Workshops at 18th International Conference on Intelligent Environments (IE2022)
H.H. Alvarez Valera and M. Luštrek (Eds.)

doi:10.3233/AISE220016

A Framework for Socially Aware Navigation Based on Robocentric Perception of People Social Cues

Manoel ANDRADE [a], José DIAZ-AMADO [a,c,1], Raphaell SOUSA [b], Dennis BARRIOS-ARANIBAR [c], Fábio LEITE [a], Daniel ALMEIDA [b], Gabriel ALVES [a]

[a] *Electrical Engineering, Instituto Federal da Bahia, Vitoria da Conquista, Brazil*
[b] *Instituto Federal da Paraíba, Cajazeiras, Brazil*
[c] *Electrical and Electronics Engineering Department, Universidad Católica San Pablo, Arequipa, Peru*

Abstract. In the last few years, the application of robotics in environments common to human beings has been increasing. There exists several proposals of socially aware navigation frameworks for mobile robots, however all of them generally focused in a specific aspect of social relantionships between humans and robots; so, there exist a lack of approaches that integrate all the aspects related to social navigation. This work aims to propose an autonomous navigation framework based on the integration of social perception elements (from a robocentric perspective) with proxemics modelling, considering the presence of human beings and the perception of their needs, feelings or intentions. We verified the feasibility of our approach by implementing it in ROS and Gazebo, and making a qualitative evaluation of its performance in two simulated scenarios where we included people with different fellings about robot prescence, that triggered changes in the path planned by the robot in real time. So, it was concluded that this framework is feasible for implementing social navigation in mobile robots.

Keywords. Framework, Social Navigation, Proxemic, Adjustment Social Robots, Emotion Analysis

1. Introduction

In the last few years, the application of robotics in environments common to human beings has been increasing. Whether to assist in routine activities or in industrial applications, robots have been inserted in contexts of coexistence among groups of people, which has stimulated the development of works about social robots and their relations with humans. In order for this integration to occur in a more natural way, many works were developed in different componentes about social interaction and navigation; for example there exist in literature approaches for perception of humans presence and their social characteristics and cues [1,2,3,4], for modeling proxemics in humans and robots [5,6,7,8], and for navigation considering social constraints [9,2].

[1]Corresponding Author; E-mail:jose_diaz@ifba.edu.br

In environments with many people, robots trajectory is restricted by spaces occupied by humans (proxemics) or interactions between groups of people, which is fundamental for autonomous navigation in environments containing humans [9]. So, the first step of implementing such navigation is related to sensoring and perception, where robots must be capable of identifying obstacles [1] and classify them, taking into account their nature (differentiate objects from humans) and probable interactions with the robot and vice-versa [3]. When an "object" in the work space of the robot is recognized as being people, it is needed new kind of processes, like sentiment analysis, which can provide characteristics that can help and give context to the decisions. So that, the extracted information can be used to trace paths that take into account this type of characteristic [2]. For example, if after the sentiment analysis it is identified that a certain person, or a group of people are expressing fear because of the robot's approach, it might be more interesting to use a path that avoids as much as possible the space near the referred individuals; in the other hand, if people have happier expressions, they may be more receptive to an approach within the limits of the personal or group space [4].

Some works have inserted perception elements integrated to autonomous navigation techniques (recurrent in the literature), such as in [1] where the author included human beings and object deviation to improve navigation, or in [2] where the author uses facial expression detection to define actions that the robot will take (by inserting cultural elements in the trajectory planning), so that the robot could move obeying the social rules of the environment in which it was inserted. Also, that integration must be based on the modeling of interaction spaces (proxemics) which are commonly included in wors about social navigation. For example some authors have investigated how the individual space of each person can be affected in situations of social interactions [5]. Also, in [6], the authors presented a study on proxemic zones (PZs) in interactions between robots and humans, so that it is possible to note how PZs depend on external factors, such as the degree of intimacy between people, or even the dependence on the environment in which individuals are embedded. As well, in [7] authors made considerations about other aspects that affect the individual space, as in the case of the sounds emitted by the robots and their effects on social distancing, there is also a relationship between the proxemic and the activity that the person is performing at the moment. Finally, it is important to note that proxemic zones have been modeled using Gaussians, which can be symmetric or asymmetric as shown by [8]. In [9] the authors introduce the definition of social space of groups of people by means of the intersection between individual Gaussians.

