A Deep Learning Approach To Multi-Context Socially-Aware Navigation

Santosh Balajee Banisetty[1], Vineeth Rajamohan[2], Fausto Vega[3], and David Feil-Seifer[4]

***Abstract*— We present a context classification pipeline to allow a robot to change its navigation strategy based on the observed social scenario. Socially-Aware Navigation considers social behavior in order to improve navigation around people. Most of the existing research uses different techniques to incorporate social norms into robot path planning for a single context. Methods that work for hallway behavior might not work for approaching people, and so on. We developed a high-level decision-making subsystem, a model-based context classifier, and a multi-objective optimization-based local planner to achieve socially-aware trajectories for autonomously sensed contexts. Using a context classification system, the robot can select social objectives that are later used by Pareto Concavity Elimination Transformation (PaCcET) based local planner to generate safe, comfortable and socially-appropriate trajectories for its environment. This was tested and validated in multiple environments on a Pioneer mobile robot platform; results show that the robot was able to select and account for social objectives related to navigation autonomously.**

I. Introduction

Human-human interpersonal navigation behavior is governed by social rules, which depend heavily on the environmental context. Robots must follow social rules governing the use of space when in close proximity to humans. A socially-aware navigation (SAN) planner could allow a robot to consider social information in order to plan its movement. Advancements in planning, control, etc. allow robots to extend their operation from a controlled lab environment to real-world dynamic environments. Changes in environment mean that the social rules governing navigation interaction might also change. For social robots to be deployed and be successful in human environments, they should be able to adapt to various interaction situations. Context-aware social behavior related to navigation is important for a successful human-robot interaction (HRI). A comprehensive solution is required for socially-aware navigation, challenges common to SAN should be dealt with holistically [1], [2].

Our prior work has demonstrated in simulation that a local planner utilizing Pareto Concavity Elimination Transformation (PaCcET) could generate SAN trajectories accounting for personal space in a hallway [3]. We extended the PaCcET local planner's multi-objective optimization capabilities to different contexts like art gallery interaction, O-formations and standing in a line and validated it on a mobile robot platform [4]. We propose a learning approach using CNN and SVM to detect the on-going interaction context; we integrated the context classifier with a non-linear optimization-based local planner [4], [3] to achieve context-appropriate robot trajectories. Using autonomous context classification and a PaCcET-enabled local planner, we can achieve socially-aware navigation behaviors not just for a single context but for multiple contexts. We realize and validate our unified socially-aware navigation (USAN) architecture [2]. The remainder of this paper is structured as follows. In the next section, we review related works. In Section III, we discuss the technical details of the architecture. In Section IV, we apply our method to various scenarios on a real robot to validate the proposed approach. Finally, in Section V, discussion and future directions are presented.

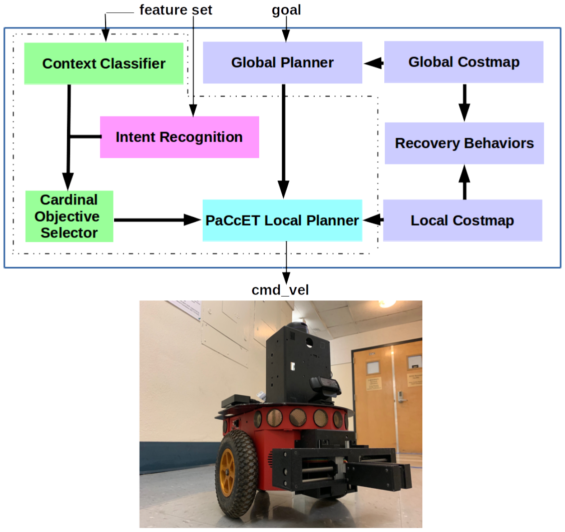


Fig. 1: **Left:** The proposed Unified Socially-Aware Navigation (USAN). The dotted lines are the modification to ROS navigation stack that we propose, blocks in blue are from ROS navigation framework. The PaCcET local planner (cyan) is from our prior work [3]. Context classification and objective selection (green) are the contributions of this paper. The intent recognition module, magenta block is an on-going work. **Right:** A Pioneer robot used to implement and validate the proposed method

II. Related Work

Methods for generating a collision-free path for robot navigation [5] do not include social norms in their algorithms. Incorporating social norms and proxemics into robot path planning algorithms, SAN, can help address HRI missteps [6]. This is especially important in dynamic human

[1]Santosh Balajee Banisetty, [2]Vineeth Rajamohan, and [4]David Feil-Seifer are with the Department of Computer Science and Engineering, University of Nevada, Reno, 1664 N. Virginia Street, Reno, NV 89557-0171, USA santoshbanisetty@nevada.unr.edu, vrajamohan@nevada.unr.edu, and dave@cse.unr.edu

