Improving Socially-aware Multi-channel Human Emotion Prediction for Robot Navigation

Aniket Bera[1], Tanmay Randhavane[1], and Dinesh Manocha[2]

Abstract— **We present a real-time algorithm for emotion-aware navigation of a robot among pedestrians. Our approach estimates time-varying emotional behaviors of pedestrians from their faces and trajectories using a combination of Bayesian-inference, CNN-based learning, and the PAD (Pleasure-Arousal-Dominance) model from psychology. These PAD characteristics are used for long-term path prediction and generating proxemic constraints for each pedestrian. We use a multi-channel model to classify pedestrian characteristics into four emotion categories *(happy, sad, angry, neutral)*. In our validation results, we observe an emotion detection accuracy of 85.33%. We formulate emotion-based proxemic constraints to perform socially-aware robot navigation in low- to medium-density environments. We demonstrate the benefits of our algorithm in simulated environments with tens of pedestrians as well as in a real-world setting with Pepper, a social humanoid robot.**

I. INTRODUCTION AND OVERVIEW

Recent advances in technology predict that humans will soon be sharing spaces in public places, sidewalks, and buildings with mobile, autonomous robots. Recently, mobile robots are increasingly used for surveillance, delivery and warehousing applications. It is important that such robots navigate in socially acceptable ways, meaning that they seamlessly navigate through pedestrian traffic while responding dynamically – and appropriately – to other pedestrians.

A robot navigating through the world alone is primarily a physical problem (compute collision-free and efficient paths that satisfy the kinematics and dynamics constraints of the robot) because it has to make its way around obstacles. However, when there are other pedestrians in this environment, navigation becomes just as much about social navigation as it does about physical navigation. Humans act as both dynamic and social obstacles to a robot and have their own intentions, desires, and goals, which can affect a robot's progress. Additionally, a robot's movement may also affect humans' comfort and/or emotional state.

To predict other people's goals, people use a variety of cues, including past behavior, speech utterances, and facial expressions [15]. One of the most important predictors of people's behavior is their emotions [24], and therefore understanding people's emotional states is essential for making your way through the social world [1]. The ability to understand people's emotions is called "emotional intelligence" [34] and is useful in many social situations, including predicting

[1]Authors from the Department of Computer Science, University of North Carolina at Chapel Hill, USA
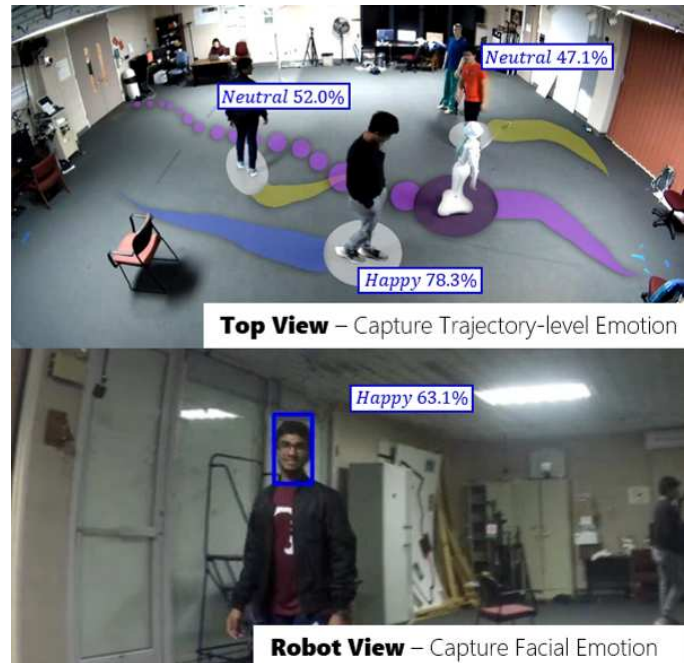[2]Author from the Department of Computer Science, University of Maryland at College Park, USA

Fig. 1: *The Emotionally Intelligent Robot: We present a real-time data-driven planning algorithm that learns the emotion state of the pedestrians to perform socially-aware navigation. (Top) The robot learns pedestrians' emotions and their proxemic constraints to improve both social comfort and navigation. (Bottom) The robot extracts facial expressions using an onboard camera and combines it with the trajectory information from the camera in (Top) to efficiently compute the overall emotions of the pedestrians and perform socially-aware navigation.*

behavior and navigation. As more robots are introduced in social settings, techniques to develop emotional intelligence of robots become increasingly important in addition to merely satisfying the physical constraints.