This paper is organized as follows: section 2 presents works related to the scope of this proposal, then section 3 presents a description of the proposal presented in this work. Finally, sections 4 and 5 present the results and conclusions obtained, respectively.

2. Related Works

There exists several proposals of socially aware navigation frameworks for mobile robots, however all of them generally focused in a specific aspect of social relantionships between humans and robots. For example, Kivrak et. al. [10] proposed a framework for social navigation that includes the capability of taking into account the motion of people (by using a so called Collision Prediction based Social Force model) in order to include people motion as a restriction in the robots navigation. Also, Hurtado et. al. [11] proposed a framework for robots learn social behavior from humans observations or demonstrations, with a special focus in diminishing bias in robot social behavior, since in social human behavior it is possible to exist societal unfairness, such as discrimination and segregation and robots could replicate, promote, and amplify them. As well, Mavrogiannis et. al. [12] proposed a framework for social Navigation called Social Momentum which incorporates into path planning the way how humans may perceive and react to robot motion with the so called social momentum value based on the estimation of the most likely intended avoidance protocols of others based on their past movement behaviors. Che et. al. [13] proposed a framework with an idea similar to Social Momentum which uses a combination of implicit (robot motion) and explicit (visual/audio/haptic feedback) in order to aproximates continuous movements and discrete behavior modes in human navigation. Finally, Pham et. al. [14] proposed a framework for a social aware navigation in robots based on the pose analysis of humans by observing human skeleton to predicts human social activities including human running, walking, standing, sitting and laying.

Cultural characteristics, personal space patterns (proxemics), etiquette rules, among others, are some of the parameters related to the interaction between humans and autonomous robots [15].

Proxemics applied to robotics aims to improve the interaction between humans and robots, for this it is necessary that robots respect the so-called proxemic zones of each individual in the environment [16]. To maintain this balance, there are several ways to determine the proxemic zones of people in an environment. Aghaei et. al. [17] applied computer vision to estimate these theoretical spaces. The technique used is called "homography", in which the slope of the ground under the people present is identified and proxemic zones with predetermined sizes are drawn on the image.

The applications of proxemics in the determination of personal space also try to relate how emotional characteristics of people can influence the determination of personal space. Ginés et. al. [7] developed a model in which proxemic zones are estimated based on the individual's facial expression. Using algorithms such as Yolo, it is possible to analyze people's feelings, based on this information, adaptive personal spaces can be determined, depending on the elements of perception.

3. A Framework for Socially Aware Navigation

This paper has as main proposal the development of social navigation considering the presence of human beings and the perception of their needs, feelings or intentions. To achieve this goal, a framework was developed, represented by the block diagram of Figure 1, where we are proposing a basic set of steps to define an autonomous social navigation. To reach this navigation, it was necessary to implement different objectives, such as filtering the data coming from the sensors, merging information, representing the proxemic zones and finally finding a path using a robocentric visual perspective.

Figure 1 is composed of three main blocks: the first is the PEPPER SENSORS block, which includes an RGB camera and a laser sensor, and the second block is the COMPUTER VISION, this layer is implemented by software, this block is responsible for human detection. The last block is CREATE ROUTES, it receives data from Pepper's sensors and the computer vision layer, and executes algorithms for path planning. Next, we are going to explain each component of the proposed framework.

