Modelling Multi-Channel Emotions using Facial Expression and Trajectory Cues for Improving Socially-Aware Robot Navigation

Aniket Bera
University of North Carolina
Chapel Hill
ab@cs.unc.edu

Tanmay Randhavane
University of North Carolina
Chapel Hill
tanmay@cs.unc.edu

Dinesh Manocha
University of Maryland
College Park
dm@cs.umd.edu

Abstract

Using facial expressions and trajectory signals, we present an emotion-aware navigation algorithm for social robots. Our approach uses a combination of Bayesian-inference, CNN-based learning and the Pleasure-Arousal-Dominance model from psychology to estimate time-varying emotional behaviors of pedestrians from their faces and trajectories. For each pedestrian, these PAD characteristics are used to generate proxemic constraints. We use a multi-channel model to classify pedestrian features into four categories of emotions (happy, sad, angry, neutral). We observe an emotional detection accuracy of 85.33 *percent in our validation results. In low-to medium-density environments, we formulate emotion-based proxemic constraints to perform socially conscious robot navigation. With Pepper, a social humanoid robot, we demonstrate the benefits of our algorithm in simulated environments with tens of pedestrians as well as in a real world setting.*

Figure 1: *We present a real-time data-driven planning algorithm that learns about pedestrians ' emotional state in order to navigate socially aware. The robot learns in real time about the emotions of pedestrians and their proxemic constraints. These distances limit the movement and navigation of the robot to avoid intrusion into peripersonal and interpersonal social spaces of the pedestrian. Combining emotional and proxemic constraints enhances social comfort as well as navigation.*

1. Introduction

Recent technological advances predict that people will soon be sharing spaces with mobile, autonomous robots in public places, sidewalks, and buildings. Mobile robots are increasingly being used recently for applications for monitoring, delivery, and warehousing. It is important that such robots navigate in ways that are socially acceptable, meaning that they navigate seamlessly through pedestrian traffic while responding to other pedestrians dynamically-and appropriately. This is done easily and naturally by humans, and in fact so easily that we can take this ability for granted. On the other hand, this is not particularly well done by robots. Instead, they move through crowds slowly and awkwardly, often stopping and getting into the way of people. These awkward social interactions limit the robots ' expanding role and undermine these machines ' social acceptance.

A robot navigating the world alone is primarily a physical problem (calculating collision-free and efficient paths that meet the robot's kinematics and dynamics constraints) because it has to overcome obstacles. However, navigation becomes just as much about social navigation as it does about physical navigation when there are other pedestrians in this environment. Human beings act as both dynamic and social barriers to a robot and have their own intentions, desires, and goals that may affect the progress of a robot. Furthermore, the movement of a robot may also affect the comfort and/or emotional state of humans.

There are two aspects of navigation in dynamic environments composed of pedestrians;

1. **Physical navigation:** Calculate collision-free paths that meet the robot's constraints on kinematics and dynamics.

2. **Social Navigation:** Compute paths to cause least social-discomfort among pedestrians.

