Efficient Context-Aware Model Predictive Control for Human-Aware Navigation

Elisa Stefanini, *Member, IEEE*, Luigi Palmieri, *Member, IEEE*, Andrey Rudenko, *Member, IEEE*, Till Hielscher, *Graduate Student Member, IEEE*, Timm Linder, *Member, IEEE*, and Lucia Pallottino, *Senior Member, IEEE*

***Abstract*—With the goal of creating efficient human-aware robot navigation systems, we present a Context-aware Model Predictive Control (MPC) formulation designed specifically for dynamic and crowded environments. State-of-the-art approaches use mainly geometric information and predictions of human motion, thus being proactive about human positions and intents but less aware of high-level behaviors encoded in contextual cues of human activities. In contrast to that, we propose a holistic planning solution that considers additional contextual information such as 3D human body poses with their velocities and recognized activities. We carefully design the MPC formulation to facilitate the integration of modern perception systems and the usage of fast solvers for embedded robot motion optimization. Compared to a set of baselines, our proposed system not only ensures safety but also significantly improves task and computational efficiency. Through extensive simulations and real-life experiments, our planner demonstrates reliable operation in terms of smooth and efficient navigation in human-populated areas.**

***Index Terms*—Human-aware motion planning, motion and path planning, autonomous agents.**

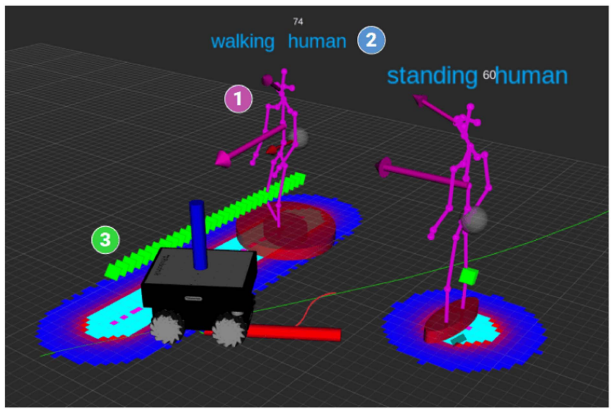


Fig. 1. Context-aware collision avoidance of a mobile robot, considering (1) full-body 3D human skeleton poses projected as red ellipses, (2) detected activities, and (3) 2D motion predictions. The mobile robot proactively clears the path of a walking human, while preparing to bypass a standing person.

I. Introduction

IN DYNAMIC and densely populated environments, robots need to recognize and adapt to human presence and activities, prioritizing safety, legibility, and social norms in their navigation systems [1], [2], [3]. Understanding human motion and anticipating their future actions is essential for improving human-aware robot navigation [4]. In turn, advanced perception systems are developed for robots to detect, track, predict, and recognize human intentions, enabling them to anticipate human actions for harmonious coexistence [5].

Model Predictive Control (MPC) is a key technology for human-aware planning, allowing robots to anticipate future events and adjust paths while considering multiple constraints [6], [7], and multi-modal 2D human motion prediction [4].

Historically, research often focused on 2D trajectory predictions, but recent studies emphasize the importance of integrating 3D data and contextual cues for improved robot navigation in human-shared spaces [8], [9].

This work explores the complexities and evolving dynamics of human-robot interaction in such environments. We present a novel Context-aware MPC approach formulation, which integrates not only geometric information about human positions but also additional cues to enhance robot navigation in densely populated areas. Starting with the premise that incorporating human motion prediction into the planner results in efficient and smooth navigation with active and reliable collision avoidance [4], [10], we propose an MPC formulation that considers additional contexts such as human activities and 3D human body poses provided by state-of-the-art 3D human perception systems. The proposed formulation allows us to use state-of-the-art fast solvers for embedded optimization [11], thus achieving

Received 8 May 2024; accepted 29 August 2024. Date of publication 13 September 2024; date of current version 24 September 2024. This article was recommended for publication by Associate Editor Pamela Carreno and Editor Gentiane Venture upon evaluation of the reviewers' comments. This work was supported in part by the EU Horizon 2020 research and innovation program (DARKO) under Grant 101017274 in part by the Next Generation EU project under Grant ECS00000017 in part by Ecosistema dell'Innovazione' TuscanyHealth Ecosystem (THE, PNRR, Spoke 4: Spoke 9: Robotics and Automation for Health) and in part by the Italian Ministry of Education and Research (MIUR) in the framework of FoReLab project (Departments of Excellence). *(Corresponding author: Luigi Palmieri.)*

