A Review of Human Robot Proxemics

Robot Autonomy and Interactive Learning Lab
School of Interactive Computing
Georgia Institute of Technology

Osvaldo Armas

Spring 2019

Undergraduate Thesis

Abstract

Robots are increasingly becoming ubiquitous, but are currently limited in their social capabilities. For robots to become ubiquitous in social environments they need to have an understanding of proxemics. Proxemics is a quickly evolving field in robotics that is rooted in anthropology and psychology. Although it initially sought to explain human-human interaction zones, it is now deeply influential in robotics research and its applications in human-robot interaction problems. This paper examines recent developments in human robot proxemics, how it is affected by emerging technologies and key limitations on further research.

Introduction

Human robot interaction is near a turning point where robots will start to become ubiquitous. Homes are already being filled by different forms of artificial intelligence (AI), such as vacuum cleaners, personal assistants, and toy robots; however they are solely responsive to a person's whims rather than self-motivated to interact with said person. Additionally, other social spaces such as hotels, stores, and hospitals are beginning to see the how larger more useful robots could improve their functionality. In order for robots to exist in these spaces, they need to be able to function respectfully and safely around humans. This is where proxemics, the study of human spatial behavior, is needed.

Robots need to be polite to humans and respect their space, they need not walk through a group of people interacting with each other and they need to be able to adapt to people of different shapes and sizes. By allowing AI to reason about proxemics, robots will be able to navigate social environments more freely, allowing innovation for human robot interaction to flourish in spaces where robots work alongside with humans as industrial robots or care for humans in the form of companion robots [6, 9].

Current research involving human robot proxemics can be divided into two sections, distance focused and navigation focused. Distance focused research extracts empirical values for reaction bubbles based on how humans react under different circumstances, such as standing or sitting [7, 9, 8, 20]. Navigation focused research works on applications and enhancements of robot navigation using the Social Force Model and in combination with Reaction Bubbles [1, 2, 3, 4, 5, 11, 18]. Proxemics seeks to ensure not only safety but also the comfort of humans near the robot.

Kruse T. et al in Human-Aware Robot Navigation goes in depth on existing models up 2013 [24]. They divide navigation approaches into sociability, comfort and naturalness and discuss these in depth. In the more recent years, models have been developed to include these navigation with more than one of these approaches in mind [4, 12, 14]. However, Kruse T. et al discusses the comfort is used very loosely and the most problematic.

Recent developments in computer vision and machine learning have bolstered machine perception [16, 23, 25, 26]. This has inspired new approaches to human robot proxemics either through deep reinforcement learning [21] or deep supervised learning [14]. Unfortunately, these new approached are limited by the availability of data. The remaining of this paper will discuss proxemics in further detail, current research, how machine learning and cloud computing are opening new doors, and how gathering data sets that illustrate human robot interactions should be an open and cooperative effort in order to propel human-robot interaction forwards.

Proxemics Background

Origin of Reaction Bubbles

Human special behavior is explained by prior work on anthropology and psychology and is now used as the basis of reasoning for human robot proxemics. Edward T Hall is known for coining the term proxemics. He defines proxemics as the study of the interpretation, manipulation, and dynamics of human spatial behavior in co-present social encounters in *The Hidden Dimension* [18]. Hall also went on to describe reaction bubbles and how people have differently sized bubbles based on who and how they are interacting with other people. Figure 1 below illustrates how there are different psychophysical aspects to each of the zones proposed by Hall. Mead et al. developed a Bayesian network to determine if a person was expressing social queues based on these psychophysical aspects [11].