However, understanding the emotions of pedestrians is a challenging problem for a robot. There has been considerable research on using non-verbal cues such as facial expressions to perceive and model emotions [10]. However, recent studies in the psychology literature question the communicative purpose of facial expressions and doubt the reliability of emotions perceived only from these expressions [33]. There are many situations where facial data is only partially available or where facial cues are difficult to obtain. For example, a pedestrian may not be directly facing the robot or may be far away. Therefore, combining facial expressions with a more implicit channel of expression
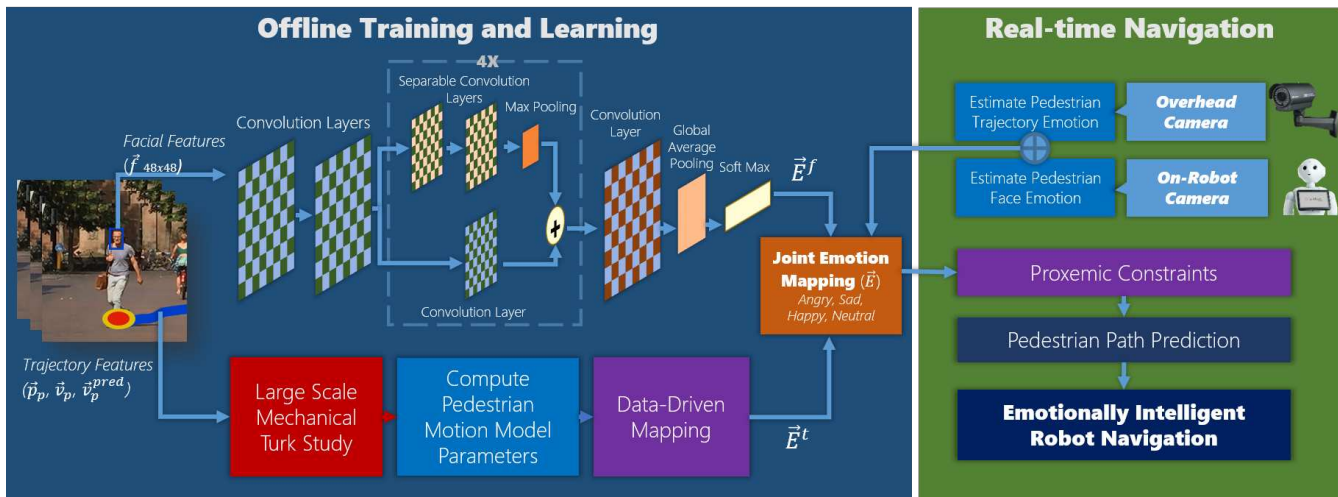
Fig. 2: *Overview: Our method takes a streaming video as input from two channels, 1) fixed, overhead camera and 2) onboard robot camera. We perform a large scale Mechanical Turk study on a crowd dataset to precompute a data-driven mapping between a motion model and their emotions. At runtime, we use this mapping along with the trajectory to compute emotion, $\vec{E}^t$, for the pedestrians. Using 2) we use a fully-convolutional neural network (which has been trained on the FER-2013 emotion dataset [14]) to compute the emotion based on facial cues, $\vec{E}^f$. We combine these multi-channel emotions with proxemic constraints and a collision-avoidance algorithm to perform socially-aware robot navigation through pedestrians.*

such as trajectories is vital for more accurate prediction of humans' emotional states.

Main Results: We present a real-time data-driven planning algorithm that takes the emotional state of the pedestrians into account to perform socially-aware navigation (Figure 1). We predict pedestrians' emotions based on the Pleasure-Arousal-Dominance (PAD) model, a 3-dimensional measure of emotional state used as a framework for describing individual differences in emotional traits/temperament [25], using information from two different channels of expression: faces and trajectories. We extract the trajectory of each pedestrian from a video stream and use Bayesian learning algorithms to compute their motion model and emotional characteristics. This trajectory-based computation is based on the results of a perception user study that provides emotion labels for a dataset of walking videos. In our validation studies, we observe an accuracy of 85.33% using 10-fold cross-validation. We also compute the facial expression-based emotion using a convolution-neural network (CNN) classifier trained on the FER-2013 emotion dataset [14]. We combine these results into a multi-channel model to classify emotion into four categories *(happy, sad, angry, neutral)*.

