

CoMet: Modeling Group Cohesion for Socially Compliant Robot Navigation in Crowded Scenes

Adarsh Jagan Sathyamoorthy¹, Utsav Patel², Moumita Paul³, Nithish K Sanjeev Kumar²,
Yash Savle³, and Dinesh Manocha²

Abstract—We present CoMet, a novel approach for computing a group’s cohesion and using that to improve a robot’s navigation in crowded scenes. Our approach uses a novel cohesion-metric that builds on prior work in social psychology. We compute this metric by utilizing various visual features of pedestrians from an RGB-D camera on-board a robot. Specifically, we detect characteristics corresponding to the proximity between people, their relative walking speeds, the group size, and interactions between group members. We use our cohesion-metric to design and improve a navigation scheme that accounts for different levels of group cohesion while a robot moves through a crowd. We evaluate the precision and recall of our cohesion-metric on pedestrian datasets. We highlight the performance of our social navigation algorithm on a Turtlebot robot and demonstrate its benefits in terms of multiple metrics: freezing rate (57% decrease), deviation (35.7% decrease), and path length of the trajectory (23.2% decrease).

I. INTRODUCTION

Mobile robots are increasingly being used in crowded scenarios in indoor and outdoor environments. Applications for these robots include surveillance, delivery, logistics, etc. In such scenarios, the robots need to navigate in an unobtrusive manner and also avoid issues related to sudden turns or freezing [1].

Extensive research in social and behavioral psychology suggests that crowds in real-world scenarios are composed of (social) groups. A group is generally regarded as a meso-level concept and corresponds to two or more pedestrians with similar goals over a short or long period of time. As a result, the pedestrians or agents in a group exhibit similar movements or behaviors. It is estimated that up to 70% of observed pedestrians in real-world crowds are part of a group [2], [3]. Therefore, it is important to understand group characteristics and dynamics to perform socially-compliant robot navigation [4], [5], [6].

The problem of efficient robot navigation among pedestrians has been an active area of research. Most existing robot navigation algorithms consider walking humans or pedestrians as separate obstacles [4], [6], [8], [9]. Some techniques tend to predict trajectories of each pedestrian using learning-based methods [9] but do not account for the influence of group characteristics on individuals. This could lead to obtrusive trajectories that may cut through groups of friends or families. Other methods use simple and conservative methods to detect locally sensed clusters



Fig. 1: Our novel navigation method uses our CoMet metric to compute a collision-free trajectory for a robot in real-world scenarios. CoMet identifies groups in crowds and detects intra-group proximity, walking speed, group size and interactions to estimate a group’s cohesion. [Left] In dense scenarios, our navigation algorithm identifies a low-cohesion group (green bounding box) and navigates between its group members (green path) by assuming human cooperation for navigation. [Right] Our method detects a high cohesion group (red bounding box) and plans a trajectory around it. In both cases, other methods such as Frozone [6] (yellow path) and a DRL method [7] (orange path) approach groups with high-cohesion and halt the robot. Overall, our method improves social-compliance and the naturalness of the trajectory.

of pedestrians and compute paths around them [6]. However, they do not work well as the crowd density increases.

One characteristic of groups that could be utilized to address these problems is the social-cohesion or the closeness of the group’s members. This is directly linked to the interpersonal relationships between group members. For example, a group of friends or family has higher cohesion than a group of strangers [10], [11]. Cohesion is inversely related to the *permeability* of the group in social settings, i.e. whether another individual can cut through the group while walking [12]. Many theories have been proposed in psychology and sociology to identify the human behaviors or features that are good indicators of group cohesion. Such features include proximity between group members [13], walking speed [14], group size [12], environmental context [15], etc. Estimating cohesion could help a robot plan a better or more socially compliant trajectory based on the context. For example, in dense crowds (i.e. pedestrian density ≥ 1 person/ m^2), the robot could navigate around a group that has high cohesion or the robot could move between members of a group that has low cohesion, similar to how humans navigate in crowded scenarios.

Main Contributions: We also present a novel group cohesion metric and a formulation to compute this metric in arbitrary crowds. Our approach uses perception algorithms to identify groups in a crowd using visual features. We

This work was supported in part by ARO Grants W911NF1910069, W911NF2110026 and U.S. Army Grant No. W911NF2120076.

¹Author is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA asathyam@umd.edu

²Authors are with the Department of Computer Science, University of Maryland, College Park, MD 20742 USA upatel22@umd.edu, nithish@terpmail.umd.edu, dmanocha@gmail.com

³Authors are with the Institute of Systems Research, University of Maryland, College Park, MD 20742 USA {moumita, ysavle}@umd.edu

combine our cohesion metric with learning-based techniques to perform socially compliant robot navigation in crowded scenes. The novel components of our approach include:

- We present CoMet, a novel metric for estimating group cohesion. Our approach is based on social psychology studies and exploits features such as proximity between people, walking speeds, group sizes, and interactions. Our method uses an RGB-D sensor to detect groups and these visual features. CoMet has a near 100% precision and recall when identifying low-cohesion groups, when evaluated in real-world pedestrian or crowd datasets.
- We present a novel CoMet-based navigation method that accounts for group cohesion, while ensuring social-compliance in terms of naturalness, low deviations and robot freezing behaviors. Our formulation assumes human cooperation in dense crowds and plans less conservative trajectories than prior methods. We prove that the deviation angles computed by our method are less than or equal to the deviation angles computed using a prior social navigation algorithm [6].
- We implement CoMet on a real Turtlebot robot equipped with a commodity RGB-D sensor and demonstrate improvements in terms of social navigation. Our qualitative evaluations in dense scenes indicate that CoMet accurately identifies the cohesion in different groups (see Fig. 1). This enables the Turtlebot robot to navigate through a group based on our cohesion metric. Compared to prior social navigation algorithms, we demonstrate improved performance in terms of following metrics: freezing rate (up to 57% decrease), path deviation or turns (up to 35.7% decrease), and path length (up to 23.2% decrease).

We refer the reader to [16] for extended analysis and proofs.

II. RELATED WORKS

In this section, we briefly review prior work in robot navigation among crowds, pedestrian and group detection, and group interactions.

A. Clustering-based Group Detection

Many techniques have been proposed in computer vision for pedestrian and group detection in a crowd. The first step in group detection is to detect individuals in the images or videos. Works in this domain include many deep learning-based approaches for pedestrian detection and tracking [17] and improved methods for high density crowds [18]. These methods have been extended from individual pedestrian detection to group detection [19], [20]. Such methods typically use different kinds of clustering based on the proximities between people, their trajectories and their velocities to segregate them into groups [20], [21].