Elisa Stefanini is with the SoftBots, Fondazione Istituto Italiano di Tecnologia, 16163 Genova, Italy, and also with the Centro di Ricerca "E. Piaggio", Dipartimento di Ingegneria dell'Informazione, Universitá di Pisa, 56122 Pisa, Italy (e-mail: elisa.stefanini@phd.unipi.it).

Luigi Palmieri, Andrey Rudenko, and Timm Linder are with the Robert Bosch GmbH, Corporate Research, 71272 Stuttgart, Germany (e-mail: Luigi.Palmieri@de.bosch.com; timm.linder@de.bosch.com; luigi.palmieri@de.bosch.com).

Till Hielscher is with the Socially Intelligent Robotics Lab, Institute for Artificial Intelligence, University of Stuttgart, 70569 Stuttgart, Germany (e-mail: till.hielscher@ki.uni-stuttgart.de).

Lucia Pallottino is with the Centro di Ricerca "E. Piaggio", Dipartimento di Ingegneria dell'Informazione, Universitá di Pisa, 56122 Pisa, Italy (e-mail: lucia.pallottino@unipi.it).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2024.3461552, provided by the authors.

Digital Object Identifier 10.1109/LRA.2024.3461552

high computational efficiency even with high dimensional social cues.

The proposed system is rigorously tested against a set of baselines in various human avoidance scenarios in crowded areas. Results of our evaluation show the effectiveness of our planner in ensuring safe, efficient and human-aware robot motion: our approach improves significantly against the state of the art in terms of task and computational efficiency while keeping a comfortable distance from humans. Finally, we deploy the planner on the DARKO[1] robot platform and demonstrate reliable operation in terms of smooth and efficient navigation across several real-world experiments.

II. Related Work

Traditional MPC approaches achieve collision avoidance by considering online human motion detection [12], [13], incorporating collision avoidance constraints, considering both robot-forbidden areas and social proximity rules for human comfort [14], or by integrating human motion prediction into the MPC optimization objective [15]. Deep Reinforcement Learning (DRL) methods such as GA3C-CADRL or SA-CADRL are also employed for navigation in dynamic settings [16], [17], but they may not comprehensively tackle the complexities of robot and environment dynamics. We show in our previous work [4] that DRL-based approaches are outperformed by model-based techniques.

Predicting human motion is crucial for ensuring smooth navigation and safety [1]. Predictive planning approaches allow robots to make informed real-time decisions by adjusting trajectory and behavior based on predicted human motion [4], [7], [10], [18]. In our previous work HuMAN-MPC [10], we use fast-embedded optimization methods with 2D human motion prediction to achieve superior computational efficiency. However, a significant shortcoming in the field is the limited consideration of 3D human data, which is crucial for accurately capturing complex human dynamics. Existing research often relies on 2D projections even when utilizing 3D sensors, potentially compromising a natural and safe interaction [13].

Advancements in 3D perception technologies, including pose estimation and activity recognition, are necessary for interpreting human intentions. Human pose estimation provides detailed body structure insights by predicting skeletons with accurate body joint locations from sensor input data. A vast amount of methods exists [19], for instance, based on CNNs [20] or Transformers [21]. In [22], authors introduced DLow, a model that uses diversifying latent flows for more varied and realistic human motion predictions, while [23] proposed DMMGAN, an attention-based generative adversarial network that predicts diverse multi-modal trajectories for 3D human joints. Most relevant to our research are single-view approaches that output skeletons in absolute, metric-scale 3D coordinates and can run in real time, e.g. [24]. They are often combined with a human detector such as [25] in a top-down fashion. The outputs of the detector and pose estimator can further be temporally associated and filtered with a temporal tracking module [26], [27], e. g. based on Kalman filters, RNNs, or Transformers. Finally, human activity recognition can enhance decision-making during navigation and interaction. It can be performed directly on raw sensor data, such as images [28], or on lower-dimensional skeletal features [19]. We opt for the 3D skeleton-based approaches, which require less training data due to the abstraction provided by human pose estimation.