We combine the time-varying emotion estimates of each pedestrian with path prediction for collision-free, socially normative robot navigation. We present a new data-driven mapping, **TEM** (Trajectory-based Emotion Model), which maps learned emotions to proxemic constraints relating to the comfort and reachability distances [32]. These distances restrict the robot motion and navigation to avoid intruding through pedestrian's peripersonal and interpersonal social spaces. The combination of emotional and proxemic constraints improves both social comfort and navigation. We

have evaluated the performance of our algorithm:
- quantitatively on a dataset of real-world videos consisting of tens of pedestrians, including dense scenarios, where we measured the number of proxemic intrusions our robot avoided, and
- qualitatively in a lab setting with a Pepper humanoid robot and a total of 11 pedestrians with real-world intentions. Our subjects felt comfortable in the environment, and they could perceive the robot's subtle reaction to their emotion.

II. RELATED WORK

In this section, we discuss previous robot navigation algorithms that focus on physical and social constraints. We also review related work on emotion detection from facial expressions and trajectories.

A. Physical Navigation

Prior work on robot navigation in pedestrian environments mainly focused on solving physical constraints such as collision avoidance. Many systems for robot navigation in urban environments implemented robots that autonomously navigated previously-mapped urban paths in the presence of crowds [27], [19]. Bauer et al. [2] created a robot for navigating urban environments without GPS data or prior knowledge. Fan et al. [12] proposed a navigation framework that handles the robot freezing, and the navigation lost problems. The collision-avoidance approaches used by these navigation algorithms include potential-based approaches for robot path planning in dynamic environments [30], probabilistic methods or Bayesian velocity-obstacles methods [13], [16] for pedestrian trajectory prediction. Other methods include a partially-closed loop receding horizon control [8] for non-conservatively avoiding collisions with dynamic obstacles.

Faisal et al. [11] tuned fuzzy logic controllers with four different methods to identify which algorithm minimizes the travel time to a goal destination.

B. Socially-aware Robot Navigation

Humans navigating among crowds follow social norms relating to the speed of movement or personal space. In particular, we tend to observe the emotions displayed by others and adjust our paths in response. Correspondingly, there is much prior work on having mobile robots navigate among humans in a socially-aware manner [29], [17]. Some navigation algorithms generate socially compliant trajectories by predicting the pedestrian movement and forthcoming interactions [18] or use modified interacting Gaussian processes to develop a probabilistic model of robot-human cooperative collision avoidance [35]. Vemula et al. [37] proposed a trajectory prediction model that captures the relative importance of each pedestrian in a crowd during navigation.

C. Emotion Characteristics from Faces

In recent years, research in computer vision and AI has focused on emotion identification from facial expressions. Most of these methods use neural-network based approaches to identify emotions trained on popular datasets such as FER [14]. Liu et al. [20] presented a novel Boosted Deep Belief Network (BDBN) based three-stage training model for facial expression recognition. An annotation method, EmotioNet, predicted action units and their intensities as well as emotion category for a million facial expressions in the wild [10].

D. Emotion Characteristics from Trajectory

The direction and speed with which people move help to predict future behavior, including both pedestrians' behavior and their emotional reactions [22]. For example, people likely walk slower when they feel depressed, walk in a less direct path if they are distracted, and may change both speed and direction if they are uncertain or ambivalent about their path [28]. A clear benefit in using trajectory tracking as a way of predicting future behavior is that it is a relatively implicit measure. People are not generally aware of the information that their trajectory may be conveying, so this channel tends to have relatively high fidelity across settings and can be used for behavior or emotional classification. There is minimal work on modeling emotions from trajectories. Our work is the first approach that combines information from the trajectory channel with facial expressions to predict emotions and use them for socially-aware robot navigation.

III. EMOTION LEARNING

We propose a joint pedestrian emotion-model from trajectories and faces. In this section, we first define an emotion state, then introduce our notation, and give an overview of our approach.
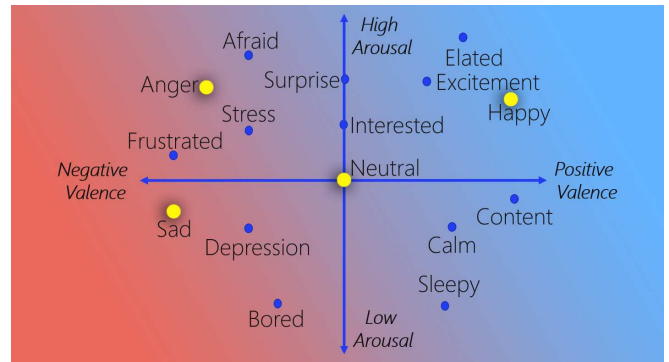