B. Group Behavior and Interaction Detection

Different techniques have been proposed for detecting group behaviors and interactions in computer vision and social psychology [15], [22] as well as event identification [23]. These works involve the analysis of different features (e.g., collectiveness, stability, uniformity) that represent how people move and interact in a crowd.

Other relevant techniques detect interactions among people in a group based on F-formations [19], [24]. These algorithms estimate features such as people's body and head

poses and identify the individuals who are facing each other. Our approach is complimentary to these methods and extends them by using other features, including proximity, walking, and interaction, to gauge group cohesion.

C. Robot Navigation and Social Compliance

Many recent works have focused on socially-compliant navigation [4], [5], [6], [25]. The underlying goal is to design methods that not only compute collision-free trajectories but also comply with social norms that increase the comfort level of pedestrians around a robot. At a broad level, the three major objectives of social navigation are comfort, naturalness, and high-level societal rules [25]. For example, a robot needs to avoid movements that are regarded as obtrusive to pedestrians by following rules related to how to approach and pass a pedestrian [4], [6].

Many techniques for social navigation have been proposed based on reinforcement learning (RL) or inverse reinforcement learning (IRL). The RL-based methods [8], [7], [6] mostly focus on treating each pedestrian as a separate obstacle to avoid collisions, sudden turns, or large deviations. IRL methods are driven by real-world natural crowd navigation behaviors [26], [27] and are used to generate trajectories with high levels of naturalness. However, they can result in unsafe trajectories and may not work well as the crowd density increases. Some methods model pedestrian behaviors by learning about their discrete decisions and the variances in their trajectories [28].

Other works have modeled human personality traits [29] or pedestrian dominance [30] based on psychological characteristics for trajectory prediction and improved navigation. Bera et al. [31] present an algorithm to avoid negative human reactions to robots by reducing the entitativity of robots. Our work on modeling group cohesion is complimentary to these methods.

III. BACKGROUND

In this section, we give an overview of prior work in social psychology, pedestrian tracking and robot navigation that is used in our approach. We also introduce the symbols and notation used in the paper.

A. Social Psychology

We use four features based on prior work in social psychology to estimate the cohesion of a group. We give a brief overview of each of these features.

Proximity: Proximity is chosen based on the proxemics principles established by Hall [13]. The underlying theory states that humans have an intimate space, a social and consultative space and a public space when interacting with others. We extend this idea to unstructured social scenarios where people or pedestrians walk, stand, or sit together. In general, humans maintain a closer proximity to other people with whom they closely interact (high cohesion).

Walking Speed: [14] studied individual and mixed gender groups' walking speeds in a controlled environment and observed significantly slower speeds when people walk with their romantic partners. *Assertion 1:* A slower-than-average walking speed in a group indicates a close relationship between group members (high cohesion).

Group Size: [12] analyzed the perceptions of people when passing through a group of two and four people in a university hallway. It was observed that people tend to

Symbols	Definitions
$\mathbf{p}^{i,t}$	Position vector of person i at time t relative to the robot/camera coordinate frame. $\mathbf{p}^{i,t} = [x^{i,t}, y^{i,t}]$
$\mathbf{v}^{i,t}$	Walking velocity vector of person i at time t relative to the robot/camera coordinate frame. $\mathbf{v}^{i,t} = [\dot{x}^{i,t}, \dot{y}^{i,t}]$
I_{rgb}^t	RGB image captured at time instant t of width w and height h .
I_{depth}^t	Depth image captured at time instant t of width w and height h .
\mathbb{B}^i	Bounding box of person with ID i .
$[x_{cen}^{\mathbb{B}^i}, y_{cen}^{\mathbb{B}^i}]$	Centroid of the bounding box \mathbb{B}^i .
$\mathbf{x}^{i,t}$	State vector used in Kalman filter for walking vector estimation for person i at time instant t . $\mathbf{x}^{i,t} = [x^{i,t}, y^{i,t}, \dot{x}^{i,t}, \dot{y}^{i,t}]^T$
ID^t	Set of person IDs detected in I_{rgb}^t .
n^k	Number of members in Group $G^{k,t}$.
N	Number of people detected in an RGB image.
$\mathbf{f}_p^i, \mathbf{f}_o^i$	Vectors for person i 's face position and orientation relative to the camera coordinate frame.
K_p, K_w, K_s K_i	Proportionality and weighing constants for each feature
$CH()$	Convex Hull function with points as its arguments.

TABLE I: List of symbols used in CoMet and their definitions.

penetrate through the 4-person group less than the 2-person group. *Assertion 2*: Humans perceive the cohesion of a bigger group to be higher (implying lower permeability) than that of a smaller group. Permeability of a group is a measure of the resistance that a moving non-group entity faces while passing between the members of a group.

Interactions: Jointly Focused Interactions (JFI) [32] entail a sense of mutual activity and engagement between people and imply their willingness to focus their attention on others. Therefore, it is chosen as an indicator for cohesion. We extrapolate JFI to signify a higher level of cohesion between group members. Visually, interactions can be detected by estimating peoples' 3-D face vectors [24] and detecting when the vectors point towards each other.

B. Notations and Definitions

We highlight the symbols and notation used, in Table I. We use i, j , and k to represent indices. All distances, angles and velocities are measured relative to a rigid coordinate frame attached to the camera (on the robot) used to capture the scene. The X-axis of this frame points outward from the camera and the Y-axis points to the left, with its origin at the center of the image. We use such a representation since our overall approach is local and has no global knowledge of the environment. The time interval between two consecutive RGB images in the stream is Δt .