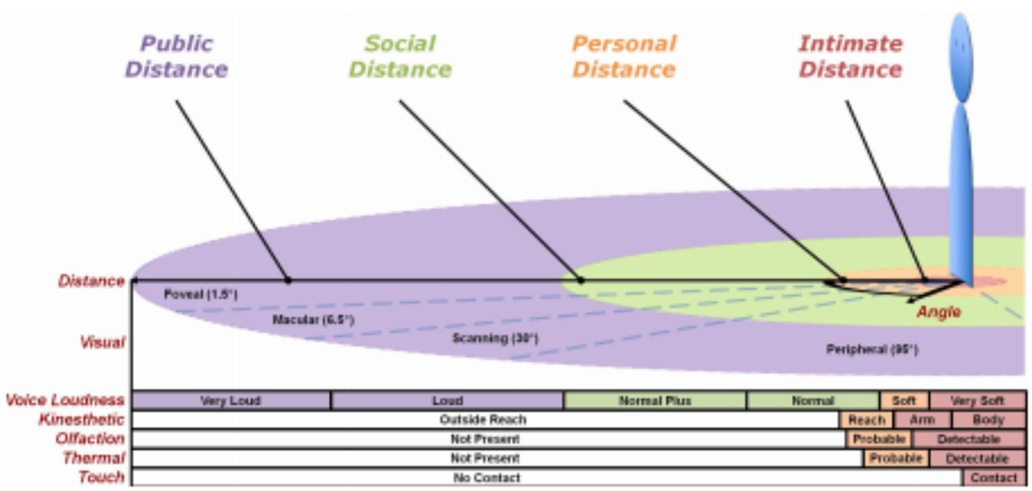

Figure 1. Mead et al.'s representation of Halls Reaction Bubbles [11].

Other studies have attempted to empirically measure changes on the size of these zones as a result of different circumstances. Mumm J. et al measures how a person's perception of a robot, and maintained gaze affect how distance they maintain between themselves and the robot [7]. Walters M. et al. measures how human-robot distances are affected by the robots appearance [8], and Dautenhahn K et al examines how the distance is affected by whether a person is sitting or standing up.

These studies develop on ways measuring human robot proxemics providing us with usable values while simultaneously revealing their variability. Additionally, they all are based on the assumption that either the speaker or the robot is static, thus introducing the Social Force Model next.

Social Force Model

In 1995 D. Helbing et al. published Social Force Model for Pedestrian Dynamics [1]. The Social Force Model (SFM) viewed pedestrian crossings and measured how people avoided each other. The study concluded that based on a person's position and velocity, one could model their behavior in a crowded environment. This model now serves as the basis of several human robot navigation studies as it is a measurable and replicable model [2, 4, 5, 6, 12]. However, the Social Force Model also has shortcomings, as it designed with the primary goal to circumvent people rather than to interact with them; this is useless if the primary goal of the robot is to interact with a human.

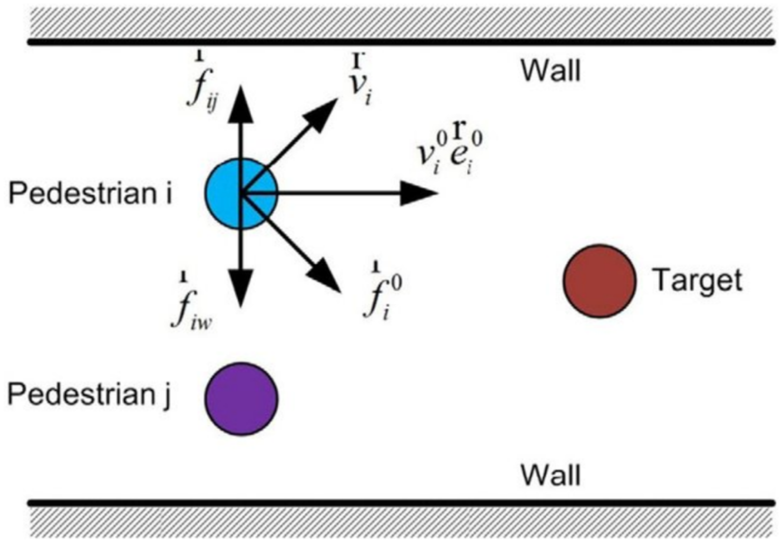

Figure 2. Depiction of the SFM where f_{ij} , f_{iw} represent the forces on pedestrian i by pedestrian j and wall w respectively. f_i^0 is the net force, and v_i is the velocity of the pedestrian.

Modern Uses of SFM and Reaction Bubbles

Hall states that proxemics "*remains a hidden component of interpersonal communication that is uncovered through observation and strongly influenced by culture.*" In the case of robots, the reaction bubbles are in fact strongly influenced by culture and thus creating a huge scalability and reproducibility problem. The empirical distance seems to range between two to four feet [22], and it is affected by both qualities of the robot (i.e humanoid or animal) and qualities of the person (i.e age or height). An examples of how such modalities affect human robot distancing can be seen on figure 3 below where the distance changes based on whether the person is sitting or standing [20]. Current approaches focus on designing these models with assumptions about the interaction space rather than creating more generic and adaptive models.