Fig. 3: *Affect Space and Discrete Emotions: All discrete emotions can be represented by points on a 2D affect space of Valence and Arousal from the PAD model [21], [9].*

A. Emotion State

Most of the previous literature has modeled emotions as either discrete categories or as points in a continuous space of emotion dimensions. Discrete categories include basic emotions such as anger, disgust, fear, joy, sadness, and surprise as well as other emotions such as pride, depression, etc. On the other hand, Ekman and Wallace [9] used "affects" to represent emotions. *Affect* is a key characteristic of emotion and is defined as a 2-dimensional space of (1) valence, the pleasure-displeasure dimension; and (2) arousal, the excited-sleep dimension. All discrete emotions can be represented by points in this 2D affect space (Fig 3). In this paper, we use the pleasure and arousal dimensions from the PAD model and use four basic emotions (*happy, angry, sad, neutral*) representing emotional states that last for an extended period and are more abundant during walking [23]. These four emotions capture the spectrum of the emotion space, and a combination of them can be used to represent other emotions [26].

B. Notation

We introduce the terminology and symbols used in the rest of the paper. We refer to an agent in the crowd as the *pedestrian* whose *state* includes his/her emotion characteristics. This state, denoted by the symbol $\vec{x}_p$, governs the pedestrian's position on the ground plane and facial features:

$$\vec{x_p} = [\vec{p}_p \ \vec{v}_p^c \ \vec{f} \ \vec{v}_p^{pred} \ \vec{E}^f \ \vec{E}^t]^\mathbf{T}; \tag{1}$$

where $\vec{p} \in \mathbb{R}^2$ is the pedestrian's position which is used to compute emotion from the trajectory; $\vec{v}^c \in \mathbb{R}^2$ is his/her current velocity; and $\vec{v}^{pred} \in \mathbb{R}^2$ is the *predicted velocity* on a 2D ground plane. A pedestrian's current velocity $\vec{v}^c$ tends to be different from their optimal velocity (defined as the predicted velocity $\vec{v}^{pred}$) that they would take in the absence of other pedestrians or obstacles in the scene to achieve their intermediate goal. $\vec{f}$ is their face pixels re-aligned to 48×48 which is used to compute the facial emotion. $\vec{E}^f \in \mathbb{R}^3$ and $\vec{E}^t \in \mathbb{R}^3$ are their facial and trajectory emotion vectors. The union of the states of all the other pedestrians and the current

positions of the obstacles in the scene is the current state of the environment denoted by the symbol $\mathbf{X}$.

We do not explicitly model or capture pairwise interactions between pedestrians. However, the difference between $\vec{v}^{pred}$ and $\vec{v}^c$ provides partial information about the local interactions between a pedestrian and the rest of the environment. Similarly, we define the robot's state, $\vec{x_r}$, as

$$\vec{x_r} = [\vec{p}_r \ \vec{v}_r^c \ \vec{v}_r^{pref}]^{\mathbf{T}}; \tag{2}$$

where $\vec{v}^{pref}$ is the *preferred velocity* of the robot, defined as the velocity the robot would take based on the present and predicted position of the pedestrians and the obstacles in the scene.

We represent the emotional state of a pedestrian by a vector $\vec{E}^t = [h, a, s]$, where h, a, and s correspond to a scalar value of happy, angry, and sad emotions (normalized to $[0, 1]$), respectively. Using the $\vec{E}^t$, we can also obtain a single emotion label e as follows:

$$e = \begin{cases} happy, & \text{if } (h > a) \land (h > s) \land (h > \theta) \\ angry, & \text{if } (a > h) \land (a > s) \land (a > \theta) \\ sad, & \text{if } (s > h) \land (s > a) \land (s > \theta) \\ neutral, & \text{otherwise} \end{cases} \tag{3}$$

where θ is a scalar threshold. In this paper, we use an experimentally determined value of $\theta = 0.55$.