[3]Fausto Vega is with the Department of Mechanical Engineering, University of Nevada, Las Vegas, 4505 S. Maryland Pkwy, Las Vegas, NV 89154, USA vegaf1@unlv.nevada.edu

environments. One of the early work in SAN, the Social Force Model (SFM) [7] uses social “forces” to consider pedestrians near a robot as external-projecting forces. This model can be extended to a group rather than just an individual to detect abnormal behavior in a crowd [8] by using a bag of words approach to classify frames as normal and abnormal. Many SAN solutions work for a single social scenario. For example, a method for hallway behavior [9], [10], a method for approaching people [11], a method for waiting in a queue [12]. Such methods solve individual challenges, but their functionalities are context-specific.

Time-dependent planning [13] combined with layered social costmap [14] generates plans that closely resembles a human-based interaction approaches. This method was applied to increase the efficiency of human-robot collaborative assembly tasks in intra-factory logistics scenarios by modeling assembly stations and operators as cost functions in a layered cost map. The preliminary experiment results showed that the system is capable of modeling both workspaces and operators in different layers and combine them with obstacle information [15]. The layered costmaps approach to SAN utilizes different costmaps for various contexts to perform socially-aware navigation by computing a master costmap [14]. However, the layered costmaps approach does not include a mechanism to autonomously select the layers (costmaps) for a sensed interaction context; thus it effectively is a single context SAN like most of the related work.

Deep reinforcement learning has been used for motion planning that accounts for social norms when navigating [16]. The robot observed and learned a policy continuously for an optimal path that will avoid collisions with humans and objects. Similar to deep reinforcement learning, inverse reinforcement learning (IRL) can plan socially-aware paths for robots based on human demonstration. By combining a feature extraction module, IRL module, and a path planning module to generate a human-like path [17]. This method was further extended for robots to navigate in a crowded environment [18] by evaluating two different IRL approaches and many feature sets in wide-scale simulation. Voronoi graph-based IRL methods can be used to efficiently explore the space of trajectories from the robots start to end position [19] for navigation in an office environment in the presence of humans. A graph-based method was applied to learn motion behavior using Bayesian IRL using sampled data [20] shows that a robot was able to learn complex navigation behaviors. Deep reinforcement learning and IRL methods for path planning problems need a considerable amount of data, computational time, and memory for a single context, let alone generalize to multiple contexts.

Our prior work modeled human navigation behavior using a Gaussian Mixture Models (GMM) using autonomously-detected features to differentiate between various interaction scenarios [21] and then extended the GMM approach to a SAN planner [9]. Taking into account interpersonal distance generates not only safe, but also comfortable social trajectories [3]. A model-based approach works well for high-level decisions, including: what context is this interaction? What objectives are essential in a sensed context, etc. On the other hand, the optimization approach requires less computational time and is suitable for low-level local planning tasks. In the next section, we will see how a combination of a model-based decision-maker and a multi-objective optimization-based local planner can be used to achieve objectives of a unified socially-aware navigation.

III. Approach

The realization of a USAN architecture presented in this paper requires visual classification of context and laser-based detection of group configurations to select appropriate navigation behavior. The USAN architecture shown in Figure 1 is implemented and tested on a pioneer mobile robot (shown in figure 1) with an upgraded camera and a long range laser setup. Appropriate behavior related to navigation can be achieved by a local planner that accounts for social normality. In this section, we discuss all the significant components of USAN architecture that include a CNN based visual context classifier, a laser-based group formation detection using SVM, and a modified local planner that utilizes non-linear optimization to generate local trajectories that are socially appropriate for an autonomously sensed context.

A. Context Dataset

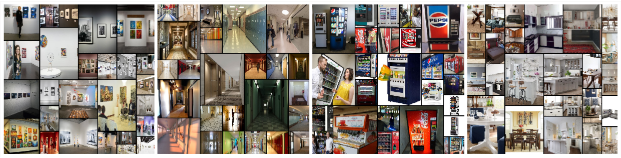


Fig. 2: A sample of images from the internet that constitute images of hallways, artwork, vending machines and other categories used for training our model.

We trained a CNN model to distinguish between four contexts (classes), art gallery, hallway, vending machine and others (anything which is not a hallway, art gallery or vending machine - we utilized images of kitchens, living rooms, and dining rooms). We collected a total of 4773 images from the internet as shown in Figure 2 and split them into training (.75), validation data (.25) and further kept aside 400 images for testing on the model as shown in Table I. The images collected were all in color, resized to 256x256 and normalized before feeding to the network. As the dataset is relatively small, data augmentation was incorporated to ensure model generalization. Augmented data includes image manipulations like zoom, shear, a shift in width, a shift in height, horizontal, and vertical flip.