C. Frozone

Frozone [6] is a navigation method that tackles the Freezing Robot Problem (FRP) [1] arising in crowds. At the same time, it can generate trajectories that are less obtrusive to pedestrians. The underlying algorithm computes a Potential Freezing Zone (PFZ), which corresponds to a configuration of obstacles where the robot's planner halts the robot and starts oscillating for a period of time when it deems that all velocities could lead to a collision. Frozone's formulation is conservative and not suitable for dense crowd navigation (> 1 person/ m^2). In addition, it treats pedestrians as individual obstacles. The main steps in Frozone's formulation are: (1) Identifying *potentially freezing* pedestrians who could cause freezing based on their walking vectors and predicting

their positions after a time horizon t_h ; (2) Constructing a Potential Freezing Zone (PFZ_{froz}) as the convex hull of the predicted positions $\hat{\mathbf{p}}_{pf}^{i,t+t_h}$ (see 5-sided convex polygon in Fig. 3) for all the *potentially freezing* pedestrians. PFZ_{froz} is formulated as,

$$PFZ_{froz} = \text{ConvexHull}(\hat{\mathbf{p}}_{pf}^{i,t+t_h}), \quad i \in 1, 2, \dots, P_f. \quad (1)$$

and (3) Computing a deviation angle for the robot to avoid PFZ_{froz} if its current trajectory intersects with it. PFZ_{froz} corresponds to the set of locations where the robot has the maximum probability of freezing and being obstructive to the pedestrians around it. The deviation angle to avoid it is computed as

$$\phi_{froz} = \min(\phi_1, \phi_2), \quad (2)$$

where ϕ_1 and ϕ_2 are given by,

$$\phi_1 = \underset{R_{z,\phi_1} \mathbf{v}_{rob} t_h \notin PFZ_{froz}}{\text{argmin}} \left(\text{dist}(R_{z,\phi_1} \cdot \mathbf{v}_{rob} \cdot t_h, \mathbf{g}_{rob}) \right), \quad (3)$$

$$\phi_2 = \tan^{-1}(y^{near,t}/x^{near,t}), \quad \phi_2 \neq 0. \quad (4)$$

Here, R_{z,ϕ_1} is the 3-D rotation matrix about the Z-axis (perpendicular to the plane of the robot), $\mathbf{v}^{rob}, \mathbf{g}^{rob}$ represent the current velocity and the goal of the robot, and $[x^{near,t}, y^{near,t}]$ denotes the current location of the nearest freezing pedestrian relative to the robot. This point is in the PFZ's exterior. For navigating the robot towards its goal, and handling static obstacles and dense crowds, a Deep Reinforcement Learning (DRL)-based method [7] is used. However, the resulting navigation may cut through groups regardless of their cohesion (see Fig. 1[left]). As a result, the robot's trajectory may not be socially compliant.

IV. CoMET: MODELING GROUP COHESION

In this section, we present our group cohesion metric, which first classifies pedestrians into groups and then measures their cohesion. Our method runs in real-time, taking a continuous stream of RGB and depth images as input and detecting the group features highlighted in Section III-A. We assume that there is sufficient distance between a group and the robot, for it to detect features and compute a group's cohesion from its perspective. Our overall approach based on these features is shown in Fig. 2.

A. Detecting Group Features

In this section, we first describe how we track and localize people, detail conditions for a set of people to be classified as a group, and then explain efficient techniques to detect group features from RGB and depth images.

B. Pedestrian Tracking and Localization

A key issue in detecting the features mentioned in Section III-A is to first detect, track, and localize each pedestrian position relative to the camera frame in a continuous stream of RGB images. We use YOLOv5 [33] and Deep Sort [34] algorithms to detect and track people, respectively. YOLOv5 outputs a set of bounding boxes $\mathcal{B} = \{\mathbb{B}^i\}$ for each detected pedestrian i in an RGB image at time instant t (denoted as I_{rgb}^t). \mathbb{B}^i is denoted using its top-left and bottom-right corners in the image-space or pixel coordinates. In addition, we also assign a unique integer number as an ID for each detected pedestrian.

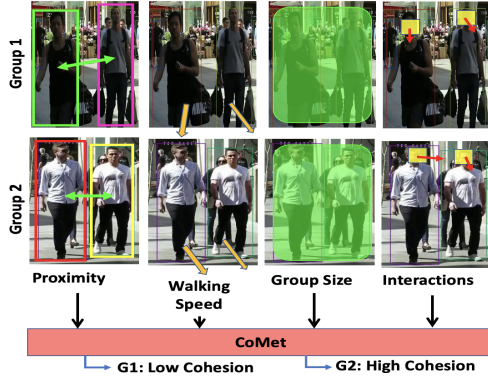


Fig. 2: We highlight how our approach is used on two different groups of pedestrians at a given time instant. Group 1 consists of two pedestrians in close proximity, walking away from each other. Their distance is increasing in subsequent frames and they are looking in different directions. This implies little interaction and low group cohesion. Group 2 has two people walking together with their faces oriented towards each other, which indicates high interaction and a high group cohesion.

Next, to accurately localize people, the distance of each detected person relative to the camera coordinate frame must be estimated. To this end, we use a depth image I_{depth}^t , every pixel of which contains the proximity (in meters) of an object at that location of the image. The pixels in I_{depth}^t contain values between a minimum and maximum distance range, which depends on the camera used to capture the image.

1) *Group Classification*: Let us consider any set of people's IDs $G^{k,t} \subseteq ID^t$ (set of IDs at time t). At any time t , if the following conditions hold,

$$\|\mathbf{p}^{i,t} - \mathbf{p}^{j,t}\|_2 \leq \Gamma \quad \forall i, j \in G^{k,t} \quad (5)$$

$$\|\mathbf{p}^{i,t} - \mathbf{p}^{j,t}\|_2 \geq \|\mathbf{p}^{i,t} + \mathbf{v}^{i,t}\Delta t - (\mathbf{p}^{j,t} + \mathbf{v}^{j,t}\Delta t)\|_2. \quad (6)$$

then the set $G^{k,t}$ is classified as a *group* in the image I_{rgb}^t . Here, Γ is a distance threshold set manually. The first condition ensures that people are close to each other and the second condition ensures that the group members walk in the same direction. When $\|\mathbf{v}^i\|, \|\mathbf{v}^j\| = 0$ (static groups), only the first condition is used for grouping.

2) *Estimating Proximity*: To estimate the proximity between people at time instant t , first the bounding boxes detected in the RGB image by YOLOv5 are superimposed over the depth image. To estimate the distance of a person i from the camera ($d^{i,t}$), the mean of all the pixel values within a small square centered around $[x_{cen}^{i,t}, y_{cen}^{i,t}]$ is computed. The angular displacement $\psi^{i,t}$ of person i relative to the camera can be computed as, $\psi^{i,t} = \left(\frac{w - x_{cen}^{i,t}}{w}\right) * FOV_{RGBD}$.