C. Overview

We present an overview of our approach in Figure 2. Our method takes a streaming video as input from two camera channels, a fixed, overhead camera, and an onboard robot camera. We perform a large-scale Mechanical Turk study on a crowd dataset to establish a linear mapping *(TEM)* between the motion model parameters and pedestrian emotion using multiple linear regressions after we obtained the labels using a perception user study. Later we use this mapping and compute trajectory-based emotions for the pedestrians. We also use a fully-convolutional neural network (which has been trained on the FER-2013 emotion dataset [14]) to compute the facial expression based-emotions. We combine these multi-channel emotions along with proxemic constraints and a collision-avoidance algorithm to perform socially-aware robot navigation.

IV. EMOTION LEARNING

In this Section, we describe our joint pedestrian emotion model that combines emotion learning from trajectories and facial features.

A. Emotion Learning from Trajectories (TEM)

Our goal is to model the levels of perceived emotion from pedestrian trajectories. We use a data-driven approach to model pedestrians' emotions. We present the details of our
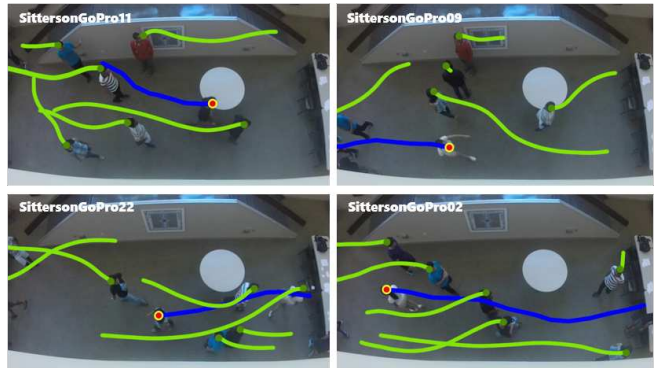

Fig. 4: *To compute a data-driven emotion mapping, we collected* 23 *videos of pedestrians walking in a corridor for our Mechanical Turk perceptual user study. Users were asked to label the emotion of one pedestrian (marked in blue).*

perception study in this section and derive the trajectory-based pedestrian emotion model (TEM) from the study results.

1) Study Goals: This web-based study is aimed at obtaining the emotion labels for a dataset of pedestrian videos. We use a 2D motion model [36] based on reciprocal velocity obstacles (RVO) to model the motion of pedestrians. We obtained scalar values of perceived emotions for different sets of motion model parameters.

2) Participants: We recruited 100 participants (77 male, 23 female, $\bar{x}_{age} = 33.24$, $s_{age} = 7.81$) from Amazon MTurk to answer questions about a dataset of simulated videos.

3) Dataset: We collected 23 videos of pedestrians walking in a corridor. In each video, a single pedestrian was highlighted by a circle and his/her trajectory (Figure 4). We computed the motion model parameters of the pedestrians using a Bayesian learning approach [16]. The motion model corresponds to the local navigation rule or scheme that each pedestrian uses to avoid collisions with other pedestrians or obstacles. In particular, we consider the following motion parameters for each pedestrian:

- Planning Horizon (how far ahead of the agent plans),
- Effective Radius (how far away an agent stays from other agents), and
- Preferred Speed.

We represent these motion model parameters as a vector ($\vec{P} \in \mathbb{R}^3$): *Planning Horiz, Radius, Pref Speed*. Table I shows the range of the values of the parameters used. These values cover the range of values observed in the real world. We will release this dataset and the motion model parameters.

Parameter (unit)	Min	Max	Average	Variance
Planning Horiz (s)	0.09	2.21	1.25	0.57
Radius (m)	0.30	0.92	0.61	0.05
Pref Speed (m/s)	0.93	2.33	1.39	0.11

TABLE I: *Values of Motion Parameters: We present the range and average values of motion parameters obtained from the dataset.*

4) Procedure: In the web-based study, participants were asked to watch a random subset of 8 videos from the dataset. Participants then answered whether the highlighted agent was experiencing one of the basic emotions (happy, angry, or sad) on a 5-point Likert scale from *strongly disagree (1) - strongly agree (5)*. Participants were presented the videos in a randomized order and could watch the videos multiple times if they wished. Before completing the study, participants also provided demographic information about their gender and age.

5) Analysis: We average the participant responses to each video to obtain a mean value corresponding to each basic emotion: V_h, V_s, V_a (normalized to $[0, 1]$). Using these values, we obtain the emotion vector $\vec{E} = [V_h, V_s, V_a]$ and the emotion label e using Equation 3.