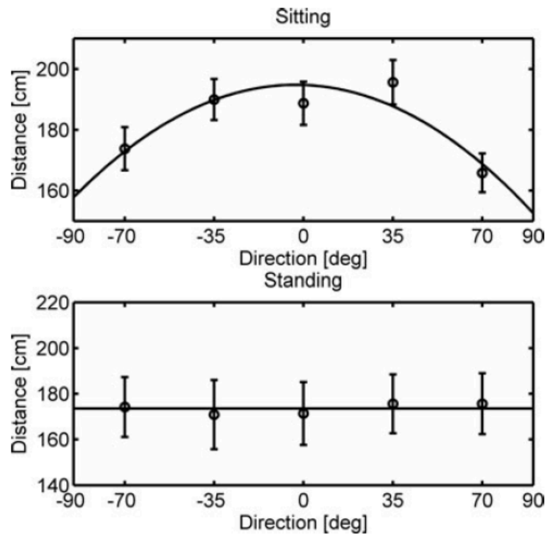

Figure 3. Example showing the difference of Human-Robot distancing between standing or sitting individuals, Torta E. et al [20].

Several extensions of the social force model have been developed [2, 4, 6, 12, 19]. These models often focus on extending social circles and creating a goal focused model with an attractive force. Ferrer G. et al uses a basic form of a social force model in an open environment and compares the *social work* required when using SFM and teleoperated approaches, more interestingly is the *social work* is a measurable and useful metric that can be used to measure the success of other approaches as well[3]. G. Ferrer et al also addresses how the social force model can be modified to improve robot that work as guides or companions [6].

Truong et al developed what is called a Proactive Social Motion Model (PSMM) in 2017, which works very similarly to the social force model, except it takes into account social circles where multiple people, moving individuals or groups, and objects people may be using [4, 12]. The PSMM is primarily designed to address humans already standing up and creates it is social circles based on overlapping reaction bubbles. This model, although at an early stage, is the state of the art in regards of robot navigation in dynamic social environments. The main limitations of the models mentioned is that they are purely deterministic instead of stochastic, and do not treat the person as a person but rather and interaction object. In other words, they do not take into account gaze, pose, and use a fixed value to represent their reaction bubbles. Since reaction bubbles been proven to be a variable value and exist in a continuous space [8, 9, 11, 20] , they could be modeled through machine learning instead.

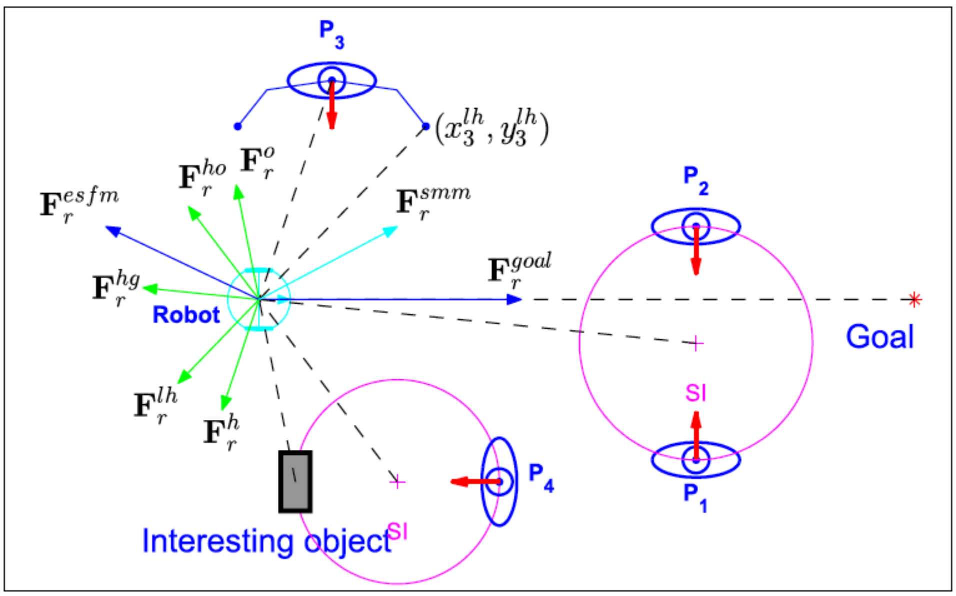

Figure 4. Illustration of the PSMM showing the social forces acting on the robot, the social groupings detected, and the human-object grouping detected. Truong X. et al. [4].

The Roles of Machine Learning and Cloud Computing in Human Robot Proxemics

The success of today's AI is largely a result on advancements in machine learning (ML) and cloud computing. One simple fact of AI being more and more ubiquitous is that the cost of making a capable AI, such as an echo dot, has decreased significantly by off-loading all heavy processing to the cloud. Meanwhile advancements in Computer vision have shown that deep convolutional networks can learn hierarchy of features, and the ability for neurons to specialize [23]. Recurrent networks have also shown impressive predictive abilities for sequential data [23]. Therefore, machines are getting better at extracting data and understanding their environments at lower computational costs.