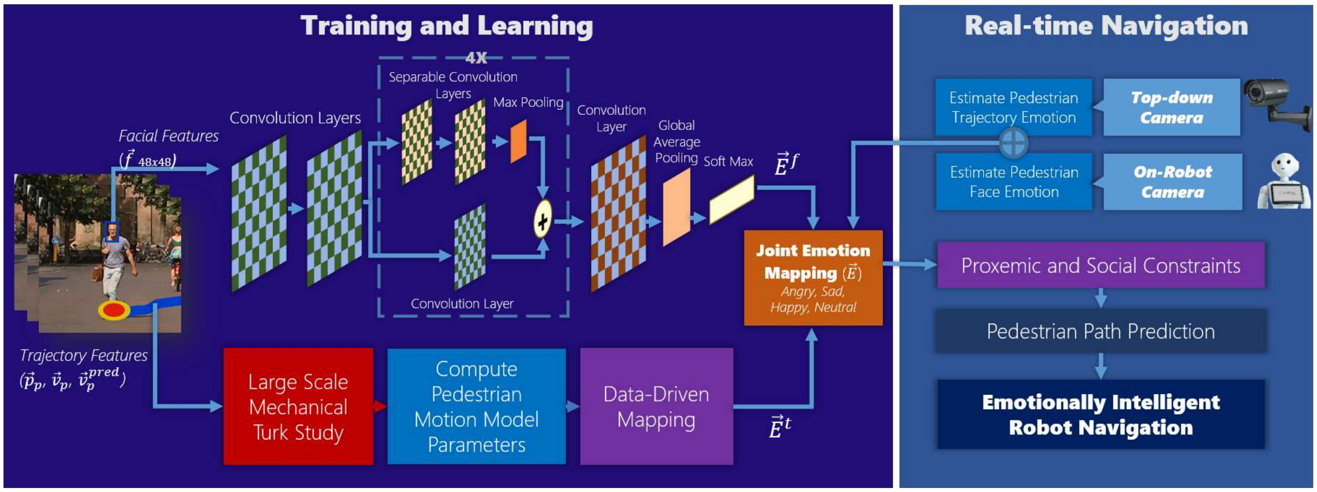

Figure 2: ***Overview****: Our method takes as input from two channels a streaming video, 1) a fixed, overhead camera, and 2) a robot camera on board. We conduct a large-scale Mechanical Turk study on a crowd dataset to pre-calculate a data-driven mapping between a movement model and its emotions. At 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 [12]) 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.*

Different individuals provide their own goals and desires with dynamic obstacles that can hinder your progress and even get angry with you. One of the human brain's key skills is its ability to predict other people's behaviors [7]. People use a variety of indications to predict the goals of other people, including past behavior, speech, and facial expressions. [13]. 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 [2]. The ability to understand people's emotions is called "emotional intelligence" [36] and is useful in many social situations, including predicting behavior and navigation. As more robots are introduced in social settings, techniques for developing robots' emotional intelligence are becoming increasingly important as well as merely satisfying physical constraints.

Comprehension of pedestrian emotions, however, is a challenging issue for a robot. Significant research has been carried out on the use of nonverbal indications such as facial expressions to perceive and model emotions [10]. Recent studies in psychological literature, however, question the communicative purpose of facial expressions and question the reliability of emotions perceived solely from these expressions [34]. There are instances when facial expressions can be unreliable [8]. There are many situations where facial data is only partially available or where it is challenging to obtain facial indications. A pedestrian, for example, may not face the robot directly or may be far from the robot. Therefore, combining facial expressions with a more implicit expression channel such as trajectories is vital to predicting the emotional states of humans more accurately. People do not always spontaneously make facial expressions, and it is unclear how much facial expressions are tied to actual behavior [17]. Walking style is not something that can be easily manipulated in the case of trajectories, making it an ideal measure and being very predictive of behaviors in analogous settings i.e. in pedestrian environments. In other words, we suggest that a combination of walking trajectories of people and their facial expression will predict their future walking behavior and their reactions to anomalous events in a given pedestrian setting.

Pedestrian/walking style is highly susceptible to change due to personal emotional experience; therefore, there is a multitude of trajectory creation applications using emotional and personality extraction. Some of many examples include increasing public safety, being able to respond to and monitor health needs and preventing dangerous situations. If a robot can learn from pedestrian trajectories by using their emotion as predictive measures, we can design more efficient evacuation plans to help the public more efficiently and safely escape dangerous situations. If we think someone tries to commit suicide while walking on a bridge or evaluating mental health signals in general, robots could predict increasing methods of providing assistance. Robots could predict if, while crossing the street, a pedestrian is likely to jaywalk and put themselves and drivers in danger. Also, robots might potentially spot criminals robbing stores, hijacking cars, or pulling a gun. The use of robots to improve the quality of our lives is constantly growing, and the need for social psychology to inform its applications is growing along with it.

Main Results: We present a real-time data-driven planning algorithm that takes into account the emotional condition of pedestrians to navigate socially (Figure 1). We predict the emotions of pedestrians based on the Pleasure - Arousal - Dominance (PAD) model, a 3-dimensional emotional state measure used as a framework to describe individual differences in emotional traits / temperament [25], using information from two different channels of expression: faces and trajectories. We extract each pedestrian's trajectory from a video stream and use Bayesian learning algorithms to calculate their emotional and movement model. This computation based on trajectories is based on the results of a perception user study that provides emotion labels for a walking video dataset. We observe an accuracy of 85.33% in our validation studies using 10-fold cross-validation. We also use a convolution - neural network (CNN) classifier trained on the FER-2013 emotion data set to calculate facial expression - based emotion [12]. We combine these results into a multi-channel model to classify emotion into four categories *(happy, sad, angry, neutral)*.