Apart from the data collected from the internet, we collected real-world data at the University of Nevada, Reno to further test the model. The locations on campus, where we collected data, include buildings in the Colleges of Engineering, Science, and Humanities. The real-world data used for testing, but not part of the training process includes all the classes - hallway, art gallery, and vending machines.

Class	Train	Validation	Test
Art Gallery	1080	360	100
Hallway	804	268	100
Other	793	265	100
Vending Machine	602	201	100
Total	3279	1094	400

TABLE I: Amount of data collected for training and testing.

Other social contexts do not depend on the environmental features, but depend on the non-verbal spatial communication among people – for example, social contexts like *waiting in a queue* and *O-formations when joining a group*. To account for such non-verbal spatial communication, we collected both simulation and real-world data of people standing in a queue and O-formations using a laser scanner. We collected approximately 170 samples of each context (173 queue and 168 O-formation). A total of 341 samples, split into 80% training and 20% test data, are collected both from simulations and real-world interactions.

For the real-world samples, the *leg_tracker* package [22] detected the positions of people that were later used to calculate circularity and linearity features to train a Support Vector Machines (SVM) model to distinguish between standing in a queue and group formations.

B. Context Model

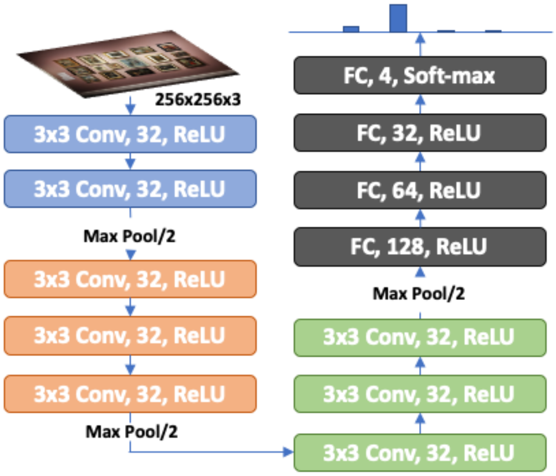


Fig. 3: USAN Context Classifier neural network architecture with 8 convolution layers, 3 max-pooling layers and 4 fully connected layers.

USAN can utilize context information to properly select the objectives specific to the sensed context for a low-level planner [3] to work with. Our approach to a context classifier is a mix of classical machine learning and neural networks. For contexts that include environmental features like hallways, we used images with a CNN architecture that resembles VGGnet [23] but with a shallow depth. For contexts that depend on non-verbal spatial communication like waiting in a queue, we used laser scanner data with a linear SVM. The CNN takes a 3-channel color image as input and outputs a probability that the image belongs to one of the four classes, as shown in Figure 3. The proposed CNN model consists of 8 convolution layers each with 32 filters, a kernel size of 3, a stride of 1x1, same padding, and ReLU activation. There are three max-pooling layers with a pool size of 2x2 to downsample between layers 2-3, 5-6, 8-9 as shown in Figure 3. The network also includes dropout regularization with every max-pooling layer and between layers 9 and 10 (between first two fully connected layers). All the fully connected layers use ReLU activation expect for the last layer which uses soft-max activation to make the predictions.

When applied to video classification task (continuous frames), the CNN model produced flickering predictions of the scene, a common problem in video classification. We used a rolling average method on the prediction probabilities to get a smooth prediction result of the scene.

As discussed earlier, there are some social contexts, such as *group formations* and *waiting in queue* which are difficult to be studied by 2-dimensional on-board cameras. However, laser data can be used to understand spatial communication [21], [9], so we used laser scan data to detect and track people in a scene [22]. The positions of the tracked people were used to calculate the following features which were later used in training a linear SVM to distinguish between *waiting in line* and *group formations*:

Circularity: It is used to describe how close a set of points should be to a true circle. The circularity of an irregular polygon formed by a set of points is given by:

$$C = (4 * \pi * area)/perimeter^2 \tag{1}$$

Where, area and perimeter of an irregular polygon are:

$$area = 1/2 \sum x_{i+1} * y_i - y_{i+1} * x_i$$

$$perimeter = \sum \sqrt{(x_{i+1} * y_i)^2 - (y_{i+1} * x_i)^2}$$

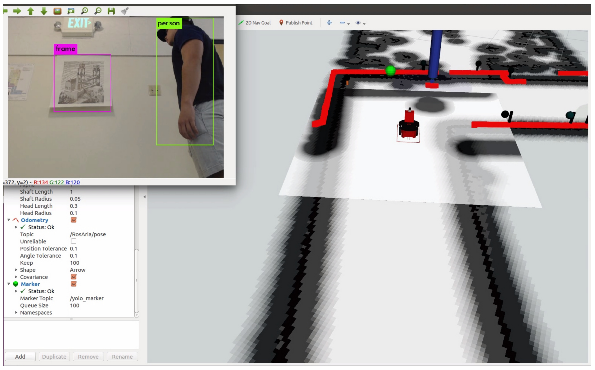
Linearity: It is the property by which a set of points can be graphically represented as a straight line. The linearity of a set of points is given by:

$$L = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \tag{2}$$

Where, n is the number of points/people.