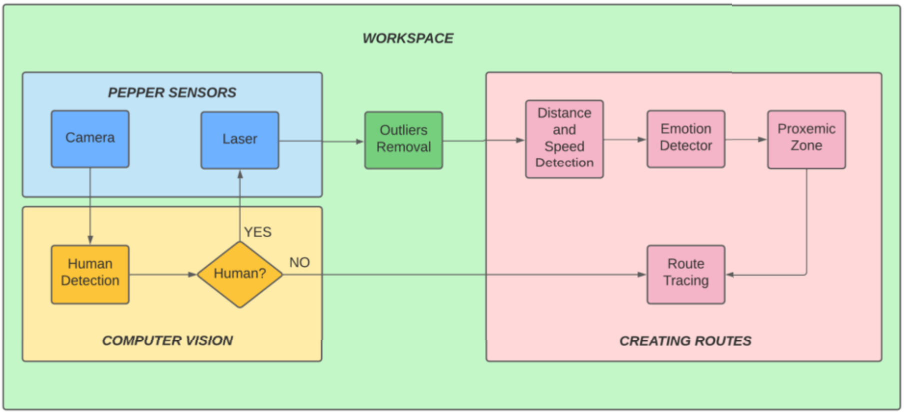


Figure 1. Block Diagram of Steps Required For Autonomous Social Navigation Framework

3.1. Workspace and Sensoring

An indispensable element for the correct functioning of the framework would be the quality of information that it will receive through the robot's sensors. For this application we are working with information coming from three Laser sensors (that can cover an area of 270 degrees around the robot) and an RGB camera (with a resolution of 640x480). All these sensors are built-in by default in the Pepper robot, this robot was used in this work for simulation and will be used in future experimental results. Initially we chose to implement our proposal, using the Robot Operating System (ROS) and the simulation software Gazebo [18] in order to validate our algorithms. Also, we implemented the framework for the Pepper robot, and for this reason this robot will also be used in the simulation process. Then we created a 3D simulation environment, which allowed us to work with the Pepper robot, people and different objects in the same environment.

3.2. Human Detection

To start the process developed in the framework, we first need to obtain information from the environment, identifying and classifying people and objects that exist in it. It is important to note that the identification and classification process needs to be done using a robocentric visual perception, that is, considering that the perspective of a robot can vary depending on several aspects, such as displacement, vibration, external agents, circular, angular movements, etc. which are part of the robot's natural process when it is moving to a certain target. To achieve this purpose, we use the YOLO Object detection algorithm [19], which allows us to efficiently differentiate between people and objects in the environment. In this work we use YOLO v3 with 30FPS. This release has a fully convolutional 106-layer underlying architecture, which makes YOLO a little slower, but with superior accuracy compared to its predecessor, YOLO v2.

3.3. Outliers Removal

When using range sensors, errors could arise in the data coming from them, for this reason some strategies were implemented to reduce such errors. The first error source identified is closely related to the laser sensor resolution. There exists a relatively large angle between each beam, which means that the greater the distance between the robot and the object, the greater the space without monitoring this sensor and, consequently, without data. For example, if the angle between the central laser ($\theta = 0$) and the next laser is $\theta = 30°$; and if an object is identified by the central laser at 5m away the next laser will only capture another value approximately 3m beside the point where the central laser finds the object. That is, the resolution of LiDAR is a problem that can cause a lot of error in the measurements. Another error identified due to the low resolution of the device occurs when one of the lasers passes between the legs of the person in front of the robot. This lateral distance between the lasers can be determined by a simple trigonometric relationship (Equation 1).

$$D_{xn} = \tan\left(\theta \left|\frac{N+1}{2} - n\right|\right) D \tag{1}$$

Equation 1 shows how the horizontal distance between each laser is related to the angle between each beam and the distance D to the object where N is the total number of lasers and n is the index of the analyzed laser. In order to solve this problem, a harmonic average calculation between the lasers n_{min} up to n_{max}. This type of average was chosen because it is less sensitive to measures that are very different from the others. The calculation of the harmonic mean is done with the equation 2.

$$M = \frac{n_{max} - n_{min}}{\frac{1}{d_{n_{min}}} + \frac{1}{d_{n_{min}+1}} + \ldots + \frac{1}{d_{n_{max}}}} \tag{2}$$

Thus, even if some laser measures a much greater distance than the actual distance to the person ($d_n >> d_{n_{min}}, d_{n_{min}+1}, d_{n_{min}+2}, \ldots, d_{n_{max}}$) the average value M will not change much, which tends to keep it as a value close to the real even if there is some error like a laser that passes between the person's legs.