An example using Machine Learning and Cloud Technology

Suppose a case of robot navigation using pose estimation. The robot needs to identify if a person is standing, sitting or facing another direction. Using OpenPose as the pose detection mechanism for this example [16, 26]. OpenPose is a deep learning model that is used for 2D pose estimation that is extremely accurate and works in a myriad of scenarios; however in order for this to be possible it requires a large amount of computing power. The case for cloud computing contains two parts, training and inference.

Training a model like OpenPose is a process that requires large computational and time intensive resources. Therefore, being able to train a model on the cloud through distributed computing improves the time requirement and frees a robot for having to perform one more task.

During inference, cloud computing is key when there are multiple robots that need to use a model like OpenPose and they all need to respond quickly. If there was a single robot, this robot would simply just have the hardware it needs, an expensive GPU. During inference multiple robots can use the same exact network, thus it is more resource efficient to connect multiple robots to the same cloud service hosting said network; not to mention this technology allows existing robots to expand their perception capabilities as they would only need a software update rather than a hardware and software update to acquire a new skill. That said, it is important to note, that in some cases it is best analyze which solution, cloud or local, is fastest as it may place the safety of people at risk.

Proxemics and Deep Learning

As of the last year there have been recent developments using deep learning and human robot proxemics. Gao Y. et al in "Investigating Deep Learning Approaches for Human-Robot Proxemics" [14] uses long shorter term memory networks (LSTM's), a form of recurrent networks, to predict a person's stopping distance. Even though their approach takes into account information that will not always be readily available to the robot, such as age and gender, it performs well.

Gao Y. et al also explores the uses of deep reinforcement learning in "Social Behavior Learning with Realistic Reward Shaping" [21], a key technology that needs to be developed in order for robots to learn from their interactions with humans. The primary limitations holding these complex models from success is that the simulations cannot accurately depict all human scenarios, there is no data set illustrating human-robot interactions or distancing in sufficiently diverse scenarios. In fact, most studies that are as data intensive often use very small and biased datasets collected in their labs [11, 14].

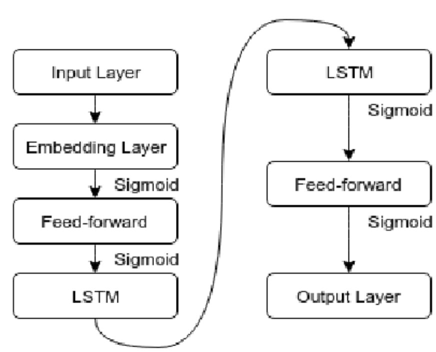

Figure 5. Illustration of the Deep Network architecture used by Gao Y. et al [14].

Conclusion

"*Proxemics remains a hidden component of interpersonal communication that is uncovered through observation and strongly influenced by culture.*" – Edward T. Hall

We have discussed about recent developments in human-robot proxemics in both static and dynamic environments. We addressed current problems in the existing models and have shown proxemics is particularly difficult as a result of inconsistency. The potential behind deep learning for human robot proxemics is endless. Deep learning has the potential to learn about the features of the hidden components of human-robot proxemics that are uncovered through observation and particularly mapping all different modalities into a continuous space.

There are other areas that remains untouched such as culture adaptability and how that implicates transfer, deep reinforcement, and unsupervised learning. There have been no studies on the performance on using transfer learning from one culture to another in a human-machine scenario; thus, it is yet unknown if a robot trained on data from one country will be able to successfully adapt to another country, begging the question of whether robots will be safe enough to be introduced prior to adjusting to the new culture and or should the agent be restrained until fully adapted.

The data needed to unlock successful human-robot proxemics is not yet readily available and its difficulty to gather is best described by: robots need to interact with people in order to get the data for robots to learn how to interact with people. Therefore, gathering data sets that illustrate human robot interactions should be an open and cooperative effort in order to propel human-robot interaction forwards.

References

[1] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E,* vol. 51, pp. 4282-4286, 05/01/ 1995.

[2] G. Ferrer, A. Garrell, and A. Sanfeliu, "Social-aware robot navigation in urban environments," in *2013 European Conference on Mobile Robots*, 2013, pp. 331-336.

[3] R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila*, et al.*, "SPENCER: A Socially Aware Service Robot for Passenger Guidance and Help in Busy Airports," in*Field and Service Robotics: Results of the 10th International Conference*, D. S. Wettergreen and T. D. Barfoot, Eds., ed Cham: Springer International Publishing, 2016, pp. 607- 622.