The range of values for C and L are [0, 1]. People forming a group (circle-like) will have a C value towards 1 and L value towards 0. People forming a line will have a C value towards 0 and L value towards 1. With Circularity and Linearity features, the data is linearly separable, and hence, a linear SVM is one of the simple and ideal models for such data.

The CNN model using camera input and the SVM model using laser data are two distinct models. Depending on the confidence scores, the cardinal objectives are selected for that



Shaft Length 1
Shaft Radius 0.05
Head Length 0.3
Head Radius 0.1
Odometry ☑
✓ Status: Ok
Topic /RosAria/pose
Unreliable □
Position Tolerance 0.1
Angle Tolerance 0.1
Keep 100
Shape Arrow
Covariance ☑
Marker ☑
✓ Status: Ok
Marker Topic /yolo_marker
Queue Size 100
Namespaces

Fig. 4: Image showing object detection and tracking using YOLO-v3 and *leg_tracker* package. The image window (top-left) shows artwork and human detection using YOLO-v3. RVIZ screenshot shows human detection (dark blue cylindrical marker) using *leg_tracker* package and localization of artwork in laser data (green spherical marker).

particular context. When the detected context is “Other”, the planner switches to a sub-optimal traditional behavior.

Scikit-Learn [24] and Keras [25] with Tensorflow [26] backend was used to implement the proposed context classifier (SVM and CNN). The models were built on a computer with an Intel Core i7-8700K CPU @ 3.70GHz x 6 cores, 32 GB of RAM and GeForce GTX 1070 Ti GPU with 8GB memory. The CNN model was trained for 500 epochs with a batch size of 64 on the GPU and took approximately two hours. The model was evaluated for accuracy; the training process included Adam optimizer with categorical cross entropy loss function. The SVM model for spatial data is build on the same hardware with linear support vector classification kernel.

C. *Object Detection and Tracking*

For detection and tracking of people using a laser scanner, we used a people tracker package [22] by Leigh *et al.* To visually detect and track picture frames (for art gallery interactions), we trained YOLO-v3 [27] on Open Images Dataset. To track picture frames in 3D, we used (x, y) pixel locations in the camera to calculate the depth in the laser scanner data.

D. *PaCcET Local Planner*

We use the global trajectory planner, and low-level collision detector [5] and make adaptations to the local trajectory planner to incorporate interpersonal distance features using PaCcET. After the context classifier determines the high-level decision of navigational context, the cardinal objectives that matter most are selected. Selected objectives are then utilized by our modified local planner to account for social norms to socially navigate an environment.

The modified local planner [3], [4] using PaCcET [28] can be summarized as follows:

1) Discretely sample the robot control space.
2) Depending on the type of the robot, for each sampled velocity (V_x, V_y and V_{theta}) perform a forward simulation from the robot’s current state for a short duration to see what would happen if the sampled velocities were applied.
3) Score the trajectories based on metrics.
    a) Score each trajectory from the previous step for metrics like distance to obstacles, distance to a goal, etc. Discard all the trajectories that lead to a collision in the environment.
    b) For all the valid trajectories, calculate the social objective fitness scores like interpersonal distance and other social features and store all the valid trajectories.
4) Perform Pareto Concavity Elimination Transformation (PaCcET) on the stored trajectories to get a PaCcET fitness score and sort the trajectories from lowest to highest PaCcET fitness score.
5) For each time step, select the trajectory with the highest fitness score.

In the above working illustration of our low-level planner, step 3b is where the social objectives are accounted for while choosing the future valid trajectory points. These social objectives change from context to context and are given by the context classifier module for an autonomously sensed interaction context.

IV. RESULTS

A. *Perception*

Our CNN based context classification model was evaluated on validation data, unseen test data and real-world test data. The results are shown in sections IV-A.1, IV-A.2 and IV-A.3 respectively. The results of the SVM model distinguishing *waiting in a queue* and *O-formations* are presented in section IV-A.4. To validate our context classifier, we used the following metrics:

- Confusion matrix, defined as a matrix with elements C_{ij} representing the percentage of observations known to be in class i but predicted as class j. For a good classifier, the main diagonal elements should have the highest percentage.
- Precision, intuitively defined as the ability of a classifier not to label a negative sample as positive. It is the ratio $t_p/(t_p + f_p)$.
- Recall, intuitively defined as the ability of a classifier to find all the positive samples. It is the ratio of $t_p/(t_p + f_n)$.
- F-1 score, can be interpreted as a weighted harmonic mean of the precision and recall. Where, 1 being best and 0 being worst.

where, t_p is the number of true positives, f_p the number of false positives and f_n the number of false negatives.