Here, FOV_{RGBD} is the field of view of the RGB-D camera. Person i 's location relative to the camera can be computed as $[x^{i,t}, y^{i,t}] = d^{i,t} * [\cos \psi^{i,t}, \sin \psi^{i,t}]$. The distance between a pair of people i and j can then be computed as, $dist(i, j) = \sqrt{(x^{i,t} - x^{j,t})^2 + (y^{i,t} - y^{j,t})^2}$.

3) *Estimating Walking Speed and Direction*: To estimate the i^{th} person's walking vector \mathbf{v}^i in I_{rgb}^t , we use a Kalman filter with a constant velocity motion model. All the detected people in I_{rgb}^t (with their IDs stored in the set ID^t) are modeled using the state vector \mathbf{x}^t defined in Table I. If ID^t contains IDs that were not present in $ID^{t-\Delta t}$, we initialize

their corresponding state vectors \mathbf{x}^t with constant values. For all the pedestrians who were detected in previous RGB images, i.e., with previously initialized states, we update their states using the standard Kalman prediction and update the steps [35]. We use a zero mean Gaussian noise with a pre-set variance to model the process and sensing noise.

4) *Estimating Group Size*: The size of the group can be trivially computed as the number of IDs in the set $G^{k,t}$.

5) *Detecting Interactions*: We use two 3-D vectors to represent the position and orientation of a person's face in I_{rgb}^t . We use OpenFace [36] to localize person i 's face position relative to the camera coordinate frame (\mathbf{f}_p^i) on an RGB image and to obtain a unit vector (\mathbf{f}_o^i) for the face orientation. Two individuals are considered to be interacting if their face positions and orientations satisfy the following condition:

$$\|\mathbf{f}_p^i - \mathbf{f}_p^j\|_2 > \|\mathbf{f}_p^i + \mathbf{f}_o^i - (\mathbf{f}_p^j + \mathbf{f}_o^j)\|_2. \quad (7)$$

This condition checks if the distance between two people's face positions is greater than the distance between the points computed by extrapolating the face orientations (see dashed lines in Fig. 1), if they are facing each other. We reasonably assume that non-interacting people do not face each other.

C. Cohesion Components

We now discuss how the detected group features can be used for cohesion estimation. Using multiple features to compute cohesion is advantageous in scenarios where not all features are properly able to be sensed.

1) *Proximity Cohesion Score*: We use the average distance between group members to model Hall's proxemics theory as previously extrapolated. As observed in Section III-A, cohesion between people is inversely proportional to the distance between them. Therefore, the cohesion score due to proximity is formulated as the reciprocal of the mean distance between group members as

$$C_p(G^{k,t}) = K_p \cdot \frac{n^k}{\sum_{i,j \in G^{k,t}} dist(i,j)}, \quad i \neq j. \quad (8)$$

2) *Walking Speed Cohesion Score*: Based on the discussion in Section III-A, we next compare the average walking speeds of each group with the average walking speed of all the detected people in I_{rgb}^t . Therefore, the cohesion score for a walking group ($\|\mathbf{v}^j\| \neq 0 \forall j \in G^{k,t}$) due to its walking speed is formulated as

$$C_w(G^{k,t}) = K_w \cdot \left(\frac{\sum_{i \in ID^t} \|\mathbf{v}^i\|}{N} \right) / \left(\frac{\sum_{j \in G^{k,t}} \|\mathbf{v}^j\|}{n^k} \right). \quad (9)$$

This reflects assertion 1 made in Section III-A, since cohesion is inversely proportional to walking speed. The average walking speed of the entire scene is included in this formulation to normalize out the effects of crowding in the scene. If $\sum_{j \in G^{k,t}} \|\mathbf{v}^j\| = 0$, i.e., the group is static, then

$C_w(G^{k,t}) = K_w \cdot \eta$, where η is a user-set large constant value.

3) *Group Size Cohesion Score*: Based on assertion 2 in Section III-A, the cohesion of a group k is directly proportional to the group size (n^k). Therefore, the group size cohesion score is computed as

$$C_s(G^{k,t}) = K_s \cdot n^k. \quad (10)$$

4) *Interaction Cohesion Score*: The interaction condition between any two people in a group (Equation 7) can be applied to all pairs in a group, and its contribution to the cohesion score of a group can be re-written as

$$C_i(G^{k,t}) = K_i \cdot \frac{1}{n^k} \cdot \sum_{\substack{i \neq j \\ i,j \in G^{k,t}}} \frac{\text{sign}(\theta_{ij})}{\cos \theta_{ij}} \quad \theta_{i,j} \in [-\frac{\pi}{4}, \frac{\pi}{4}]. \quad (11)$$

Here θ_{ij} is the angle between face orientation vectors \mathbf{f}_o^i and \mathbf{f}_o^j in the X-Y plane of the camera coordinate system. $\text{sign}()$ is the signum function. θ_{ij} is limited to $[-\frac{\pi}{4}, \frac{\pi}{4}]$ since face orientations are accurate in this range. Intuitively, we want the cohesion score to be positive and greater than 1, when people are facing towards each other and negative otherwise. Therefore, we choose the ratio $\frac{\text{sign}(\theta)}{\cos \theta}$, as it belongs to the range $[-\sqrt{2}, -1) \cup \{0\} \cup [1, \sqrt{2}]$ when $\theta_{i,j} \in [-\frac{\pi}{4}, \frac{\pi}{4}]$. Since $\cos()$ is an even function, the $\text{sign}()$ function ensures that the formulation is sensitive to the sign of the angle.

D. CoMet: Overall Group Cohesion Metric

Using the individual cohesion scores in Equations 8, 9, 10, 11, the total cohesion score for a group at time t ($G^{k,t}$) is given as

$$C_{tot}(G^{k,t}) = (C_p + C_w + C_s + C_i). \quad (12)$$

Here, $G^{k,t}$ is omitted in the RHS for readability. Note that K_p, K_w, K_s, K_i weigh the different features before adding them to the total cohesion score. If any of the features are not detectable, their contribution to C_{tot} will be zero. This acts as a measure of confidence, as our approach is able to better compute a group's cohesion when more features are detected. Our formulation is not learning-based due to the lack of extensive datasets with annotations of cohesion or related metrics for groups.

Proposition IV.1. *The value of the overall cohesion metric $C_{tot}(G^{k,t})$ for a group is bounded.*

We refer the reader to [16] for a proof of this proposition. We use the bounds on the cohesion metric to compute appropriate thresholds that are used to categorize groups as low-, medium- and high-cohesion groups.

