Expert Systems With Applications 66 (2016) 261–273

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Robot navigation in large-scale social maps: An action recognition approach

Konstantinos Charalampous*, Ioannis Kostavelis, Antonios Gasteratos

Laboratory of Robotics and Automation, Production and Management Engineering Dept., Democritus University of Thrace, Vas. Sophias 12, GR-671 00 Xanthi, Greece

ARTICLE INFO

Article history:
Received 1 November 2015
Revised 14 September 2016
Accepted 15 September 2016
Available online 15 September 2016

Keywords:
Robot navigation
Action recognition
Social mapping
Human environments
Proxemics

ABSTRACT

As robots tend to establish their presence in every day human environments the necessity for them to attain socially acceptable behavior is a condition sine qua non. Consequently, robots need to learn and react appropriately, should they be able to share the same space with people and to reconcile their operation to man's activity. This work proposes an integrated robot framework that allows navigation in a human populated environment. This is the first work that employs the performed actions of individuals so as to re-plan and design a collision-free and at the same time a socially acceptable trajectory. Expandability is another feature of the suggested mapping module since it is capable of incorporating an unconstrained number of actions and subsequently responses, according to the needs of the task in hand and the environment in which the robot operates. Moreover, the paper addresses the integration of the proposed mapping module with the rest of the robot framework in order to operate in a seamless fashion. The generic design of this architecture allows the replacement of modules with other similar ones, thus providing adaptability with respect to the environment and so on. The method utilizes off-line constructed 3D metric maps organized in terms of a topological graph. During its perambulation the robot is ample to detect humans while it exploits deep learning strategies to recognize their activities. The memorized actions are seamlessly associated with specific rules –deriving from the proxemics theory– and are organized in an efficient manner to be recalled during robot's navigation. Moreover, the paper exhibits the differences of the robot navigation in inhabited and uninhabited environments and demonstrates the alteration of the robot's trajectory with respect to the recognized actions and poses of the individuals. The system has been evaluated on a robot able to acquire RGB-D data in domestic environments. The human detection and the action recognition modules exhibited remarkable performance, the human detection one was flawless about its decision while the action recognition one confused actions regarding the number of individuals that participate in them. Last, the robot navigation component was proved capable of extracting safe trajectories in human populated environments.

1. Introduction

The incorporation of robots in human populated environments imposes the establishment of methodologies that allow them to cruise in a secure, appropriate and common manner among people. This raises new challenges in terms of research where the concept of safe navigation attains a wider interpretation and aims to facilitate human-robot coexistence beyond the established safety measures (Mitka, Gasteratos, Kyriakoulis, & Mouroutsos, 2012). In mobile robotics, mapping allows robots to construct a meaningful description of their surrounding that endows them with the capacity to accomplish high-level objectives. The emerging topic of social mapping deals with robots' maneuvering in complex environments, where the human presence and activity is unknown and the robot has to model them in an appropriate manner so as to associate its operational behavior with the human additivity and accordingly adjust its navigation strategy so as to consider the individuals convenience. Therefore, the administration of space about humans and their reaction when interacting with each other need to be comprehended and modeled. Considerable research has been conducted in the social sciences field in an attempt to formalize an individual's behavior and to estimate how the position in space affects the quality of interaction. The cultural anthropologist Edward T. Hall coined the term *proxemics* (Hall, 1969) to suggest that *physiological* influences which are formed by culture designate zones

* Corresponding author.
E-mail addresses: kchara@pme.duth.gr (K. Charalampous), gkostave@pme.duth.gr (I. Kostavelis), agaster@pme.duth.gr (A. Gasteratos).
URL: http://robotics.pme.duth.gr (K. Charalampous)

http://dx.doi.org/10.1016/j.eswa.2016.09.026
0957-4174/

of proxemic distances. The majority of this work is outlined in Argyle (2013) and proposes an intimacy equilibrium model (Argyle & Dean, 1965), explaining the interactions between proxemic behavior and mutual gaze. In a nutshell, when they are approached excessively close enough people instinctually take on a less comfortable stance. Moreover, individuals allege a certain space to perform any ordinary conversation and only the participants are allowed to access it. This space is appreciated by others, while any violation causes discomforts (Hall, 1969). According to the proxemics theory the area around a person is separated in four fundamental zones, namely the intimate, the personal, the social and the public one. The first zone is a disc of radius 0.45 m and corresponds to actions engaging physical contact, such as whispering. The second one is a circular ring of inner and outer radius 0.45 and 1.2 m, respectively and refers to interaction with relatives or friends, as well as in crowded situations. The third circular zone with radiuses in the range $[1.2 - 3.5]$ m regards the distance in public, for example when shopping or in a museum. The last zone with radius greater that 3.5 m designates the lack of interaction. Based on the work in Takayama and Pantofaru (2009), it is safe to assume that the same proxemic behavior is applied between mobile robots and humans.