However, the integration of 3D perception methods into MPC remains limited despite the benefits they offer in terms of enhancing interaction and safety in robotic navigation. Our contribution addresses this gap by introducing a context-aware MPC that integrates comprehensive 3D and contextual human cues, significantly enhancing the robot's navigation capabilities in populated environments.

III. Method

In this section, we describe the preliminaries to our method in Section III-A, processing and representation of the 3D human motion data in Section III-B, human activities in Section III-C, the novel MPC formulation in Section III-D and the system integration aspects in Section III-E.

A. Preliminaries

Our context-aware MPC method depends on several 3D perception components to infer and predict the contextual cues of nearby people, and on third-party components to solve the optimal control problem.

In particular, we assume access to the 2D positions $\mathbf{h}_i(t)$ for each human $i \in [1, N_h]$ in the environment at time t up to the prediction horizon T_p. Predictions are represented as discrete trajectories $\mathbf{T}_{i,1:T_p} = \{\mathbf{h}_i(1), \mathbf{h}_i(2), \ldots, \mathbf{h}_i(T_p)\}$. Moreover, we assume access to the 2D linear velocity estimations $[v_{h_{x_i}}, v_{h_{y_i}}]$, 3D skeleton positions of all N_s joints $s_{i,j}(t) = [x_{i,j}, y_{i,j}, z_{i,j}]$ for each human $i \in [1, N_h]$ and joint $j \in [1, N_s]$ at time t, along with the associated activity labels, such as walking, standing, and sitting.

Finally, in our method, we employ the acados framework [11]. In particular, we utilize the SQP Real-Time Iteration (RTI) scheme [29] and the HPIPM solver with partial condensing [30], which addresses nonlinearly constrained optimization problems by transforming them into a series of quadratic problems.

B. 3D Human Skeletons Poses and Velocity

Inspired by the recent approaches to using ellipses for obstacle representation [31], [32], we construct ellipses around 3D skeleton joints to encapsulate the human pose and movement. These dynamically adjusted ellipses maintain an accurate 3D representation in 2D space and are enlarged proactively based on human velocity for safety, giving the robot more time to avoid collisions, particularly with fast-moving humans. To avoid introducing non-linear distance constraints in MPC due to the inherent non-linearity of ellipse equations, we select the point h_e^* on the ellipse's perimeter closest to the robot's pose. This ensures linear distance constraints while adapting to changes in the ellipse's size and orientation increasing minimum human-robot distance during navigation. For computational efficiency, we opt for the deterministic formulation of the constraint.

For each human $i \in [1, N_h]$, given the N_s 3D skeleton joint coordinates $s_{i,j} = [x_{i,j}, y_{i,j}, z_{i,j}], j \in [1, N_s]$ (Fig. 2(1)), the minimum 2D ellipse is computed using the CGAL Library [33] based on the projection of the 3D joint Cartesian coordinates onto the XY plane ($z_{i,j} = 0$), as depicted in Fig. 2(2)). This minimum ellipse is defined as the unique ellipse with the smallest

[1]EU Project DARKO, https://darko-project.eu/

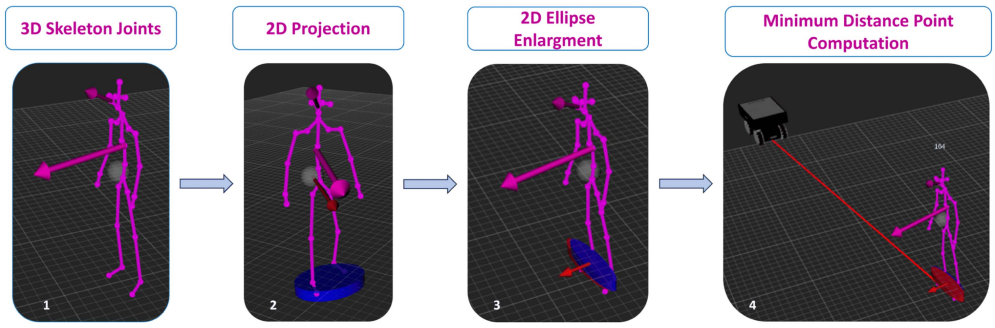


Fig. 2. Given the 3D human joints coordinates (1), the ellipse is computed by projecting the coordinates onto the XY plane (2). Then, the ellipse is enlarged based on the estimated human velocity (3). Finally, the ellipse point closest to the robot is computed (4).