3.4. Human Localization

The use of machine learning in an RGB image is enough to guarantee the detection and recognition of objects, however, this is not enough to indicate the distance from the robot to such an object. For this, the sensor called "Laser Scan" is used, which is a LiDAR but does not fire the lasers in a three-dimensional way, but even so it is sufficient for the desired application in this work. Starting from the estimate that the edges measured by the laser are aligned with the edges of the camera image, a mathematical approximation was made considering that if the lasers are overlapped on the image, they would be equally spaced from each other, with the lasers at the edges corresponding to the points $x = 1$ and $x = 640$, that is, they are in the first and last pixels of the image.

3.5. Emotion Detector

For the robot to be able to detect facial expressions in a human, it is necessary that it identifies the facial activation zones that are involved in each of the 7 basic emotions. For the identification of emotions through facial expressions, the first step is to make the robot identify which action zones are being activated on the face of the individual being analyzed. To perform the detection, the Make Human software [20] was used to create human models with facial expressions.

3.6. Poxemic Zones Representation

The way robots should approach humans in the environment for some interaction, or even the way they should keep their distance from humans are directly linked to the proxemic zones of each individual. A proxemic area is basically a spatial boundary where people tend to feel uncomfortable if someone else crosses it. The comfort area of a person is linked to several factors that can change its shape and tolerance. In this paper, the asymmetric Gaussian distribution is used to determine the [21] proxemic zones of individuals in the environment based on their facial expressions. This form of distribution receives as main parameters the coordinates of the thresholds of its sides, front and back.

After performing the steps described above, a form of visualization is required for the results obtained in the fusion of the sensors. Using as a reference to represent the proxemic zones, where we can analyze the robot's perspective towards the people in the environment. The algorithm was designed in such a way that data such as average distance (harmonic average result), class of the detected object and indices of the pixels in which the person is situated appear in real time through ROS.

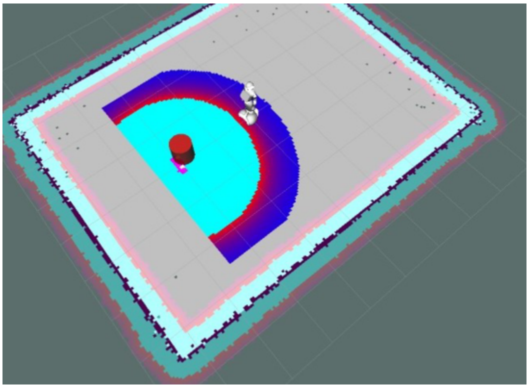


Figure 2. Proxemic Zones Obtained From Robocentric View

In Figure 2 we can see the calculated proxemic zones, where Dark blue represents a public zone, in this case the robot can invade this area, but the social zone (light blue) and the intimate zone (Red), the robot will not be able to enter on it, leading to a possible modification of the robot's navigation path. The social zone area is directly influenced in size and shape by the emotion detected by the robot of the person in front of it.

3.7. Route Tracing

Autonomous navigation is based on the planning and execution of a trajectory in which the robot travels the shortest possible distance, ensuring the integrity of previously established proxemic zones and avoiding collisions with other objects present in the environment. In a dynamic environment, with mobile objects inserted, the path planning must consider that after the creation of the first route, the robot must be able to calculate new routes according to changes in the environment, since there is the possibility of some object or some person interfering. The definition of the proxemic zones of people who are in front of the robot will determine a dynamic and real-time navigation.

4. Results and Discussion

The work carried out has, among other aspects, a strong relationship with the robot's perception of the environment, which includes the detection and classification of objects around it and determination of the proxemic zones of the detected humans and relationship with autonomous navigation, which is closely linked to the perception stage and is responsible for carrying out the movement of the robot from an initial point to an end point by the shortest possible path without exceeding the limits of the proxemic zones of the people present in the environment. Considering this subdivision of the work as a whole in the two mentioned topics, the results obtained regarding the perception of the environment (detection and classification of objects and determination of the proxemic zones of the detected individuals) and autonomous navigation (definition of the shortest route avoiding the collision with objects and invasion of established proxemic zones).