We obtain emotion vectors $\vec{E}_i^t$ corresponding to each variation of the motion model parameters $\vec{P}_i$ for the 23 data points corresponding to 23 videos in the simulated dataset. We use this labeled data to fit a model for emotion computation using multiple linear regression. We chose linear regression because it is computationally inexpensive and easily invertible. Other forms of regressions can also be employed. TEM takes the following form:

$$\vec{E}^t = \begin{pmatrix} -0.15 & 0.00 & -0.12 \\ 0.24 & -0.61 & 0.20 \\ -0.02 & 0.79 & 0.11 \end{pmatrix} * \vec{P}. \qquad (4)$$

We can make several inferences from the values of the coefficients of the mapping between perceived emotion and the motion model parameters. The radius of the pedestrian representation affects the perception of anger negatively and the perception of sadness positively, whereas it doesn't affect the perception of happiness significantly. Therefore, increasing the radius makes pedestrians appear sad and decreasing it makes them appear angry. Similarly, we can control the value of the planning horizon to control the perception of happiness and anger. Increasing the planning horizon makes pedestrians appear angrier whereas decreasing it makes them appear happier. The preferred speed affects the perception of happiness positively and the perception of anger and sadness negatively.

We use the linear model to predict the value of a pedestrian's emotion label given the motion model parameters. We compute the accuracy of TEM using 10-fold cross-validation on our labeled dataset. We perform 100 iterations of the cross-validation and obtain an average accuracy of 85.33% using the actual and predicted emotion values.

B. Emotion Learning from Facial Features

In this section, we discuss the architecture of the neural network used to detect faces from a video stream captured by the robot and to classify the emotions for those faces. We leverage a CNN based on the Xception [7] architecture to predict the emotions, $\vec{E}^f$, from faces. Our network is fully-convolutional and contains 4 residual depth-wise separable convolutions where a batch normalization operation and a ReLU activation function follows each convolution. The final layer of our network is followed by a global average pooling and a soft-max activation function. The network has been trained on the FER-2013 dataset, which contains $35,887$ grayscale images. We choose images that belong to one of the following classes: *angry, happy, sad, neutral*.

We use a fully-convolutional network because each face is localized to a roughly square part of the video frame and detecting its emotion does not require knowledge of the "global" information about other faces in the frame. We use depth-wise separable convolutions since they need fewer numerical computations on images with a depth of more than 1 - here, we are dealing with RGB video, so it has three input channels. For more details, we refer the readers to [7].

C. Joint Pedestrian Emotion Model

We combine the computed emotions ($\vec{E}^t$ and $\vec{E}^f$) using a reliability weighted average. Since facial features are more unreliable (faces are partially visible or far away from the camera), we define the joint pedestrian emotion as:

$$\vec{E} = \frac{\alpha \vec{E}^t + \lfloor \max(E^f) + 1/2 \rfloor \vec{E}^f}{\alpha + \lfloor \max(E^f) + 1/2 \rfloor} \qquad (5)$$

where $\alpha \in [0, 1]$ is the pedestrian tracking confidence metric based on – a local collision-avoidance algorithm – as the non-linear motion model. For more details we direct our readers to [4], [3], [5]. Based on the unreliability in the facial features, Equation 5 computes a weighted average of the emotions predicted from faces and trajectories. Whenever facial emotion is unavailable, we use $\vec{E} = \vec{E}^t$. We also compute the emotion label e using Equation 3 from $\vec{E}$.

D. Socially-Aware Robot Navigation

We present an extension to Generalized Velocity Obstacles (GVO) [38], [6] that takes into account the comfort (cd_e) and reachability (rd_e) distances and enables socially-aware collision-free robot navigation through a crowd of pedestrians. The GVO algorithm uses a combination of local and global methods, where the global metric is based on a roadmap of the environment, and the local method computes a new velocity for the robot. We compute the comfort and reachability distances using the emotional labels e (Section IV-C) and use them in the computation of the new velocity for the robot (i.e., the local method of the GVO algorithm). In this novel formulation, we also take into account the dynamic constraints of the robot. Even though our socially-aware navigation algorithm is illustrated with GVO, our approach is agnostic to the underlying navigation algorithm and can be combined with other methods like potential field methods.

V. PERFORMANCE AND ANALYSIS

We have implemented our algorithm on a semi-humanoid robot, Pepper. It is about $1.2m$ tall with a top speed of $0.83m/s$ and an on-board camera with $2592x1944$ Active Pixels. We conducted experiments in a lab setting (Fig. 1).