1) Validation Set: The model was trained on the training set and validated on the validation set over 500 epochs. Our model achieved a 96.44% training accuracy and 94.33% accuracy on validation data.

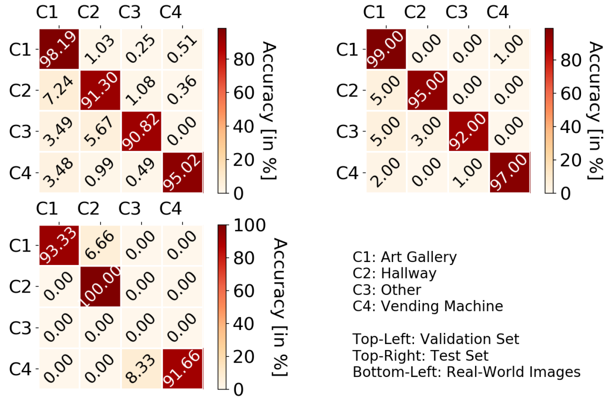


Fig. 5: Confusion matrix of validation set, test set and real-world images, showing accuracy (in percentage) for all four context.

The confusion matrix of the validation set, shown in Figure 5 shows that the model was able to learn to distinguish between an art gallery, a hallway, vending machine, and other contexts with accuracy of 98.19%, 91.30%, 95.02%, and 90.82% respectively. Table II shows performance on the validation set.

Validation set / Test set / Real-world data			
Class	Precision	Recall	F1-Score
C1	0.92 / 0.89 / 1.0	0.98 / 0.99 / 0.93	0.95 / 0.94 / 0.97
C2	0.93 / 0.97 / 0.97	0.91 / 0.95 / 1.0	0.92 / 0.96 / 0.99
C3	0.98 / 0.99 / 0.00	0.91 / 0.92 / 0.00	0.94 / 0.95 / 0.00
C4	0.98 / 0.99 / 1.0	0.95 / 0.97 / 0.92	0.97 / 0.98 / 0.96
C1: Art Gallery, C2: Hallway, C3: Other, C4: Vending Machine			

TABLE II: Performance of the CNN based context classifier.

2) Unseen Test Set: The confusion matrix of the unseen test set (Images from the internet that we kept aside) is shown in Figure 5 shows that the model was able to generalize to unseen data and was able to distinguish between an art gallery, a hallway, vending machine, and other contexts with accuracy of 99.00%, 95.00%, 97.00%, and 92.00% respectively. Table II shows performance on the unseen test set.

3) Real-World Data: To see if the model generalizes to real-world images that it has not seen, we collected 15 art gallery, 33 hallway, and 12 vending machines, a total of 60 images on campus. The "other" category is only a place-holder for any other context apart from the learned hallway, art gallery, and vending machine, so we omitted it from this test set. When in an unknown context, the planner can select default, but likely sub-optimal, objectives that will reward safe movement from one place to another. As seen in Figure 5, the model performed well on real-world images as well. The accuracy for an art gallery, hallway, and vending machine categories are 93.33%, 100.0%, and 91.66%, respectively. The performance on real-world data is presented in Table II.

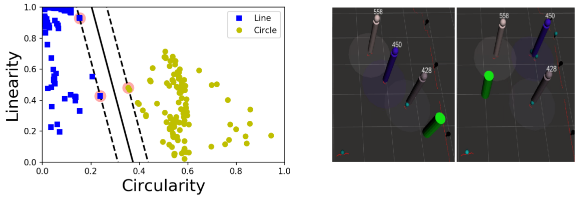


Fig. 6: **Left:** trained SVM classifier, **Right:** Social goal determined by the robot in *waiting in queue* and *O-formation* contexts.

4) Group and Queue Formations: We trained a linear SVM on location data collected from a laser scanner to classify if a group of people as *waiting in a queue* or forming a *O-formation*. We selected features like circularity, linearity, and the radius of the best-fit circle (with standardization). Later, we trained the SVM omitting the radius feature as circularity and linearity are sufficient to differentiate between the two classes, as shown in Figure 6 (left). The trained SVM achieved 100% accuracy on both training and test data. Precision, recall, and f1-scores are all 1.00 for both training and test sets. Figure 6 (right) shows an rviz screenshot of the computed social goal (green marker) determined by the robot in *waiting in queue* and *O-formation*.