In this work, a simulation environment was developed on Gazebo to validate our framework. The scenario chosen for this work was a gas station (Figure 3), in which we obtained from the pre-existing models in Gazebo, people were placed in random positions to simulate a casual condition. The human models were made using the Make Human app, where we generated each one with a face mask.

In the robocentric perception of the environment, one can see in Figure 3, the framework is able to show several people, using different colors, and robot can identify the distance and emotion of each of them. To prove the effectiveness of the proxemic zones, two tests were carried out in which, first, the robot had to avoid passing in the middle of a group of two people, creating from the trajectory, a path around the group to its final destination. Thus, it is possible to visualize in Figure 4 the first scenario in the Gazebo, where, from the sensors of the robot, a group of two people was identified and represented in the RVIZ with a red marking, so the trajectory taken by the robot in relation to the Gaussian generated around people. It is important to note that one person has a larger Gaussian area than the other, this is due to the fact that the robot detected a emotion of sadness in her, and in the other, with the smaller Gaussian, a feeling of happiness. In Figure 5, the second scenario in the Gazebo is presented, where a group of three people in a row was identified from the sensors and represented in the RVIZ through a red marking. It is noted that, first, the robot chooses to go on the right side of the second person, and as the sensors detect the other people in the queue and the Gaussian transforms according to her feelings, it continues its way in the same path. With this, it is clear that in both tests the Pepper robot respected the proxemics in the proposed scenarios and, thus, the personal space of the people, according to their feeling.

In this work, it was proposed a framework for socially aware navigation, in which a robocentric perspective has been considered. The general idea was to integrate perception (sensors and computer vision), and a path planning strategy based on provided perception, it was discussed the effect that this perception could cause in socially aware navigation, creating a more natural path planning, based on social conventions. As discussed in the simulations results, our purpose was able to provide a suitable path planning for the Pepper robot, in a social casual scene, where there were different people, the chosen feature was the emotional state, which was used to modify proxemics zones and create a dynamic social zone.

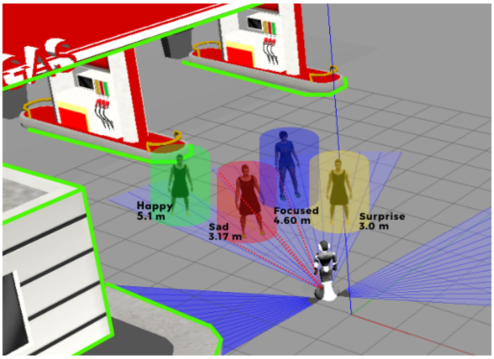
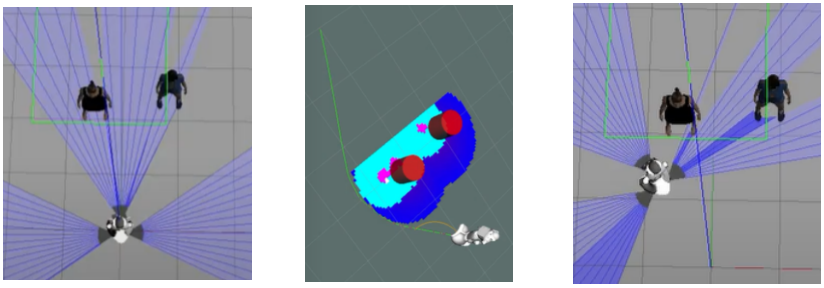


Figure 3. Framework can identify several people individually using different colors, in addition to defining the distance and their emotion