For collision-free, socially normative robot navigation, we combine the time-varying emotion estimates of each pedestrian with path prediction. We present a new data-driven mapping, **TEM** (Trajectory-based Emotion Model), mapping emotions to proxemic constraints related to distances of comfort and accessibility [33]. These distances limit the movement and navigation of the robot to avoid intrusion through the peripersonal and interpersonal social spaces of the pedestrian. 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.

The rest of the paper is organized as follows. Section II provides an overview of related work. We introduce the terminology and present our algorithm to model emotional constraints and use them for path prediction in Section III. In Section IV, we present our socially-aware robot navigation scheme. We highlight our algorithm's performance on benchmarks and describe results in Section V.

2. Related Work

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

2.1. Socially-aware Robot Navigation

People navigating among crowds are following social norms concerning motion or personal space. Specifically, we tend to observe other people's emotions and adjust our paths in response. There is, therefore, a lot of work ahead of time to have mobile robots navigate among people in a socially conscious manner [28, 18, 26, 11, 19].Some navigation algorithms generate trajectories that are socially compatible by predicting pedestrian movement and future interactions [19] or using modified Gaussian interactions to develop a probabilistic model of robot-human cooperative collision avoidance [38]. Vemula et al. [40] proposed a trajectory prediction model that captures the relative importance of each pedestrian in a crowd during navigation. Different methods tend to model interactions and personal space for human-aware navigation [1] or use learning-based approaches to account for social conventions in robot navigation [22, 29, 30]. Many explicit models have been proposed for social constraints to enable person-acceptable navigation [37, 15]. However, these methods do not take into account the psychological constraints or emotions of pedestrians.

2.2. 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 [12]. 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]. A dataset, EMOTIC, was proposed to facilitate emotion recognition given the context of the environment [16]. A recent survey on state-of-the-art methods for recognizing emotion characteristics from facial expressions is given in [32]. In this paper, we use a convolutional neural network based on the Xception [6] for emotion recognition based on facial expressions.

2.3. Emotion Characteristics from Trajectory

The direction and speed with which people move help predict future behavior, including the behavior and emotional reactions of pedestrians [23]. For example, people are likely to walk slower when depressed, walk on a less straightforward path if distracted, and may change speed and direction if they are uncertain or ambivalent about their path [27]. A clear benefit of using trajectory tracking as a way to predict future behavior is that it is a measure that is relatively implicit. People are generally not aware of the information their trajectory may convey, so this channel tends to have relatively high settings-wide fidelity and can be used for behavioral or emotional classification. There is minimal work on trajectory modeling of emotions. Our work is the first approach that combines trajectory channel information with facial expressions to predict and use emotions for robot navigation that is socially aware.

3. 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.

3.1. Emotion State

Most of the previous literature has modeled emotions either as discrete categories or as points in an emotional dimension continuous space. Discrete categories include basic emotions like anger, disgust, fear, joy, sorrow, and surprise, as well as other emotions like pride, depression, etc. On the other hand, Ekman and Wallace [9] used "affects" to represent emotions. *Affect* is a key feature of emotion and is defined as a 2-dimensional space of (1) valence, the dimension of pleasure-displeasure; and (2) excitement, the dimension of excited-sleep. All discrete emotions can be represented by points in this 2D space affect (Fig 3).

3.2. 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

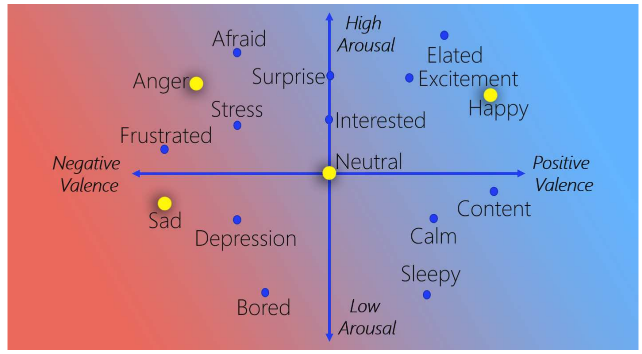

Figure 3: ***Affect Space and Discrete Emotions****: All discrete emotions can be represented by points on a 2D affect space of Valence and Arousal from the PAD model [21, 9].*

a 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:

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

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

3.3. Overview

In Figure 2, we present an overview of our approach. Our method takes a streaming video from two camera channels, a fixed overhead camera, and a robot camera on-board. Using multiple linear regressions, we perform a large-scale Mechanical Turk study on a crowd dataset to establish a linear mapping *(TEM)* between the parameters of the motion model and pedestrian emotion after obtaining the labels using a user perception study. We use this mapping later and calculate the pedestrians' trajectory-based emotions. We also use a fully-convolutional neural network (trained on the FER-2013 emotion dataset [12]) to calculate emotions based on facial expression. To perform socially-aware robot navigation, we combine these multi-channel emotions with proxemic constraints and a collision-avoidance algorithm.

4. Emotion Learning

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

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

4.1.1 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 [39] 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.

4.1.2 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.

4.1.3 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 [14]. 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:

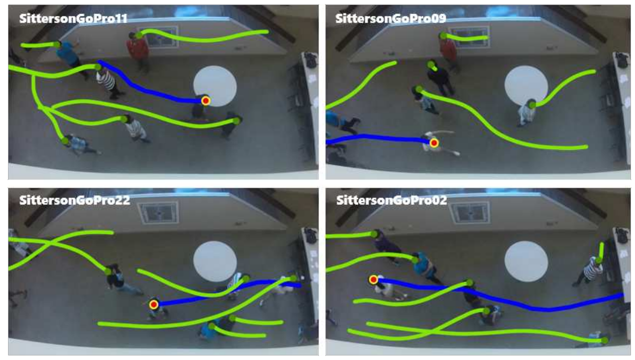

Figure 4: *To compute a data-driven emotion mapping, we collected* 23 *videos of pedestrians walking in a corridor for our Mechanical Turk perceptual user study. Users were asked to label the emotion of one pedestrian (marked in blue).*

- 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 1 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 1: ***Values of Motion Parameters****: We present the range and average values of motion parameters obtained from the dataset.*

4.1.4 Procedure

Participants were asked to view a random subset of 8videos from the dataset in the web-based study. Participants then replied whether the highlighted agent experienced one of the basic (happy, angry, or sad) emotions on a 5-point Likert scale from *strongly disagree (1) strongly agree (5)*. The videos were presented to the participants in a randomized order, and if they wanted, they could watch the videos multiple times. Participants also provided demographic information on their gender and age before completing the study.

4.1.5 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:

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

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.

4.2. 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 [6] 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*.

4.3. 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:

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

where $\alpha \in [0, 1]$ is the pedestrian tracking confidence metric based on [3]. 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 $\vec{E}$.

5. Proxemic Constraints and Path Prediction

We present our real-time approach, which calculates and uses a combination of physical and social constraints for navigation (Figure 2). Most of the previous robot navigation work is limited to avoiding collisions while taking into account kinematics and dynamic constraints. However, additional social and emotional constraints such as proximity also need to be respected when navigating in an environment alongside humans. These constraints should also be met in addition to the physical constraints, depending on the emotions or behaviors of humans in the environment. We extract pedestrian trajectories due to a pedestrian tracker that works well on crowds of low to medium density. In the trajectories extracted by the pedestrian trackers, we use Bayesian inference to compensate for the noise. A zero-mean Gaussian distribution is assumed to follow both the sensor error and the prediction error. We use an Ensemble Kalman Filter (EnKF) and Expectation Maximization (EM) [14] to estimate each pedestrian's most likely $\vec{P}$ motion model parameters. Using these parameters, we learn with TEM (Equation 4) the emotions of the pedestrians.

We formulate emotion-based proximity constraints and combine them with collision-avoidance constraints for robot navigation. Pedestrians' emotion predictions are also used for path prediction and a socially-aware robot navigation algorithm.

5.1. Pedestrian Path Prediction

Our pedestrian path prediction algorithm takes the state of all pedestrians in the environment $\mathbf{X}$ for the previous n timesteps at time t and predicts the path of a pedestrian i for a future time, $\vec{x}_i^{t+\Delta t}$. This path is predicted in the form of their motion parameters $\vec{P}_i$ because it is the best estimator of their motion in a dense environment.