We recruited 11 participants and asked them to assume that they are experiencing a certain emotion and walk accordingly. Previous studies show that non-actors and actors are both equally good at walking with different emotions [31]. We do not make any assumptions about how accurately the subjects acted or depicted the emotions. The participants reported being comfortable with the robot in the scene. Participants with *sad* emotions reported being given a wider space to walk by the robot. Participants with *angry* emotions reported that the robot made way for the pedestrian more swiftly. Participants with *happy* and *neutral* emotions didn't report significant changes, but some noticed a minor slow down in the robot's speed.

We also quantitatively evaluate the performance of our socially-aware navigation algorithm with GVO [38], which does not take into account proxemic or emotional constraints. We compute the number of times the non-social robot intrudes on the peripersonal and the interpersonal spaces of the pedestrians, thereby resulting in emotional discomfort. We also measure the additional time a robot with our algorithm takes to reach the goal position without any intrusions on pedestrians' comfort distances (hard constraint) and reachability distances (soft constraint). Our results (Table II) demonstrate that our robot can reach its goal with $< 25\%$ time overhead while ensuring that the proxemic spaces of the pedestrians aren't violated.

Dataset	Additional Time	Performance	Intrusions Avoided
NDLS-1	19.44%	2.89E-04 ms	31
NDLS-2	21.08%	2.12E-04 ms	26
NPLC-1	16.71%	2.29E-04 ms	30
NPLC-3	18.93%	3.09E-04 ms	22
UCSD-Peds1	24.89%	3.51E-04 ms	11
Students	9.12%	0.78E-04 ms	16
seq_hotel	11.89%	1.07E-04 ms	9
Street	11.09%	1.27E-04 ms	13

TABLE II: *Navigation Performance: A robot using our socially-aware navigation algorithm can reach its goal position (within $\tilde{1}m$ accuracy), while ensuring that the peripersonal/interpersonal space of any pedestrian is not intruded on with $< 25\%$ overhead. We evaluated this performance in a simulated environment, though the pedestrian trajectories were extracted from the original video.*

VI. Conclusions, Limitations, and Future Work

We present a real-time data-driven planning algorithm that takes the emotional state of the pedestrians into account to perform socially-aware navigation. We predict pedestrians' emotions based on the PAD model using information from two different channels of expression: faces and trajectories. We extract the trajectory of each pedestrian from a video stream and use Bayesian learning algorithms to compute his/her motion model and emotional characteristics. The computation of this trajectory-based emotion model (**TEM**) is based on the results of a perception user study that provides emotion labels for a dataset of walking videos. We also compute the facial expression-based emotion using a CNN classifier trained on the FER-2013 emotion dataset [14]. Our

work is the first approach that combines information from the trajectory channel with facial expressions to predict emotions and use them for socially-aware robot navigation.

References

[1] Lisa F. Barrett. *How emotions are made: The secret life of the brain.* Houghton Mifflin Harcourt, 2017.

[2] Andrea Bauer et al. The autonomous city explorer: Towards natural human-robot interaction in urban environments. *International Journal of Social Robotics*, 1(2):127–140, 2009.

[3] Aniket Bera, Sujeong Kim, and Dinesh Manocha. Efficient trajectory extraction and parameter learning for data-driven crowd simulation. In *Graphics Interface*, 2015.

[4] Aniket Bera and Dinesh Manocha. Realtime multilevel crowd tracking using reciprocal velocity obstacles. In *Proceedings of Conference on Pattern Recognition, Sweden*, 2014.

[5] Aniket Bera and Dinesh Manocha. Reach-realtime crowd tracking using a hybrid motion model. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 740–747. IEEE, 2015.

[6] Aniket Bera, Tanmay Randhavane, Rohan Prinja, and Dinesh Manocha. Sociosense: Robot navigation amongst pedestrians with social and psychological constraints. In *IROS*, 2017.

[7] F Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[8] N.E. Du Toit and J.W. Burdick. Robotic motion planning in dynamic, cluttered, uncertain environments. In *ICRA*, 2010.

[9] P Ekman and W V Friesen. Head and body cues in the judgment of emotion: A reformulation. *Perceptual and motor skills*, 1967.

[10] C Fabian Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M Martinez. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *CVPR*, 2016.

[11] Mohammed Faisal, Mohammed Algabri, Bencherif Mohamed Abdelkader, Habib Dhahri, and Mohamad Mahmoud Al Rahhal. Human expertise in mobile robot navigation. *IEEE Access*, 2018.