V. COHESION-BASED NAVIGATION

In this section, we present our socially-compliant navigation algorithm, which uses the group cohesion metric.

A. Socially-Compliant Navigation

Our objective is to improve the naturalness of a robot's trajectory. We attribute three qualities to natural trajectories: (1) Not suddenly halting or freezing (avoiding FRP), (2) Low deviation angles, (3) Not cutting between high cohesion groups (friends, families etc) in a crowd. This is in accordance with humans' natural walking behaviors [12], [37].

We extend Frozone [6] (Section III-C) by considering groups and their cohesions, and prove that our proposed method leads to smaller deviations from the robot's goal, and shorter trajectory lengths. It also does not navigate the robot through high cohesion groups. We assume a higher density in the environment (in terms of crowds and static obstacles) than Frozone's formulation and human cooperation for the robot's navigation.

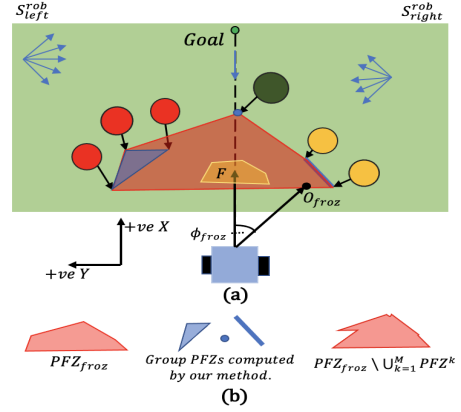


Fig. 3: (a) This scenario shows two groups (red and yellow agents in close proximity) and an individual pedestrian (green) walking towards the robot. The green rectangle denotes the robot's sensing region and the blue arrows denote the potentially freezing walking directions within each half of the sensing region ($S_{left/right}^{rob}$). (b) The PFZ computed by Frozone, our approach's group PFZs, and the set difference between them are shown. Our proposed method is less conservative and results in a smaller deviation from PFZs (no deviation needed in this case) than Frozone [6]. We also highlight F , one possible subset of $PFZ_{froz} \setminus \bigcup_{k=1}^M PFZ^k$, which contains positions with deviations $\leq \phi_{froz}$.

B. CoMet-based Navigation

To improve the naturalness of a robot's trajectory, our proposed method includes the following steps: (1) Identifying potentially freezing groups within the sensing region of the robot and predicting their positions after a time period t_h ; (2) Constructing a PFZ for each group using the predicted future locations of each group member (see blue regions in Fig. 3); (3) Computing a deviation angle to avoid the group PFZs while accounting for every group's cohesion. If a feasible solution is not found, the robot navigates between the group with the lowest cohesion in the scene.

Definition V.1 (Potentially Freezing Groups:) Groups of pedestrians that have a high probability of causing FRP after time t_h . Such groups are identified based on conditions of their average walking direction [6] (see blue arrows in Fig. 3). Groups that satisfy these conditions move closer to the robot as time progresses (proven in [6]). We predict the future positions of the potentially freezing group members as

$$\hat{\mathbf{p}}_{pf}^{i,t+t_h} = \mathbf{p}_{pf}^{i,t} + \mathbf{v}_{avg}^{G^{k,t}} t_h, \quad i \in G^{k,t}, \quad k \in \{1, 2, \dots, M\}. \quad (13)$$

Here, $\mathbf{v}_{avg}^{G^{k,t}}$ is the average group walking vector, computed as the mean of the walking vectors of the group members and, M is the total number of potentially freezing groups.

Definition V.2 (Group PFZ) The region in the vicinity of a group where the robot has a high probability of freezing. Instead of constructing the single PFZ as the convex hull of all potentially freezing pedestrians (like in [6]), we construct a PFZ for each potentially freezing group (see Fig. 3) as

$$PFZ^k = CH(\hat{\mathbf{p}}_{pf}^{i,t+t_h}), \quad i \in G^{k,t}, \quad k \in \{1, 2, \dots, M\}. \quad (14)$$

Every potentially freezing group's PFZ is computed for a future time instant $t + t_h$.

1) *Computing Deviation Angle*: If the robot's current trajectory navigates it into any of the group PFZs (implying an occurrence of FRP after time t_h), a deviation angle ϕ_{com}

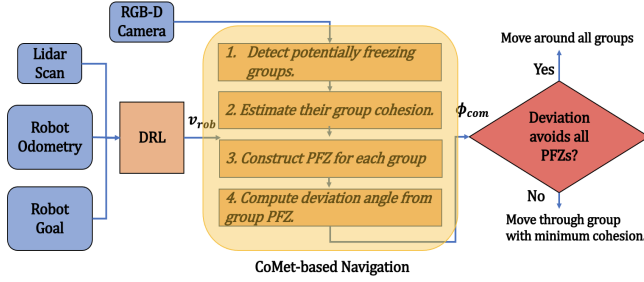


Fig. 4: The architecture of our socially-compliant navigation algorithm. Our formulation results in lower occurrence of freezing, exhibits lower deviations for the robot relative to pedestrians and groups, and avoids high-cohesion groups by moving around them.

to avoid it is computed. The robot’s current velocity \mathbf{v}_{rob} is deviated by ϕ_{com} using a rotation matrix about the Z-axis as,

$$\mathbf{v}'_{rob} = R_{z, \phi_{com}} \cdot \mathbf{v}_{rob}, \quad (15)$$

$$\phi_{com} = \underset{\mathbf{v}'_{rob} \cdot t_h \notin PFZ^k}{\operatorname{argmin}} (dist(\mathbf{v}'_{rob} \cdot t_h, \mathbf{g}_{rob})). \quad (16)$$

This equation implies that our navigation method deviates the robot by the least amount from its goal such that it does not enter any group’s PFZ. However, in dense scenarios, when the robot encounters many potentially freezing groups and their corresponding PFZs, Equation 16 may not be able to compute a feasible solution for ϕ_{com} . In such cases, a potential solution is to let the robot pass through the PFZ of a low cohesion group (see Fig. 1 left). In such cases, we formulate the deviation angle as,

$$\phi_{com} = \underset{\mathbf{v}'_{rob} \cdot t_h \in \mathcal{P}}{\operatorname{argmin}} (dist(\mathbf{v}'_{rob} \cdot t_h, \mathbf{g}_{rob})), \quad (17)$$

$$\mathcal{P} = PFZ^{min} \setminus (PFZ^{min} \cap PFZ^k) \forall k \in \{1, 2, \dots, M\}, \quad (18)$$

where PFZ^{min} is the PFZ of the group with the minimum cohesion in the scene. Since the permeability of low cohesion groups is high, the above formulation also lowers the probability of freezing.