B. Cardinal Objective Selection

We teleoperated the robot in an environment with *hallways*, *artwork*, *people in O-formations*, and *people waiting in queues* to test if the models can select objectives related to detected context. The results of the robot deciding on the objectives for an autonomously sensed context are shown in Figure 7, the transitions from one context to the other are shown using the vertical grid lines. Figure 7 shows that the robot is considering personal space and activity space in an *art gallery* situation. In a *hallway* situation, the robot accounts for personal space and staying on the right-side objectives. Similarly, in a *group (0-formation)* scenario, the robot considers the personal space of all the people, the O-space of the group, and the social goal of joining the group. In *waiting in a queue* context, the robot considers joining the end of the line along with the personal space of the people forming the line. It is also important to note that reaching the goal, and collision avoidance are other objectives of our PaCcET local planner.

The black box with the dotted line in Figure 7 shows the ambiguity of classification during the transition of the same group of people from *O-formation* to a *line formation*. This ambiguity is due to the quick change in the group dynamics, but misclassification for a fraction of a second should not affect the overall social performance of the planner.

C. Socially-Aware Navigation

In Sections IV-A and IV-B, we discussed the results of perception pipeline: performance of the CNN based visual classification, SVM based group scenario classification using laser data, by teleoperating the robot in an environment, we

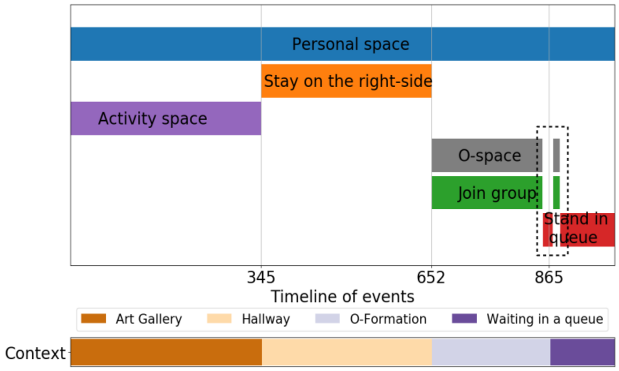


Fig. 7: Timeline showing the social objectives selected by the robot when teleoperated in an environment with hallways, artwork, people in O-formations, and people waiting in queue contexts.

showed that our method was able to detect the context accurately and thereby was able to select the cardinal objectives for that particular context.

Figure 8 shows the robot's interaction in an *art gallery* followed by a *hallway* context. In the *art gallery* context, the robot encountered one spectator viewing the art. When switching to hallway context, the robot encountered a person in a narrow hallway. The green trajectory in figure 8 **a**, **c** represents the shortest global trajectory that a traditional local planner would closely follow. In figure 8 **a**, the trajectory violates the social rule of traversing in the activity zone (space between the artwork and the spectator). In figure 8 **c**, the trajectory violates the personal space around the human in a hallway. On the other hand, in figure 8 **b**, our social planner steered the robot away from the activity space, thereby executing a socially appropriate trajectory in an art gallery. Similarly, in figure 8 **d**, our social planner steered the robot in such a way that it does not violate a person's personal space.

Figure 9 shows the robot's interaction in an *O-formation* situation followed by a *waiting in a queue* context. In both these contexts, the robot interacted with three humans. The green trajectory in figure 9 **a**, **c** represents the shortest global trajectory that a traditional local planner would closely follow. In figure 9 **a**, the trajectory planner steered the robot to the center of the group, placing it in an inappropriate location to meet with the group. In figure 9 **c**, the generated trajectory forces the robot to cut the line which is socially inappropriate. On the other hand, in figure 9 **b**, our social planner steered the robot to an appropriate location on the circle formed by the group (social goal). Similarly, in figure 9 **d**, our social planner steered the robot to the end of the line formed by the people (social goal). The social goal calculation in *O-formation* and *waiting in a queue* context is determined by geometric reasoning [4].

V. Discussion and Future Work

Our prior work [3] proposed a non-linear multi-objective optimization based PaCcET local planner using two objectives that was able to execute socially-aware behavior in a hallway setting. We then extended it to include more than two objectives to show that our PaCcET local planner can scale and extend to complex social situations like avoiding activity zones, joining a group, and waiting in a line scenarios [4]. In this paper, we concentrate on the PaCcET-enabled local planner in conjunction with a hybrid context classification method using CNN and SVM to demonstrate that architecture shown in Figure 1 can be used to exhibit socially-aware navigation behaviors in multiple social contexts.