The estimated motion parameters may vary slightly during the duration of motion of each pedestrian. We model these variations by an upper bound $\vec{P}_{ub}$ and a lower bound $\vec{P}_{lb}$. We compute the values of $\vec{P}_{ub}$ and $\vec{P}_{lb}$ using the values of the emotion vector $\vec{E}$ using TEM (Equation 4). Using the computed emotion label e from Equation 3, we compute $\vec{P}_{ub}$ by adding $\gamma\%$ to the emotion value corresponding to e and adding $(\gamma/3)\%$ to the other traits. Similarly, we compute $\vec{P}_{lb}$ by subtracting $\gamma\%$ from the emotion value corresponding to e and subtracting $(\gamma/3)\%$ from the other traits. The users can control the value of γ. In this paper, we use a value of $\gamma = 5\%$ that captures the noise and natural variance of the pedestrians' emotions.

We assume that for the duration of the navigation of the robot around a pedestrian, the emotion of that pedestrian does not dramatically change and lies within a range of variance. Under these assumptions, a substantial change in the emotion vector at successive timesteps indicates an error in the prediction. To account for this error, we clamp the estimated motion model parameters to the corresponding boundary value if they are out of the bounds $\vec{P}_{lb}$ and $\vec{P}_{ub}$. Otherwise, we use them directly for path prediction. We recompute the emotion vector $\vec{E}$ according to the updated motion model parameters.

$$\vec{P} = \begin{cases} \vec{P}_{lb}, & \text{if } \vec{P}_i \leq \vec{P}_{lb_i}, \forall i \\ \vec{P}_{ub}, & \text{if } \vec{P}_i \geq \vec{P}_{ub_i}, \forall i \\ \vec{P}, & \text{otherwise} \end{cases} \tag{6}$$

We use these motion parameters $\vec{P}$ to perform path prediction using the GLMP algorithm [4], which computes various local movement patterns of pedestrians as well as the crowd's overall global movement patterns.

6. Emotionally Intelligent Robot Navigation

We use the emotions computed using Equation 5 to perform socially-aware robot navigation using proxemic distances as described below.

6.1. Peripersonal and Interpersonal Social Spaces

Recent neuro-cognitive studies [33] have suggested a relationship between peripersonal-action (reachability distance) and interpersonal-social (comfort distance). In particular, reachability distance refers to the distance at which pedestrians feel comfortable interacting with other pedestrians,

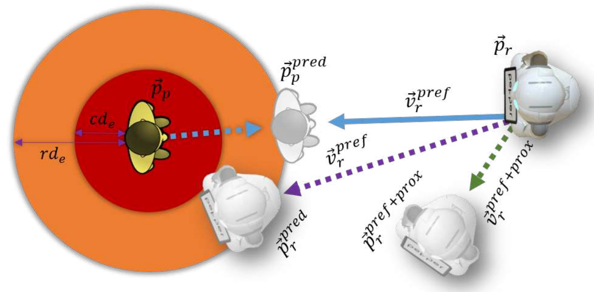

Figure 5: *Our robot navigation algorithm satisfies the proxemic distance constraints, including peripersonal space (green) and interpersonal space (blue). The trajectory computed by our algorithm does not intrude onto these spaces, whereas a robot that fails to consider the reachability distance (purple trajectory) may cause discomfort to some pedestrians.*

and comfort distance refers to the distance at which pedestrians feel comfortable with the presence of a pedestrian. These proxemic behaviors offer a window into everyday social cognition by revealing pedestrians emotional states and responses.

To enable the robot to perform socially-aware navigation, we incorporate these proxemic distances in the navigation algorithm. We compute the numerical values of *comfort distance* (cd_e) and *reachability distance* (rd_e) to perform socially-aware navigation (where e indicates the computed emotion label, Section 4.3). Though these distances depend on cultural norms, environment, or an individual's personality, we mainly focus on variations originating from the differences in emotions. Specifically, we exploit the experiments described in [33] to compute the limits on an individual's comfort and reachability distances (Table 2).

	cd_e)	rd_e
Happy	90.04	127.38
Sad	112.71	148.97
Angry	99.75	138.38
Neutral	92.03	136.09

Table 2: ***Peripersonal and Interpersonal Social Spaces:*** *The comfort and reachability distances indicate the minimum distance at which the pedestrian feels uncomfortable with the robot [33]. The distances are given in cm.*

We integrate our approach to an extension to Generalized Velocity Obstacles (GVO) [41, 5] 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.