Proposition V.1. *The deviation angles computed by CoMet-based navigation (Equations 16 or 18), and Frozone (2) satisfy the relationship $\phi_{froz} \geq \phi_{com}$.*

We refer the reader to [16] for a detailed proof for this proposition. This bound also guarantees that our new navigation algorithm generates trajectories that are more natural than Frozone [6]. We integrate our CoMet-based navigation method with a DRL-based navigation scheme to evaluate it in real-world scenarios [7] and DWA [38] to evaluate it in simulated environments. Figure 4) shows the components of our navigation algorithm used to compute trajectories that are more natural in real-world scenes.

VI. IMPLEMENTATION AND RESULTS

In this section, we describe the implementation for computing CoMet and CoMet-based navigation. We then evaluate CoMet in different standard pedestrian datasets that are annotated with perceived group cohesion levels. We highlight the performance of our navigation algorithm and show the benefits over prior methods in terms of the following quantitative metrics: freezing rate, deviation angle, and normalized path-length. We also qualitatively compare trajectories with an increased number of obstacles in the environment.

A. Implementation

In order to evaluate CoMet, we annotate groups in pedestrian datasets such as MOT, KITTI, ETH, etc. as low-, medium- and high-cohesion groups with the help of multiple human annotators. These annotated datasets are used as the ground truth since they reflect how humans perceive the groups in the videos. We choose these datasets because they depict groups in real-world scenarios with various lighting conditions, crowd densities and occlusions. We use a depth estimation method [39] with RGB images in the datasets to localize different pedestrians in the scene. We manually tune the weighing constants in the CoMet formulation (K_p, K_w, K_s, K_i) and set thresholds on the cohesion score to classify groups into the aforementioned categories based on the annotations in one of the datasets. These constants need not be tuned for different scenarios. We evaluate CoMet’s precision and recall in the groups in all other datasets. Since there are no prior methods that compute cohesion, our evaluation is only against the human annotations.

We have evaluated CoMet-based navigation using simulations created in MATLAB: (i) with tens to hundreds of groups, each with two to five members (Fig. 5) and a pre-assigned cohesion metric for each group and (ii) in corridor-like constrained scenarios with tens of pedestrians. The groups move linearly to a goal, and the simulated robot must take full responsibility to avoid collisions with them. We also evaluate CoMet-based navigation on a real Turtlebot 2 robot mounted with an Intel RealSense RGB-D camera (for pedestrian tracking and localization) and a 2-D Hokuyo lidar (used by the DRL [7] method).

B. Analysis

CoMet Classification: Table II highlights CoMet’s classification precision and recall in multiple datasets. CoMet’s parameters were tuned to improve its accuracy corresponding to detecting low-cohesion groups, which is required for navigation. This is reflected in the high precision and recall values in the second column. CoMet observes pedestrians in these datasets for ~ 5 seconds. During this period, it is able to update its initial classification as pedestrians’ trajectories change. For instance, a group initially perceived as high-cohesion may have its members move apart and is thereby classified as a low-cohesion group. Moreover, it is easier to detect features corresponding to proximity and walking speed than interactions between the pedestrians. This sometimes results in CoMet misclassifying medium- and high-cohesion groups in certain datasets. An interesting observation is that human annotators tend to classify groups in extremes, i.e. as either high-cohesion or low-cohesion groups. This leads to a low number of data points for medium-cohesion groups.

Socially-Compliant Navigation: We first qualitatively compare the trajectories (Fig. 5) of DWA [38] (in red), a DWA-Frozone hybrid [6] (in blue), and DWA and CoMet-based navigation (in green) in simulated environments with non-cooperative walking groups of obstacles. We observe that for the same set of dynamic obstacles, our approach computes smaller deviations while also avoiding collisions. DWA and the DWA-Frozone hybrids lead to conservative, and highly sub-optimal deviations from the goal direction. Although the exact set of obstacles each robot faces could be different depending on the trajectories they take, we observe that CoMet-based navigation’s deviations at any

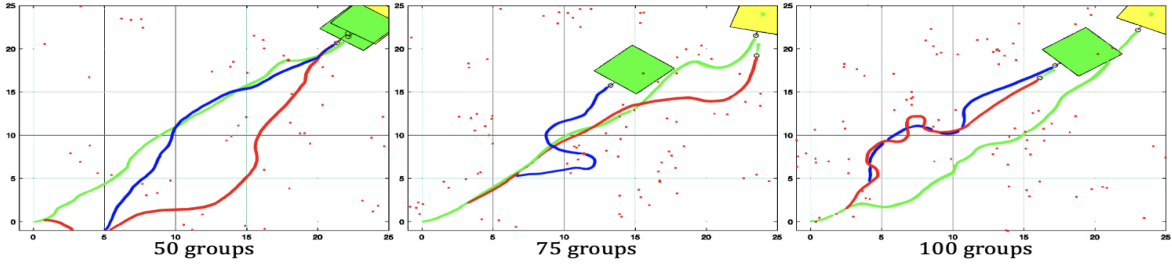


Fig. 5: Qualitative evaluations of the trajectories generated using CoMet-based navigation (green), Frozone [6](blue) and DWA [38] (red) in scenarios with 50, 75 and 100 moving groups (red dots) that are non-cooperative for collision avoidance. The green and yellow squares represent the sensing regions of Frozone and CoMet-based navigation respectively. We observe that our formulation leads to lower deviation angles and more natural trajectories even in dense environments with tens to hundreds of obstacles. Both Frozone and DWA lead to unnatural trajectories with large deviations as the number of obstacles in the environment increases.