Real-world long-term deployment of service robots require a unified socially-aware navigation method that can exhibit social navigation behavior in every social situation it might encounter in a dense human environment. Our proposed work is novel yet has certain limitations/improvements that can push USAN methods in real-world deployment. Possible improvements and future work includes the following:

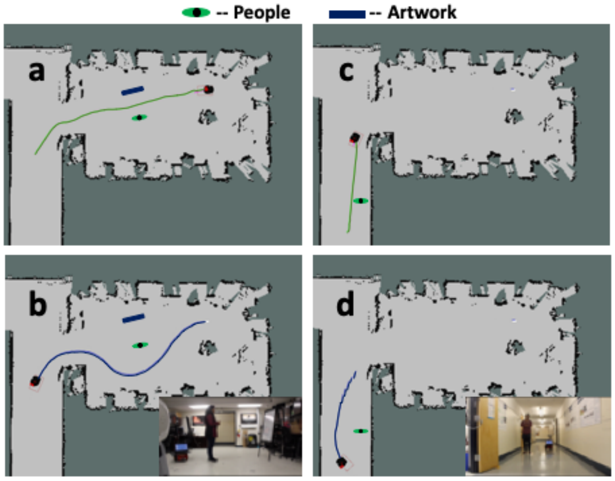


Fig. 8: Sub figures **a**, **c** shows a non-social path a robot with traditional planner would take in an *art gallery* and *hallway* contexts respectively. Sub figures **b**, **d** shows the social path our SAN planner executed.

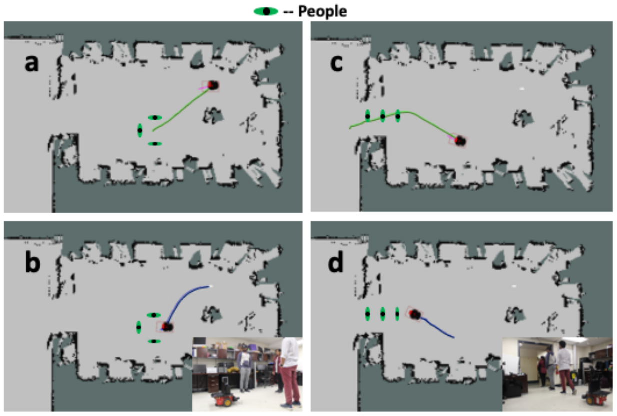


Fig. 9: Sub figures **a**, **c** shows a non-social path a robot with traditional planner would take in an *O-formation* and *waiting in a queue* contexts respectively. Sub figures **b**, **d** shows the social path our SAN planner executed.

1) The trained CNN classifier works well for the trained contexts but a better solution would be a combination of learning and reasoning. For example, the model learns what objects constitute a context, later when encountered a situation, it should reason about the correct context against a knowledge base from prior experience.
   Our ongoing efforts include building a broader knowledge base using MIT Indoor Scenes dataset [29]. Future work will augment our system to autonomously build a knowledge graph by learning the relationships between contexts and objects within the context [30].
2) The cardinal objectives are hand-picked for each trained context. Possible improvement would be to learn these objectives from human-human interactions without being explicitly told.
3) When closely observed, human-human navigational interaction benefits from intent communication and intent recognition. An intent module that can both infer and communicate navigational intentions would make our proposed method predictive system as opposed to a reactive system.

Real-world deployment of social robots that can socially navigate in a human dense human-robot environment may be far off. But it is clearly evident that social behavior in one context is not sufficient for long-term acceptance of service robots in public place. With this work, we demonstrate how differing navigation behavior is appropriate given different social and environmental contexts and that visual and laser range information can be used to autonomously sense the context.

VI. Conclusion

It is unlikely that social behavior for a single context is sufficient for long-term acceptance of service robots in public places. As robots are increasingly present in human environments, these robots need to account for social norms in various navigational contexts. There is a need for a unified architecture that can autonomously sense the ongoing navigational interaction and execute a trajectory that is socially appropriate for that particular interaction context. We presented a novel approach to a unified socially-aware navigation, discussed various subsystems, and implemented it on a robot. In this paper, we showed that a context classifier along with a low-level planner utilizing PaCcET could be used to generate socially optimal trajectories for an autonomously sensed social context. The perception system has generalized to new data and had performed well in recognizing the contexts in real human environments. The navigation results show that the robot was able to account for the social norms while performing navigational actions in various social contexts such as hallway interactions, art gallery situations, O-formations when joining a group, and waiting in queue situations.

Acknowledgement

The authors would like to acknowledge the financial support of this work by the National Science Foundation (NSF, #IIS-1719027, IIS-1757929), Nevada NASA EPSCoR (#NNX15AI02H). We would like acknowledge the help of Athira Pilla, Ashish Kasar, Anuraag Gupta, and Mounica Santhapur.

References

[1] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.

