The Emotionally Intelligent Robot: Improving Social Human-Robot Teaming in Crowded Environments

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 [4]. One of the most important predictors of people's behavior is their emotions [5], 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" [9] 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

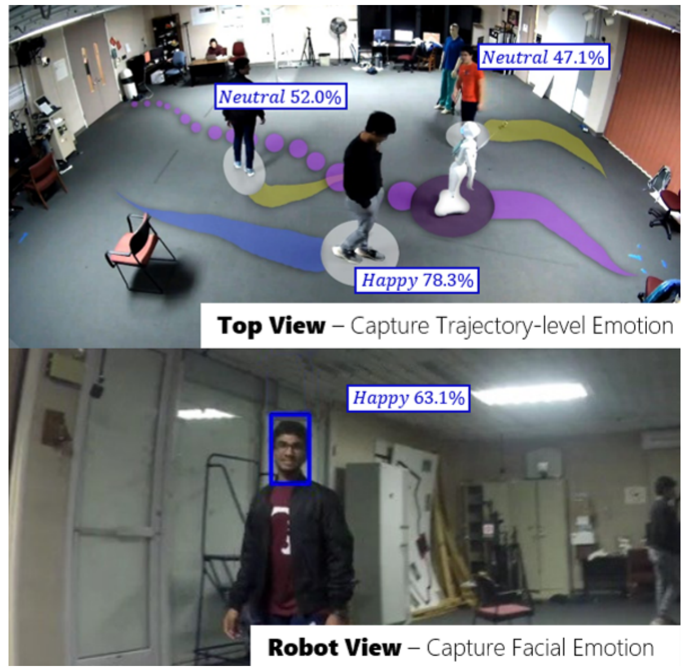

Fig. 1: ***The Emotionally Intelligent Robot****: We present a real-time data-driven planning algorithm that learns the emotion state of the pedestrians to perform socially-aware navigation. (Top) The robot learns pedestrians' emotions and their proxemic constraints to improve both social comfort and 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 [2]. 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 [8]. 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 from the robot. Therefore, combining facial expressions with a more implicit channel of expression

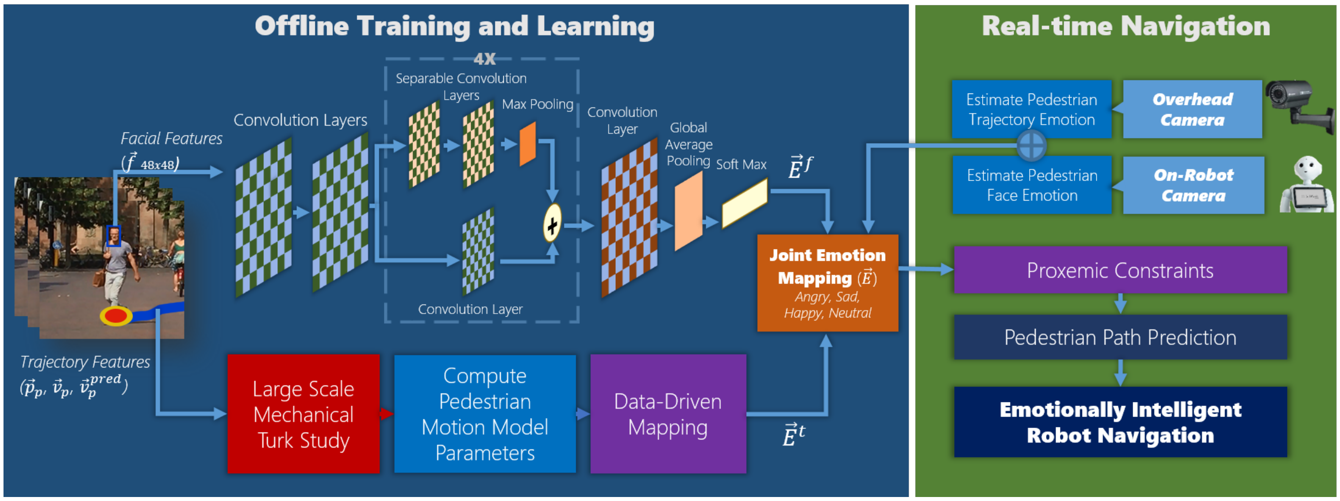

Fig. 2: ***Overview****: Our method takes a streaming video as input from two channels, 1) fixed, overhead camera and 2) onboard robot camera. We perform a large scale Mechanical Turk study on a crowd dataset to precompute a data-driven mapping between a motion model and their emotions. At runtime, we 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 [3]) 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 [6], 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 [3]. 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 [7]. 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.

References

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

[2] 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.

[3] 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.

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

[5] 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.

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

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

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

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