Dataset Video	Low-Cohesion	Medium-Cohesion	High-Cohesion
ADL-Rundle	1.00/1.00	-	1.00/0.75
AVG-TownCenter	1.00/1.00	0.50/0.67	0.50/0.33
ETH Jelmolli	1.00/1.00	-	1.00/0.80
ETH Bahnhof	1.00/1.00	1.00/0.67	0.80/1.00
KITTI-16	1.00/1.00	1.00/0.50	0.33/1.00
KITTI-17	1.00/1.00	-	1.00/0.667
MOT17-08	1.00/1.00	-	1.00/0.50
MOT17-11	1.00/1.00	0.75/1.00	1.00/0.833
MOT20-02	1.00/1.00	0.34/0.33	0.50/0.33
TUD	1.00/0.857	-	1.00/1.00
Venice	0.875/0.875	-	0.938/0.60

TABLE II: This table shows the precision (first value in each column) and recall (second) values for the three classes when CoMet observes groups in the video for ~ 5 seconds. We observe good performance for high-, and low-cohesion groups. The original datasets have fewer occurrences of medium-cohesion groups, which impacts our precision.

instant never exceed the deviations of the other two methods in comparison. These results reinforce proposition V.1.

We also quantitatively compare the aforementioned three methods in environments with varying numbers (10 – 50) of pedestrians in a corridor-like scenario (which constricts the free space available to the robot). Pedestrians are given random initial locations and velocities, based on which they are classified into groups and PFZ_{froz} and group $PFZs$ (PFZ^k) are computed. The robot needs to navigate through the pedestrians to reach its goal. We use the following metrics: (1) *Average deviation angle* measured relative to the line connecting the start and goal locations; (2) *Freezing Rate* measured as the number of times the robot halted/froze over the total number of trials; and (3) *Normalized Path Length* measured as the robot’s path length over the length of the line connecting the start and goal locations.

CoMet-based navigation results in lower values with respect to all these metrics, as compared to DWA and the DWA-Frozone hybrid on the same scenarios. As the crowd size, density, or number of groups increase, Frozone’s conservative formulation makes the robot freeze at a high rate. This is because Frozone forbids the robot from avoiding a robot from in-front of the person, to improve pedestrian-friendliness in low- to medium-density scenes. On the other hand, we observe that CoMet-based navigation produces trajectories with high social compliance and naturalness by reducing the occurrence of freezing.

Real-world Evaluations: We qualitatively compare CoMet-based navigation’s trajectories with Frozone’s and a DRL algorithm’s [7] trajectories in narrow corridors with few walking pedestrians to constrain the available free space (and obey social distancing rules during the COVID-19 pan-

Metrics	Method	10 peds	20 peds	30 peds	40 peds	50 peds
Avg. Deviation Angle (lower better)	DWA	43.80	45.25	41.78	43.64	49.88
	DWA + Frozone	46.41	46.67	42.85	50.80	53.22
	DWA + Our Method	41.07	44.31	41.11	37.47	34.19
Freezing Rate (lower better)	DWA	0	0.37	0.39	0.36	0.44
	DWA + Frozone	0	0.29	0.25	0.43	0.57
	DWA + Our Method	0	0	0	0	0
Normalized Path Length (lower better)	DWA	1.33	1.31	1.42	1.47	1.49
	DWA + Frozone	1.46	1.45	1.55	1.59	1.51
	DWA + Our Method	1.12	1.29	1.27	1.35	1.35

TABLE III: We compare different navigation methods (DWA [38], DWA-Frozone hybrid [6], and DWA-CoMet-based navigation algorithm) on three metrics. We observe that our method consistently results in lower values corresponding to all these metrics, signifying improved naturalness of the trajectories.

demic). We highlight the differences in Fig. 1. Our approach is able to identify low-cohesion groups successfully and navigate through them without interfering with high-cohesion groups. In contrast, Frozone halts the robot completely, since it does not assume pedestrian cooperation in its formulation. The DRL method prioritizes reaching the goal with the minimum path length and therefore navigates through groups regardless of their cohesion, generating obtrusive trajectories. For more real-world qualitative analysis, please refer to [16].

CoMet-based navigation implicitly assumes a certain distance between the robot and a group to sense features and compute the group’s cohesion. This implies that the robot has sufficient free space to navigate without requiring the group members to modify their trajectories. That is, a group’s behavior is not influenced by the robot’s presence until the scenario is extremely dense.

Run-time Cost: The group cohesion computation’s run-time performance mainly depend on the pair-wise operations in the formulation since all the software packages used are highly optimized. However, since the maximum number of pedestrians detected within the robot’s sensing region is low (~ 7 people), the overall method is tractable.

VII. CONCLUSIONS, LIMITATIONS AND FUTURE WORK

We present a novel method to compute the cohesion of a group of people in a crowd using visual features. We use our cohesion metric to design a novel robot navigation algorithm that results in socially-compliant trajectories. We highlight the benefits over previous algorithms in terms of the following metrics: reduced freezing, deviation angles, and path lengths. We test our cohesion metric in annotated datasets and observe a high precision and recall.

Our method has some limitations. We model cohesion as a linear combination of the features. In addition, there are other characteristics used to estimate cohesion, including age, gender, environmental context, cultural factors, etc. that

we do not consider. Our approach also depends on the accuracy of how these features are detected, which may be affected due to lighting conditions and occlusions. Our navigation assumes that under normal levels of density of the scene, groups surrounding the robot are not influenced by its presence. This needs to be studied from a human-robot interaction perspective. It also assumes that different groups exhibit varying levels of cohesion, which may not hold all the time. As part of future work, we hope to address these limitations and evaluate our approach in more crowded real-world scenes.