[2] S. B. Banisetty and D. Feil-Seifer, "Towards a unified planner for socially-aware navigation," in *AAAI Fall Symposium Series: AI-HRI*, October 2018.

[3] S. Forer, S. B. Banisetty, L. Yliniemi, M. Nicolescu, and D. Feil-Seifer, "Socially-aware navigation using non-linear multi-objective optimization," in *Proceedings of the International Conference on Intelligent Robots and Systems*, 2018.

[4] S. B. Banisetty, S. Forer, L. Yliniemi, M. Nicolescu, and D. Feil-Seifer, "Socially-aware navigation: A non-linear multi-objective optimization approach," 2019.

[5] E. Marder-Eppstein, E. Berger, T. Foote, B. Gerkey, and K. Konolige, "The office marathon: Robust navigation in an indoor office environment," in *2010 IEEE international conference on robotics and automation*, pp. 300–307, IEEE, 2010.

[6] B. Mutlu and J. Forlizzi, "Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction," in *Proceedings of the International Conference on Human-Robot Interaction (HRI)*, (Amsterdam, The Netherlands), pp. 287–294, ACM, 2008.

[7] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical review E*, vol. 51, no. 5, p. 4282, 1995.

[8] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 935–942, IEEE, 2009.

[9] M. Sebastian, S. B. Banisetty, and D. Feil-Seifer, "Socially-aware navigation planner using models of human-human interaction," in *International Symposium on Robot and Human Interactive Communication (RO-MAN)*, (Lisbon, Portugal), pp. 405–410, August 2017.

[10] F. Zanlungo, T. Ikeda, and T. Kanda, "A microscopic social norm model to obtain realistic macroscopic velocity and density pedestrian distributions," *PloS one*, vol. 7, no. 12, p. e50720, 2012.

[11] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans?: strategies for social robots to initiate interaction," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 109–116, ACM, 2009.

[12] Y. Nakauchi and R. Simmons, "A social robot that stands in line," *Autonomous Robots*, vol. 12, no. 3, pp. 313–324, 2002.

[13] M. Kollmitz, K. Hsiao, J. Gaa, and W. Burgard, "Time dependent planning on a layered social cost map for human-aware robot navigation," in *2015 European Conference on Mobile Robots (ECMR)*, pp. 1–6, IEEE, 2015.

[14] D. V. Lu, D. Hershberger, and W. D. Smart, "Layered costmaps for context-sensitive navigation," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pp. 709–715, IEEE, 2014.

[15] P. Santana, "Human-aware navigation for autonomous mobile robots for intra-factory logistics," in *Symbiotic Interaction: 6th International Workshop, Symbiotic 2017, Eindhoven, The Netherlands, December 18–19, 2017, Revised Selected Papers*, vol. 10727, p. 79, Springer, 2018.

[16] Y. F. Chen, M. Everett, M. Liu, and J. P. How, "Socially aware motion planning with deep reinforcement learning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1343–1350, IEEE, 2017.

[17] B. Kim and J. Pineau, "Socially adaptive path planning in human environments using inverse reinforcement learning," *International Journal of Social Robotics*, vol. 8, no. 1, pp. 51–66, 2016.

[18] D. Vasquez, B. Okal, and K. O. Arras, "Inverse reinforcement learning algorithms and features for robot navigation in crowds: an experimental comparison," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1341–1346, IEEE, 2014.

[19] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard, "Socially compliant mobile robot navigation via inverse reinforcement learning," *The International Journal of Robotics Research*, vol. 35, no. 11, pp. 1289–1307, 2016.

[20] B. Okal and K. O. Arras, "Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2889–2895, IEEE, 2016.

[21] S. B. Banisetty, M. Sebastian, and D. Feil-Seifer, "Socially-aware navigation: Action discrimination to select appropriate behavior," in *AAAI Fall Symposium Series: AI-HRI*, November 2016.

[22] A. Leigh, J. Pineau, N. Olmedo, and H. Zhang, "Person tracking and following with 2d laser scanners," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 726–733, IEEE, 2015.

[23] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, *et al.*, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[25] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[26] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pp. 265–283, 2016.

[27] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv*, 2018.

[28] L. Yliniemi and K. Tumer, "PaCcET: An objective space transformation to iteratively convexify the pareto front," in *Asia-Pacific Conference on Simulated Evolution and Learning*, pp. 204–215, Springer, 2014.

[29] A. Quattoni and A. Torralba, "Recognizing indoor scenes," pp. 413–420, 06 2009.

[30] R. Salek Shahrezaie, S. B. Banisetty, M. Mohammadi, and D. Feil-Seifer, "Towards deep reasoning on social rules for socially aware navigation," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 515–518, 2021.