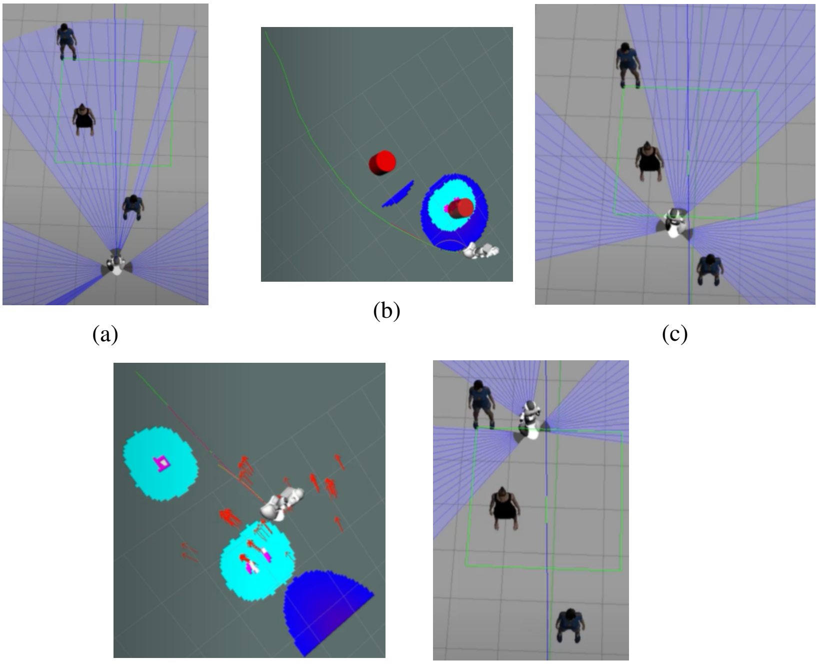
(a) (b) (c)

Figure 4. First test: (a) simulation of the scenario in the Gazebo, (b) representation of the scenario and visualization of the sensors in the RVIz, (c) trajectory taken by the robot in relation to the Gaussian presented in (b).

5. Conclusions

After the implementation and testing of our proposal, we verified its feasibility. Still, the perspective of a robot is crucial because it will bring a three-dimensional notion of the environment, but this perspective can vary depending on several aspects, such as vibrations, luminosity, environmental aspects, locomotion, among others.

For robots to be accepted traveling in the same environment of humans, it is necessary that robots respect a certain distance between them and the human. This distance can be defined by several aspects such as gender, age, ethnicity, emotion and even physical aspects. We characterize this distance as a proxemic zone. Our proposal was tested in virtual environments containing humans with different facial expressions, where the robot would have to analyze the expression and determine the minimum distance it could move from the person and at the same time it analyzes it, it traces its best trajectory. Because we are using an existing camera on the robot, its resolution is low, which makes facial analysis difficult at distances greater than 2m. However, the behavior of the robot in the face of the generation of the proxemic zone from the detected facial expression and the traced trajectory, we can conclude that the framework obtained the expected results.



(a) (b) (c)

(d) (e)

Figure 5. Second test: (a) simulation of the scenario in the Gazebo, (b) Representation of the people identified in the scenario with a red mark in the RVIz, (c) First trajectory decision taken by the robot in relation to the Gaussian formed around the people, (d)) representation in the RVIz of the development of the Gaussian according to the robot's trajectory, (e) Arrival at the destination.

In future works, we will implement these results in environments containing people in motion and with static objects, where the robot will have to trace its route avoiding people and objects. With this, the robot will have to analyze the direction in which the person is going.

Acknowledgement

This research was supported by FONDO NACIONAL DE DESARROLLO CIENTÍFICO, TECNOLÓGICO Y DE INNOVACIÓN TECNOLÓGICA - FONDECYT as executing entity of CONCYTEC under grant agreement no. 01-2019-FONDECYT-BM-INC.INV in the project RUTAS: Robots for Urban Tourism centers, Autonomous and Semantic-based.

References

[1] Nin MC, Osório F. Navegação de robôs móveis autônomos e detecção de humanos baseada em sensor laser e câmera térmica. 2011.

[2] Cunhal MJA. Sistema de visão para a interação e colaboração humano-robô: reconhecimento de objetos, gestos e expressões faciais; 2014.

[3] Redmon J, Divvala SK, Girshick RB, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:779-88.