REFERENCES

- [1] P. Trautman and A. Krause, "Unfreezing the robot: Navigation in dense, interacting crowds," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2010, pp. 797–803.
- [2] L. He, J. Pan, S. Narang, and D. Manocha, "Dynamic group behaviors for interactive crowd simulation," in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Eurographics Association, 2016, p. 139–147.
- [3] S. Bandini, A. Gorrini, L. Manenti, and G. Vizzari, "Crowd and pedestrian dynamics: Empirical investigation and simulation," in *Proceedings of Measuring Behavior*, 08 2012.
- [4] Y. F. Chen, M. Everett, M. Liu, and J. How, "Socially aware motion planning with deep reinforcement learning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 09 2017, pp. 1343–1350.
- [5] L. Tai, J. Zhang, M. Liu, and W. Burgard, "Socially compliant navigation through raw depth inputs with generative adversarial imitation learning," in *ICRA*, May 2018, pp. 1111–1117.
- [6] A. J. Sathyamoorthy, U. Patel, T. Guan, and D. Manocha, "Frozone: Freezing-free, pedestrian-friendly navigation in human crowds," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4352–4359, 2020.
- [7] P. Long, T. Fan, X. Liao, W. Liu, H. Zhang, and J. Pan, "Towards Optimally Decentralized Multi-Robot Collision Avoidance via Deep Reinforcement Learning," *arXiv e-prints*, p. arXiv:1709.10082, Sep 2017.
- [8] Y. F. Chen, M. Liu, M. Everett, and J. P. How, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *ICRA*. IEEE, 2017, pp. 285–292.
- [9] A. J. Sathyamoorthy, J. Liang, U. Patel, T. Guan, R. Chandra, and D. Manocha, "Densecavoid: Real-time navigation in dense crowds using anticipatory behaviors," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 11 345–11 352.
- [10] N. Ashton, M. E. Shaw, and A. P. Worsham, "Affective reactions to interpersonal distances by friends and strangers," *Bulletin of the psychonomic society*, vol. 15, pp. 306–308, 1980.
- [11] S. Karau and K. Williams, "The effects of group cohesiveness on social loafing and social compensation," *Group Dynamics: Theory, Research, and Practice*, vol. 1, pp. 156–168, 1997.
- [12] E. Knowles, "Boundaries around group interaction: the effect of group size and member status on boundary permeability," *Journal of personality and social psychology*, vol. 26, no. 3, p. 327–331, June 1973. [Online]. Available: <https://doi.org/10.1037/h0034464>
- [13] E. T. Hall, R. L. Birdwhistell, B. Bock, P. Bohannon, A. R. Diebold, M. Durbin, M. S. Edmonson, J. L. Fischer, D. Hymes, S. T. Kimball, W. La Barre, J. E. McClellan, D. S. Marshall, G. B. Milner, H. B. Sarles, G. L. Trager, and A. P. Vayda, "Proxemics [and comments and replies]," *Current Anthropology*, vol. 9, no. 2/3, pp. 83–108, 1968. [Online]. Available: <https://doi.org/10.1086/200975>
- [14] J. Wagnild and C. M. Wall-Scheffler, "Energetic consequences of human sociality: Walking speed choices among friendly dyads," *PLOS ONE*, vol. 8, no. 10, pp. 1–6, 10 2013.
- [15] J. Shao, C. C. Loy, and X. Wang, "Scene-independent group profiling in crowd," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2227–2234.
- [16] A. Jagan Sathyamoorthy, U. Patel, M. Paul, N. K. Sanjeev Kumar, Y. Savle, and D. Manocha, "CoMet: Modeling Group Cohesion for Socially Compliant Robot Navigation in Crowded Scenes," *arXiv e-prints*, p. arXiv:2108.09848, Aug. 2021.
- [17] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, 2018. [Online]. Available: <http://arxiv.org/abs/1804.02767>
- [18] R. Chandra, U. Bhattacharya, A. Bera, and D. Manocha, "Densepeds: Pedestrian tracking in dense crowds using front-rvo and sparse features," *arXiv preprint arXiv:1906.10313*, 2019.
- [19] W. Choi, Y.-W. Chao, C. Pantofaru, and S. Savarese, "Discovering groups of people in images," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 417–433.
- [20] A. K. Chandran, L. A. Poh, and P. Vadakkepat, "Identifying social groups in pedestrian crowd videos," in *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, 2015, pp. 1–6.
- [21] R. M. Favaretto, L. L. Dihl, and S. R. Musse, "Detecting crowd features in video sequences," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2016, pp. 201–208.
- [22] Z. Ebrahimpour, W. Wan, O. Cervantes, T. Luo, and H. Ullah, "Comparison of main approaches for extracting behavior features from crowd flow analysis," *ISPRS International Journal of Geo-Information*, vol. 8, no. 10, 2019. [Online]. Available: <https://www.mdpi.com/2220-9964/8/10/440>
- [23] M. Yang, L. Rashidi, A. S. Rao, S. Rajasegarar, M. Ganji, M. Palaniswami, and C. Leckie, "Cluster-based crowd movement behavior detection," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, 2018, pp. 1–8.
- [24] M. Cristani, L. Bazzani, G. Paggetti, A. Fossati, D. Tosato, A. Del Bue, G. Menegaz, and V. Murino, "Social interaction discovery by statistical analysis of f-formations," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2011, pp. 23.1–23.12, <http://dx.doi.org/10.5244/C.25.23>.
- [25] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey," *Robotics Auton. Syst.*, vol. 61, pp. 1726–1743, 2013.
- [26] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, "Activity forecasting," in *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 201–214.
- [27] M. Pfeiffer, U. Schwesinger, H. Sommer, E. Galceran, and R. Siegwart, "Predicting actions to act predictably: Cooperative partial motion planning with maximum entropy models," 10 2016, pp. 2096–2101.
- [28] H. Kretschmar, 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. [Online]. Available: <https://doi.org/10.1177/0278364915619772>
- [29] A. Bera, T. Randhavane, R. Prinja, and D. Manocha, "Sociosense: Robot navigation amongst pedestrians with social and psychological constraints," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 7018–7025.
- [30] T. Randhavane, A. Bera, E. Kubin, A. Wang, K. Gray, and D. Manocha, "Pedestrian dominance modeling for socially-aware robot navigation," in *2019 International Conference on Robotics and Automation (ICRA)*, 2018.
- [31] A. Bera, T. Randhavane, A. Wang, D. Manocha, E. Kubin, and K. Gray, "Classifying group emotions for socially-aware autonomous vehicle navigation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.
- [32] E. Goffman, *Encounters: Two studies in the sociology of interaction*. Ravenio Books, 1961.
- [33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *arXiv e-prints*, p. arXiv:1506.02640, June 2015.
- [34] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Real-time Tracking with a Deep Association Metric," *arXiv e-prints*, p. arXiv:1703.07402, Mar. 2017.
- [35] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
- [36] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 2018, pp. 59–66.
- [37] J. Van Den Berg, S. Guy, M. Lin, and D. Manocha, "Reciprocal n-body collision avoidance," in *Robotics Research - The 14th International Symposium ISRR*, ser. Springer Tracts in Advanced Robotics, no. STAR, June 2011, pp. 3–19.
- [38] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics Automation Magazine*, vol. 4, no. 1, pp. 23–33, March 1997.
- [39] K. Harsányi, A. Kiss, A. Majdik, and T. Sziranyi, "A hybrid cnn approach for single image depth estimation: A case study," in *Multimedia and Network Information Systems*, K. Choroś, M. Kopel, E. Kukla, and A. Siemiński, Eds. Cham: Springer International Publishing, 2019, pp. 372–381.