Thus, social mapping studies the embodiment of social intelligence in robots, which potentially can be highly associated with the rules of proxemics. Robot navigation relies on any path planning approach which determines the trajectory based on the description of the environment represented as a map. Maps that provide information which is sufficient solely for elementary cruising capabilities are characterized as metric, while the ones endowed with rich knowledge regarding semantic features are branded as semantic maps (Kostavelis & Gasteratos, 2015), attaining increasing attention along the last years. This growing interest may be justified by the fact that maps bestowed with wealth of information lead to improved navigation strategies closer to human perception. Under this point of view, the social mapping field pursues to enhance robots with social capabilities adjacent to ones displayed by humans. Thorough studies in the social mapping field have been recently described in Martinez (2013) and Kirby (2010). Yet, the archetypal paper of Tadokoro, Hayashi, Manabe, Nakami, and Takamori (1995) presented a planning methodology regarding robots that exist with humans, capable of applying a danger estimation procedure, computing time and magnitude of danger. Afterwards, the calculated risk is given as an input along with the length of the remaining path, its respective linearity and the change in energy to an evaluation function. The robot's path is derived by optimizing the evaluation function via genetic programming methodologies. The work in Pacchierotti, Christensen, and Jensfelt (2006) suggests a control strategy incorporating in robots acceptable behavior in hallways under the presence of humans. First, when the robot infiltrates the personal space of an individual, it signals that his/her presence has been noticed. Subsequently, the robot moves as right as possible (w.r.t. his frame) until the person passes by. Once it is completed, the robot switches to ordinary navigation and continues its journey. The work in Sisbot, Marin-Urias, Alami, and Simeon (2007) proposes a motion planner that considers the presence of humans in terms of their vision field, their accessibility and their personal choices regarding the human-robot placement. Albeit this system considers an individual as a static one, the computation time of those modules is small enough, permitting the online recalculation of the path. A motion planner that considers the dynamic presence of people is presented in Svenstrup, Bak, and Andersen (2010), where the location and movement of humans is modeled as potential fields and the most feasible trajectory is calculated using a *Rapidly-exploring Random Tree* (RRT) algorithm. Moreover, a predictive control mechanism is exploited and instead of completing the entire path, only a portion is followed until the calculation of the subsequent iteration of the RRT. In a similar manner dynamic floor fields are utilized in Boukas, Kostavelis, Gasteratos, and Sirakoulis (2015) to assess crowded areas in a hall, so as for a robot to counterbalance the congestion and to facilitate evacuation. Torta, Cuijpers, Juola, and van der Pol (2011) suggest a planning architecture that provides a socially acceptable notion and, for this reason, the personal space is modeled taking into account both inter-distance and the direction of the individual. The next location of the robot is calculated in an online fashion based on a Bayesian filtering procedure which is then passed to the navigation module. The work in O'Callaghan, Singh, Alempijevic, and Ramos (2011) presents a navigation method based on the examination of pedestrians motion patterns. A *Gaussian* learning procedure calculates a probabilistic function that concludes the robot's direction depending on its location. Additionally, a filtering procedure removes noise from the initial data which leads to a smoother function, thus providing more natural trajectories. The latter are coupled with prior ordinary planning approaches while it can be improved by aggregating more and more data. Luber, Spinello, Silva, and Arras (2012) described a learning methodology that exploits the measurements of walking people from available datasets to solve an unsupervised learning problem. In particular, they formally define the relative motion prototypes and cluster them via a hierarchical technique, which employs a distance function relying on a modified *Dynamic Time Warping* (DTW) module. Afterwards, relative motion prototypes are used in a model selection to extract social context. This in turn is utilized for the formation of the cost map serving as an input to a path planner. The work in Rios-Martinez, Renzaglia, Spalanzani, Martinelli, and Laugier (2012) presents a navigation strategy in populated environments based on the prediction of people's movement and the level of discomfort as imposed by the proxemics theory. The function that incorporates those two terms is optimized via the *Cognitive Adaptive Optimization* (CAO) approach (Kosmatopoulos & Kouvelas, 2009), a stochastic method appropriate for optimization tasks when an analytical expression is absent. As a result, the optimal result lies solely on the sensor's measurements as input to the state configuration. Moreover, this approach facilitates the expansion of the criterion with respect to dynamical or environmental constrains.

Concerning unhampered and efficient human-robot interaction (HRI), a robot should be capable of perceiving proxemics and adopt its behavior accordingly. Towards this direction, Mead (2012) proposes a methodology considering full pose, speech and gesture data, the distributions and interdependencies of which are modeled as a *Dynamic Bayesian Network* (DBN). Afterwards, the robot infers about probable poses that facilitate the production and perception of social stimuli. Another study (Mead, Atrash, & Mataric, 2011) analyzes factors that contribute to social behavior, namely attentional, interpersonal, physiological, and organizational ones. This analysis is exploited in a real-time system able to recognize and annotate those behaviors in a human populated environment. The work in Papadakis, Spalanzani, and Laugier (2013) represents the social zones in terms of isocontours of an implicit function capable of describing complex social interaction. Such zones are shaped through non-linear probability functions which derive as solutions to a learning problem in the kernel space. Moreover, the work in Takayama and Pantofaru (2009) presents a study regarding the factors that affect HRI with respect to proxemics. Particularly, three different hypotheses are examined in the form of controlled experiments: the first one considers individuals approaching a robot, whilst the second and third ones regard an autonomous and a teleoperated robot, respectively, reaching the people. Luber and Arras (2013) address the problem of detecting and recognizing social connections among people as well as tracking the clusters they form. This work relies solely on a 2D range sensor and is capable