[4] Ginés J, Martín F, Vargas D, Rodríguez FJ, Matellán V. Social navigation in a cognitive architecture using dynamic proxemic zones. Sensors. 2019;19(23):5189.

[5] Patompak P, Jeong S, Nilkhamhang I, Chong NY. Learning Proxemics for Personalized Human–Robot Social Interaction. International Journal of Social Robotics. 2020;12. Available from: https://doi.org/10.1007/s12369-019-00560-9.

[6] Daza M, Barrios-Aranibar D, Diaz-Amado J, Cardinale Y, Vilasboas J. An Approach of Social Navigation Based on Proxemics for Crowded Environments of Humans and Robots. Micromachines. 2021;12(2). Available from: https://www.mdpi.com/2072-666X/12/2/193.

[7] Ginés J, Martín F, Vargas D, Rodríguez FJ, Matellán V. Social Navigation in a Cognitive Architecture Using Dynamic Proxemic Zones. Sensors. 2019;19(23). Available from: https://www.mdpi.com/1424-8220/19/23/5189.

[8] Pérez-Higueras N, Ramón-Vigo R, Merino L, Caballero F. NAVEGACION LOCAL DE UN ROBOT MOVIL CON COSTES SOCIALES APRENDIDOS POR DEMOSTRACION.

[9] Daza M, Barrios-Aranibar D, Diaz-Amado J, Cardinale Y, Vilasboas J. An approach of social navigation based on proxemics for crowded environments of humans and robots. Micromachines. 2021;12(2):193.

[10] Kivrak H, Cakmak F, Kose H, Yavuz S. Social navigation framework for assistive robots in human inhabited unknown environments. Engineering Science and Technology, an International Journal. 2021;24(2):284-98. Available from: https://www.sciencedirect.com/science/article/pii/S2215098620308727.

[11] Hurtado JV, Londoño L, Valada A. From Learning to Relearning: A Framework for Diminishing Bias in Social Robot Navigation. Frontiers in Robotics and AI. 2021;8. Available from: https://www.frontiersin.org/article/10.3389/frobt.2021.650325.

[12] Mavrogiannis C, Alves-Oliveira P, Thomason W, Knepper RA. Social Momentum: Design and Evaluation of a Framework for Socially Competent Robot Navigation. J Hum-Robot Interact. 2022 feb;11(2). Available from: https://doi.org/10.1145/3495244.

[13] Che Y, Okamura AM, Sadigh D. Efficient and Trustworthy Social Navigation via Explicit and Implicit Robot–Human Communication. IEEE Transactions on Robotics. 2020;36(3):692-707.

[14] Pham NA, Nguyen LA, Truong XT. Socially aware robot navigation framework: Social activities recognition using deep learning techniques. In: 2021 8th NAFOSTED Conference on Information and Computer Science (NICS); 2021. p. 381-5.

[15] Mumm J, Mutlu B. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In: Proceedings of the 6th international conference on Human-robot interaction; 2011. p. 331-8.

[16] Hall ET. A system for the notation of proxemic behavior. American anthropologist. 1963;65(5):1003-26.

[17] Aghaei M, Bustreo M, Wang Y, Bailo G, Morerio P, Bue AD. Single Image Human Proxemics Estimation for Visual Social Distancing; 2020.

[18] Quigley M, Conley K, Gerkey B, Faust J, Foote T, Leibs J, et al. ROS: an open-source Robot Operating System. In: ICRA workshop on open source software. vol. 3. Kobe, Japan; 2009. p. 5.

[19] Du J. Understanding of object detection based on CNN family and YOLO. In: Journal of Physics: Conference Series. vol. 1004. IOP Publishing; 2018. p. 012029.

[20] Tang J, Luo J, Tjahjadi T, Guo F. Robust arbitrary-view gait recognition based on 3D partial similarity matching. IEEE Transactions on Image Processing. 2016;26(1):7-22.

[21] Lu DV, Allan DB, Smart WD. Tuning cost functions for social navigation. In: International Conference on Social Robotics. Springer; 2013. p. 442-51.