[12] Tingxiang Fan, Xinjing Cheng, Jia Pan, Pinxin Long, Wenxi Liu, Ruigang Yang, and Dinesh Manocha. Getting robots unfrozen and unlost in dense pedestrian crowds. *IEEE RAL*, 2019.

[13] C. Fulgenzi, A. Spalanzani, and C. Laugier. Dynamic obstacle avoidance in uncertain environment combining pvos and occupancy grid. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 1610–1616, 2007.

[14] Ian J Goodfellow et al. Challenges in representation learning: A report on three machine learning contests. In *International Conference on Neural Information Processing*, pages 117–124. Springer, 2013.

[15] James M Kilner, Karl J Friston, and Chris D Frith. Predictive coding: an account of the mirror neuron system. *Cognitive processing*, 2007.

[16] Sujeong Kim, Stephen J Guy, Wenxi Liu, David Wilkie, Rynson WH Lau, Ming C Lin, and Dinesh Manocha. Brvo: Predicting pedestrian trajectories using velocity-space reasoning. *The International Journal of Robotics Research*, 34(2):201–217, 2015.

[17] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726–1743, 2013.

[18] Markus Kuderer, Henrik Kretzschmar, Christoph Sprunk, and Wolfram Burgard. Feature-based prediction of trajectories for socially compliant navigation. In *Robotics: science and systems*, 2012.

[19] Rainer Kümmerle et al. Autonomous robot navigation in highly populated pedestrian zones. *Journal of Field Robotics*, 2015.

[20] Ping Liu, Shizhong Han, Zibo Meng, and Yan Tong. Facial expression recognition via a boosted deep belief network. In *CVPR*, 2014.

[21] A Loutfi, J Widmark, et al. Social agent: Expressions driven by an electronic nose. In *VECIMS*. IEEE, 2003.

[22] Umberto Castiello Luisa Sartori, Cristina Becchio. Cues to intention: The role of movement information. *Cognition*, 119(2):242–252, 2011.

[23] Y Ma, H M Paterson, et al. A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior research methods*, 2006.

[24] Seger M. Breugelmans Rik Pieters Marcel Zeelenberg, Rob M. A. Nelissen. On emotion specificity in decision making: Why feeling is for doing. *Judgment and Decision Making*, 3(1):18–27, 2001.

[25] Albert Mehrabian. Pleasure-arousal-dominance. *Current Psychology*, 14:261–292, 1996.

[26] J Mikels, B Fredrickson, et al. Emotional category data on images from the international affective picture system. *Behavior research methods*, 2005.

[27] Yoichi Morales et al. Autonomous robot navigation in outdoor cluttered pedestrian walkways. *Journal of Field Robotics*, 2009.

[28] Black SA Goodwin JS Ostir GV, Markides KS. Emotional well-being predicts subsequent functional independence and survival. *Journal of the American Geriatrics Society*, 5(48):473–478, 2000.

[29] Amit Kumar Pandey and Rachid Alami. A framework towards a socially aware mobile robot motion in human-centered dynamic environment. In *IROS*, pages 5855–5860. IEEE, 2010.

[30] N. Pradhan, T. Burg, and S. Birchfield. Robot crowd navigation using predictive position fields in the potential function framework. In *American Control Conference (ACC), 2011*, pages 4628–4633, 2011.

[31] C Roether, L Omlor, et al. Critical features for the perception of emotion from gait. *Vision*, 2009.

[32] Gennaro Ruggiero et al. The effect of facial expressions on peripersonal and interpersonal spaces. *Psychological research*, 81(6):1232–1240, 2017.

[33] J Russell, J Bachorowski, et al. Facial & vocal expressions of emotion. *Rev. of Psychology*, 2003.

[34] Mayer J. D. Salovey P. Emotional intelligence. *Imagination, Cognition and Personality*, 9(3):185–211, 1990.

[35] Peter Trautman et al. Robot navigation in dense human crowds: the case for cooperation. In *ICRA*, 2013.

[36] Jur Van Den Berg, Stephen J Guy, Ming Lin, and Dinesh Manocha. Reciprocal n-body collision avoidance. In *Robotics Research*, 2011.

[37] Anirudh Vemula, Katharina Muelling, and Jean Oh. Social attention: Modeling attention in human crowds. In *ICRA*. IEEE, 2018.

[38] David Wilkie, Jur van den Berg, and Dinesh Manocha. Generalized velocity obstacles. In *IROS*, 2009.