area capable of enclosing a finite set of points in two-dimensional Euclidean space. Formally, the resulting ellipse satisfies the condition that all points $s_{i,j} = [x_{i,j}, y_{i,j}, 0]; j \in [1, N_s], i \in [1, N_h]$ meet:

$$ax_{i,j}^2 + bx_{i,j}y_{i,j} + cy_{i,j}^2 + dx_{i,j} + ey_{i,j} + f \leq 0 \quad (1)$$

where a, b, c, d, e, f are the coefficients of the ellipse. From these coefficients, we derive the canonical ellipse parameters for each human i, i.e. its center coordinates c_{x_i}, c_{y_i}, orientation c_{θ_i}, major and minor semi-axes a_i, b_i, achieved through the general ellipse geometric formulation [34].

Once computed, the next step involves enlarging the ellipse based on estimated human velocities $v_{h_{x_i}}$ and $v_{h_{y_i}}$ as shown in Fig. 2(3). Given the linear human velocity $v_{h_i} = \sqrt{v_{h_{x_i}}^2 + v_{h_{y_i}}^2}$, the enlargement e_i of both semi-axes $a_i = a_i + e_i$ and $b_i = b_i + e_i$ is determined using the hyperbolic tangent (tanh) function $e_i = |\tanh(1.5v_{h_i})|$.

This enlarges the ellipse while keeping the orientation constant and maintaining the ratio between major and minor axes. As human velocity increases, the ellipse enlarges significantly to enhance safety, enabling effective human avoidance by the robot. However, to prevent overly restrictive planning constraints, the enlargement is limited to one meter when human velocity exceeds 1.8 m/s [35], ensuring the planner can still find feasible solutions even in dynamic scenes.

Newton's method [36] is employed to compute the point $h_{e_i}^*$. This technique optimizes the squared distance D^2 between a point h_{e_i} on the ellipse's perimeter and the robot's pose $\mathbf{p_r}$, as shown in Fig. 2(4):

$$D^2 = (h_{e_i} - \mathbf{p_r}) \cdot (h_{e_i} - \mathbf{p_r}) \quad (2)$$

Upon convergence, the ellipse point $h_{e_i}^*$ closest to the robot, is identified. This point, determined for each human i, is then integrated as a constraint in the MPC (see Section III-D).

C. Human Activities

Based on detected human activities, we dynamically adjust the robot's speed accordingly, i.e. slowing down in response to more dynamic movements for safety and predictability in interactions. This adjustment is crucial not only in dynamic scenarios but also around static activities like standing or sitting to ensure comfortable and safe interactions [37], [38].

We assume that a list of human activities, $Activity = \{\text{walking}, \text{standing}, \text{sitting}\}$, is provided, each mapped to an activity factor α using a mapping function h_α. An aggregate factor $\mathcal{A}$ is computed for all $N_p < N_h$ nearby humans within $4\,\text{m}$ radius, as speed control is only necessary when humans are close (otherwise the cost term is not included):

$$\mathcal{A} = \frac{1}{N_p} \sum_{i=1}^{N_p} h_{\alpha_i}(Activity_i), \quad (3)$$

The choice to use the average of activity factors was made to avoid creating a bias towards a single activity, ensuring a balanced response to the overall environment.

Utilizing the socially acceptable maximum velocity ($v_{sm} = 0.5$ m/s) [37], the desired robot velocity (v_{des}) is calculated based on $\mathcal{A}$, subsequently used in the formulation of a new cost functional term (see Section III-D):

$$v_{des} = v_{sm} \cdot \mathcal{A} \quad (4)$$

Table I lists activity factors and corresponding robot velocities for single-human scenarios, determined via (4) after informal validation. For other activities, the system defaults to the walking scenario to determine the desired velocity.