[4] X. T. Truong and T. D. Ngo, "Toward Socially Aware Robot Navigation in Dynamic and Crowded Environments: A Proactive Social Motion Model," *IEEE Transactions on Automation Science and Engineering,* vol. 14, pp. 1743-1760, 2017.

[5] E. Marder-Eppstein, E. Berger, T. Foote, B. Gerkey, and K. Konolige, "The Office Marathon: Robust navigation in an indoor office environment," in *IEEE International Conference on Robotics and Automation*, pp. 300-307.

[6] G. Ferrer, A. Garrell, and A. Sanfeliu, "Robot companion: A social-force based approach with human awareness-navigation in crowded environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1688-1694.

[7] J. Mumm and B. Mutlu, "Human-robot proxemics: Physical and psychological distancing in human-robot interaction," in *6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2011, pp. 331-338.

[8] Walters , M L , Dautenhahn , K , Te Boekhorst , R , Koay , K L , Syrdal , D S & Nehaniv , "An empirical framework for human-robot proxemics." in *Procs of New Frontiers in Human-Robot Interaction*: symposium at the AISB09 convention. pp. 144-149 .

[9] K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, A. Sisbot*, et al.*, "How may I serve you?: a robot companion approaching a seated person in a helping context," presented at the Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, Salt Lake City, Utah, USA, 2006.

[10] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "An affective guide robot in a shopping mall," in *4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2009, pp. 173-180.

[11] R. Mead and M. Mataric, *Perceptual Models of Human-Robot Proxemics* vol. 109, 2016.

[12] X.-T. Truong, V. N. Yoong, and T.-D. Ngo, "Socially aware robot navigation system in human interactive environments," *Intelligent Service Robotics,* vol. 10, pp. 287-295,

2017/10/01 2017.

[13] Mead, R. & Matarić, “M.J. Autonomous human–robot proxemics: socially aware navigation based on interaction potential”
Available: https://doi.org/10.1007/s10514-016-9572-2

[14] Gao, Y., Wallkötter, S., Obaid, M., & Castellano, G. (2018). “Investigating Deep Learning Approaches for Human-Robot Proxemics.” *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 1093-1098.

[15] Park, Seonwook & Spurr, Adrian & Hilliges, Otmar. (2018). Deep Pictorial Gaze Estimation.

[16] Cao, Zhe & Simon, Tomas & Wei, Shih-En & Sheikh, Yaser. (2017). Realtime Multi person 2D Pose Estimation Using Part Affinity Fields. 1302-1310. 10.1109/CVPR.2017.143.

[17] Bohus, D., Horvitz, E. “Learning to Predict Engagement with a Spoken Dialog System in Open-World Settings”, *in Proceedings of SIGdial'09*

[18] E. T. Hall. *The hidden dimension*. 1966 Chicago: Doubleday Company

[19] F. Farina, D. Fontanelli, A. Garulli, & A. Giannitrapani, & D. Prattichizzo. “Walking Ahead: The Headed Social Force Model.” 2017. PLOS ONE. 12. e0169734. 10.1371/journal.pone.0169734.

[20] E. Torta, R. Cuijpers, J. Juola. “Design of a Parametric Model of Personal Space for Robotic Social Navigation” Int J Soc Robot, 2013. 5:357–365 DOI 10.1007/s12369-013-0188-9

[21] Y. Gao , F. Yang , M. Frisk , D. Hernandez , C. Peters and G. Castellano. Social Behavior “Learning with Realistic Reward Shaping.”, Oct 16 2018.
Available: https://arxiv.org/pdf/1810.06979.pdf

[22] M. L. Walters, K. Dautenhahn, R. te Boekhorst, K. L. Koay, D. S. Syrdal and C. L. “Nehaniv An Empirical Framework for Human-Robot Proxemics.”
Available:https://pdfs.semanticscholar.org/ca6b/08930a5ce23a3173aeaf40a3db6488447b76.pdf

[23] Y. LeCun, Y. Bengio, G. Hinton. “Deep Learning Review.” doi:10.1038/nature14539

[24] T. Kruse, A. Pandey, R. Alami, A. Kirsch. Human-Aware Robot Navigation: A Survey. Robotics and Autonomous Systems, Elsevier, 2013, 61 (12), pp.1726-1743. ffhal-01684295f

[25] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 770-778.
doi: 10.1109/CVPR.2016.90

[26] S. Qiao, Y. Wang and J. Li, "Real-time human gesture grading based on OpenPose," *10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Shanghai, 2017, pp. 1-6.
doi: 10.1109/CISP-BMEI.2017.8301910