBMM) to combine different posterior probabilities from a set of classifiers with interesting performance on datasets. A biologically inspired approach adopting artificial neural network to combine pose and motion features for action perception is proposed by [9]. The approach presented in [10] uses HMMs combined with Gaussian Mixture Models (GMM) to model the multimodality of continuous joint positions over time for activity recognition. All the aforementioned works have in common the fact that they attempt to recognize daily activities from one individual performing an activity or interacting with some object during the activity. Nowadays, publicly available RGB-D datasets for ADL present only one subject performing the activities. In this work, we are going further, focusing on social interaction between two subjects, since this topic is still challenging in robotics and when it comes to RGB-D data, it is still little explored.

Approaches based on other types of sensors (one or a network of monocular cameras or IMUs) can be found in the literature for social interaction analysis, however when IMUs are used in social interaction, most of datasets analyze only one individual using wearable technology. In [11], the authors show a model based on the orientation of the lower part of the body to recognize conversational groups. In [12], the authors resort to Laban Movement Analysis (LMA) to recognize the social role of a human in a social interaction. In [13], proxemics theory is adopted to define qualitative features for social behavior. In this work, we also take support from proxemics theory, which was introduced by Edward T. Hall [14] to associate proximity features with social space surrounding a person as a key-feature to study the effect of distance on communication and how the effect varies between cultures and other environmental factors. The space is divided into intimate, personal, social and public spaces. In robotics, this is a topic that was carried out by [15], [16], [17], yet in a simpler way, using only the concept of defined distances based on thresholds observed from social science. Differently from others, our approach extracts proximity-based features learned from social interaction as prior for the recognition module.

There are three main contributions in this paper: (i) A probabilistic approach that merges spatio-temporal features from individual bodies and social features from the relation-

This work has been supported by the European (H2020-PHC) project: ENRICHME[1] and by the Portuguese project AMS-HMI, COMPETE[2]. Claudio Coppola and Nicola Bellotto are with 1: L-CAS, School of Computer Science, University of Lincoln, Brayford Pool, Lincoln, Lincolnshire, LN6 7TS, United Kingdom, and Diego Faria and Urbano Nunes are with 2: Institute of Systems and Robotics, DEEC, University of Coimbra, Polo II, 3030-290 Coimbra, Portugal. (emails: {ccoppola, nbellotto}@lincoln.ac.uk and {diego, urbano}@isr.uc.pt).