TABLE I
ACTIVITY FACTOR α AND ROBOT DESIRED VELOCITY v_{des} VALUES ACCORDING TO THE HUMAN ACTIVITY

Activity	Activity Factor α	v_{des}
Walking	0.4	0.2
Standing	0.6	0.3
Sitting	0.8	0.4
Unknown	0.4	0.2

D. Context-MPC Formulation

The motion planner is designed as a model predictive controller that addresses a discretized optimal control problem (OCP) in every iteration [11]. The OCP is defined as follows:

$$\arg\min_{x(\cdot),u(\cdot)} \; \mathcal{J}_T(\mathbf{x}(T_p), \mathbf{p}(T_p)) + \sum_{t=0}^{T_p-1} \mathcal{J}_S(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}(t)) \quad (5a)$$

$$\text{subject to} \quad \mathbf{x}(t) \in X \qquad t \in [0, T_p] \quad (5b)$$

$$\mathbf{u}(t) \in U \qquad t \in [0, T_p - 1] \quad (5c)$$

$$\mathbf{x}(t+1) = f(\mathbf{x}(t), \mathbf{u}(t)) \qquad t \in [0, T_p - 1] \quad (5d)$$

$$d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{h}_{e_i(0)}^*) \geqslant d_h \qquad i \in [1, N_h], t \in [0, T_p - 1] \quad (5e)$$

$$d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{o}_{s(0)}) \geqslant d_s \qquad t \in [0, T_p - 1] \quad (5f)$$

where $\mathbf{x}(t)$ and $\mathbf{u}(t)$ denote respectively the state and the control at a specific time step n on the horizon which is split into T_p shooting nodes, with X and U being the allowed state and control spaces. Equation (5d) represents the dynamic constraint of the system, wherein we employ a differential drive model to describe the robot's dynamics. The additional parameter $\mathbf{p}(t)$ includes the goal $\mathbf{g}(t)$, the desired trajectory, the predicted human position $\mathbf{h}_i(t)$ over the horizon, the computed point $\mathbf{h}^*_{e_i(0)}$ on the ellipse at the minimum distance from the robot's pose, the activity factor $\mathbf{h}_{\alpha_i(0)}$ for all N_h considered humans, and the position of the nearest static obstacle $\mathbf{o}_{s(0)}$. Moreover, $\mathbf{h}^*_{e_i(0)}$, $\mathbf{h}_{\alpha_i(0)}$, and $\mathbf{o}_{s(0)}$ are only given for the initial state of the robot and considered constant for the entire prediction horizon. The two distance functions, $d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{h}^*_{e_i(0)})$ and $d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{o}_{s(0)})$, calculate the Euclidean distance between the robot's position $\mathbf{p}_{\mathbf{r}(t)} \in \mathbf{x}(t)$ and the given positions $\mathbf{h}^*_{e_i(0)}, \mathbf{o}_{s(0)}$, where d_h represents the minimum safe distance from a human, and d_s denotes the general collision avoidance distance.

The objective function is split into terms for the stage cost $\mathcal{J}_S$ and the terminal cost $\mathcal{J}_T$. The cumulative stage cost is composed of four terms:

$$\mathcal{J}_S(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}(t)) = \mathcal{J}_g(\mathbf{x}(t), \mathbf{p}(t)) + \mathcal{J}_u(\mathbf{u}(t)) + \mathcal{J}_{vel}(\mathbf{x}(t), \mathbf{p}(t)) + \mathcal{J}_{col}(\mathbf{x}(n), \mathbf{p}(t)). \tag{6}$$

Penalizing the distance to the given goal is achieved with the goal cost term:

$$\mathcal{J}_g(\mathbf{x}(t), \mathbf{p}(t)) = ||\mathbf{x}(t) - \mathbf{p}(t)||^2_{W_g} \quad t \in [0, T_p] \tag{7}$$

weighted by the diagonal matrix W_g.

The control is penalized with the control cost term:

$$\mathcal{J}_u(\mathbf{u}(t)) = ||\mathbf{u}(t)||^2_{W_u} \quad t \in [0, T_p - 1] \tag{8}$$

weighted by the diagonal matrix W_u.

The new velocity cost term to control the robot velocity v_r is formulated as:

$$\mathcal{J}_{vel}(\mathbf{x}(t), \mathbf{p}(t)) = w_{vel}||v_r - v_{des}||^2 \tag{9}$$

weighted by a coefficient w_{vel} and using v_{des} as reference velocity.