7. Performance and Analysis

On a semi-humanoid robot, Pepper, we have implemented our algorithm. It's about $1.2m$ tall with $0.83m/s$ top speed and 2592×1944 Active Pixels on-board camera. In a laboratory setting, we conducted experiments (Fig. 1). We recruited participants from 11and asked them to assume they had a certain emotion and walk accordingly. Previous studies show that both non-actors and actors walk equally well with different emotions [31]. We make no assumptions as to how accurately the emotions were acted upon or depicted by the subjects. The participants reported that they were comfortable with the scene robot. Participants with *sad* emotions reported a wider room for the robot to walk. Participants with*angry* emotions reported that the robot was making way faster for the pedestrian. Participants with *happy* and *neutral* emotions reported no significant changes, but some noticed a slowing down in the speed of the robot. We also quantitatively evaluate the performance with GVO [41] of our socially-aware navigation algorithm, which does not take into account proxemic or emotional limitations. We calculate the number of times the non-social robot intrudes on the pedestrians' peripersonal and interpersonal spaces, leading to emotional discomfort. We also measure the extra time that a robot takes with our algorithm to reach the target position without interference with the comfort distances (hard constraint) and reachability distances (soft constraint) of the pedestrians. Our results (Table 3) show that our robot can achieve its goal with an overhead time of < 25 percent while ensuring that the pedestrians' proxemic spaces are not 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 3: ***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.*

8. Conclusions, Limitations, and Future Work

We present a real-time data-driven planning algorithm that takes into account the emotional condition of the pedestrians to perform socially conscious navigation. Using information from two different channels of expression: faces and trajectories, we predict pedestrian emotions based on the PAD model. We extract each pedestrian's trajectory from a video stream and use Bayesian learning algorithms to calculate their emotional and movement model. This trajectory-based emotion model (**TEM**) calculation is based on the results of a perception user study providing emotion labels for a walking video dataset. We also use a CNN classifier trained on the FER-2013 emotion dataset [12] to calculate the facial expression-based emotion. Our work is the first approach that combines trajectory channel information with facial expressions to predict and use emotions for robot navigation that is socially aware. There are some limitations to our approach. We assume that an overhead camera captures the pedestrian trajectories and the onboard robot camera captures the facial expressions. Both of these channels must be corrected from perspective. The classification of emotions is based on the PAD model, and only four emotions are chosen, i.e., *happy, sad, angry, and neutral*. These may not be enough to understand and capture the range of emotions observed. We want to learn emotions from full-body gaits for future work and integrate the three channels of sensors. We would also like to take into consideration group behaviors and cultural norms for socially conscious navigation. We also want to compare with an alternative representation of emotion such as the Circumplex model [35] to show that this work can be applied to different representations of emotion.

References

[1] M Barnaud et al. Proxemics models for human-aware navigation in robotics: Grounding interaction and personal space models in experimental data from psychology. In *IROS Workshop on Assistance and Service Robotics in a Human Environment*, 2014. 3

[2] Lisa F. Barrett. *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt, 2017. 2

[3] Aniket Bera et al. Adapt: real-time adaptive pedestrian tracking for crowded scenes. In *ICRA*, 2014. 6

[4] Aniket Bera, Sujeong Kim, Tanmay Randhavane, Srihari Pratapa, and Dinesh Manocha. GLMP- real-time pedestrian path prediction using global and local movement patterns. *ICRA*, 2016. 7

[5] Aniket Bera, Tanmay Randhavane, Rohan Prinja, and Dinesh Manocha. Sociosense: Robot navigation amongst pedestrians with social and psychological constraints. In *IROS*, 2017. 7

[6] 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. 3, 6

[7] Andrei A. Rusu Clifford A. Robbins Raymond A. Mar Daniel L. Schacter Demis Hassabis, R. Nathan Spreng. Imagine all the people: How the brain creates and uses personality models to predict behavior. *Cerebral Cortex*, 8(8):1979–1987, 2014. 2

[8] P. Ekman. Facial expression and emotion. *American Psychologist*, 48(4):384–392, 1993. 2

[9] P Ekman and W V Friesen. Head and body cues in the judgment of emotion: A reformulation. *Perceptual and motor skills*, 1967. 4

[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. 2, 3

[11] Gonzalo Ferrer, Anais Garrell, and Alberto Sanfeliu. Robot companion: A social-force based approach with human awareness-navigation in crowded environments. In *IROS*, 2013. 3

[12] 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. 2, 3, 5, 8

[13] James M Kilner, Karl J Friston, and Chris D Frith. Predictive coding: an account of the mirror neuron system. *Cognitive processing*, 2007. 2

[14] 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. 5, 6

[15] Rachel Kirby et al. Companion: A constraint-optimizing method for person-acceptable navigation. In *RO-MAN*, 2009. 3

[16] Ronak Kosti, Jose M Alvarez, Adria Recasens, and Agata Lapedriza. Emotion recognition in context. In *CVPR*, 2017. 3

[17] R. E. Kraut and R. E. Johston. Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, 39(9):1539–1553, 1979. 2

[18] 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. 3

[19] Markus Kuderer, Henrik Kretzschmar, Christoph Sprunk, and Wolfram Burgard. Feature-based prediction of trajectories for socially compliant navigation. In *Robotics: science and systems*, 2012. 3

[20] Ping Liu, Shizhong Han, Zibo Meng, and Yan Tong. Facial expression recognition via a boosted deep belief network. In *CVPR*, 2014. 3

[21] A Loutfi, J Widmark, et al. Social agent: Expressions driven by an electronic nose. In *VECIMS*. IEEE, 2003. 4

[22] Matthias Luber, Luciano Spinello, Jens Silva, and Kai O Arras. Socially-aware robot navigation: A learning approach. In *IROS*, 2012. 3

[23] Umberto Castiello Luisa Sartori, Cristina Becchio. Cues to intention: The role of movement information. *Cognition*, 119(2):242–252, 2011. 4

[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. 2

[25] Albert Mehrabian. Pleasure-arousal-dominance. *Current Psychology*, 14:261–292, 1996. 3

[26] Billy Okal and Kai O Arras. Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning. In *ICRA*. IEEE, 2016. 3

[27] 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. 4

[28] 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. 3

[29] N. Pérez-Higueras et al. Robot local navigation with learned social cost functions. In *ICINCO*, volume 02, pages 618–625, Sept 2014. 3

[30] R. Ramón-Vigo, N. Pérez-Higueras, F. Caballero, and L. Merino. Transferring human navigation behaviors into a robot local planner. In *Ro-Man*, pages 774–779, Aug 2014. 3

[31] C Roether, L Omlor, et al. Critical features for the perception of emotion from gait. *Vision*, 2009. 8

[32] Philipp V Rouast, Marc Adam, and Raymond Chiong. Deep learning for human affect recognition: Insights and new developments. *IEEE Transactions on Affective Computing*, 2019. 3

[33] Gennaro Ruggiero et al. The effect of facial expressions on peripersonal and interpersonal spaces. *Psychological research*, 81(6):1232–1240, 2017. 3, 7

[34] J Russell, J Bachorowski, et al. Facial & vocal expressions of emotion. *Rev. of Psychology*, 2003. 2

[35] James A Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980. 8

[36] Mayer J. D. Salovey P. Emotional intelligence. *Imagination, Cognition and Personality*, 9(3):185–211, 1990. 2

[37] Emrah Akin Sisbot, Luis F Marin-Urias, Rachid Alami, and Thierry Simeon. A human aware mobile robot motion planner. *IEEE Transactions on Robotics*, 23(5):874–883, 2007. 3

[38] Peter Trautman et al. Robot navigation in dense human crowds: the case for cooperation. In *ICRA*, 2013. 3

[39] Jur Van Den Berg, Stephen J Guy, Ming Lin, and Dinesh Manocha. Reciprocal n-body collision avoidance. In *Robotics Research*, 2011. 5

[40] Anirudh Vemula, Katharina Muelling, and Jean Oh. Social attention: Modeling attention in human crowds. In *ICRA*. IEEE, 2018. 3

[41] David Wilkie, Jur van den Berg, and Dinesh Manocha. Generalized velocity obstacles. In *IROS*, 2009. 7, 8