The collision cost term $\mathcal{J}_{col}(\mathbf{x}(t), \mathbf{p}(t))$ is formulated by taking into account the motion prediction trajectories of the humans $h_i, i \in [1, N_h]$, i.e. $\mathbf{T}_{i,1:T_p}$:

$$\mathcal{J}_{col}(\mathbf{x}(t), \mathbf{p}(t)) = \sum_{i=1}^{N_h} f_h(d(\mathbf{p}_{\mathbf{x}(t)}, \mathbf{h}_i(t))) \tag{10}$$

where $d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{h}_i(t)$ is the Euclidean distance between the robot's pose and the human's predicted position $\mathbf{h}_i(t)$ and $f_h(d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{h}_i(t)))$ is a function defined as:

$$\begin{cases} \left(-\frac{\kappa q}{4}\right) d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{h}_i(t)) \\ + \left(\frac{q}{2} + \frac{\kappa q}{4} r_{\text{th}}\right) & \text{if } d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{h}_i(t)) \le r_{\text{th}} \\ \dfrac{q}{1 + e^{\kappa(d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{h}_i(t) - r_{\text{th}})}} & \text{if } d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{h}_i(t)) > r_{\text{th}} \end{cases} \tag{11}$$

where k and q are positive non-zero tuning parameters and r_{th} is a threshold distance where the returned cost is $\frac{q}{2}$. Moreover, r_{th} accounts for shifting the distance on which f_h is computed, q adjusts cost amplitude and k affects cost steepness at the threshold distance.

The collision cost term formulation, adapted from [39], where it is formulated as the equation in the $d(\mathbf{p}_{\mathbf{r}(t)}, \mathbf{h}_i(t)) > r_{\text{th}}$ condition in (11), was modified to meet acados requirements in our previous work [10], addressing the original term's lack of twice continuity differentiability and positive semi-definite Hessian, unsuitable for efficient SQP-RTI use due to potential unfeasibility issues from a negative definite Hessian in rapid embedded optimization [39].

The optimization includes several inequality constraints. Equation (5f) is used as a soft constraint for general collision avoidance, while a hard constraint, (5e), is added to ensure human safety by constraining the state space based on the distance between the robot and the current point h^*_e on the human-representing ellipse for nearby humans. Due to space constraints, we have omitted the details of the slack variable. For further information, please refer to [4], [10]. Moreover, we chose to integrate human motion predictions solely into the objective function to enhance overall performance, as observed in [4], where using predictions as a cost rather than a constraint led to smoother movements and decreased computational load. Finally, uncertainties in the threshold value of d_h are considered to compensate for potential inaccuracies in human perception and fluctuations in the robot's localization system, as suggested in [40].

In conclusion, the terminal cost for the final shooting node is:

$$\mathcal{J}_T(\mathbf{x}(T_p), \mathbf{u}(T_p), \mathbf{p}(T_p)) = \mathcal{J}_g(\mathbf{x}(T_p), \mathbf{p}(T_p)) + \mathcal{J}_{vel}(\mathbf{x}(T_p), \mathbf{p}(T_p)) + \mathcal{J}_{col}(\mathbf{x}(T_p), \mathbf{p}(T_p)). \tag{12}$$

E. System Integration

The context-aware navigation planning system integrates the MPC as local planner plugin within the ROS 1 navigation stack. A context information module connects the plugin with humans data in the environment, including 3D human poses, velocities, activities, and trajectory predictions.

3D Perception of Humans: In this work, we utilize a 3D perception system developed in the EU project DARKO for real-time 3D human detection, pose estimation, tracking, and trajectory prediction. To detect 3D human centroids, we use a TensorRT-accelerated variant of RGB-D YOLO [25]. Centroids are fed into a Kalman filter-based trajectory tracker [26], while associated 2D bounding boxes are passed into a top-down absolute 3D human pose estimator based upon volumetric heatmaps [24]. The resulting per-frame 3D skeletons are temporally associated and assigned to the same identities as used by the trajectory tracker. Both trajectories and temporally associated 3D skeletons are fed into a human activity classifier, implemented using an RBF kernel SVM trained on skeletal features (pairwise joint angles, angles between skeletal bones, and ground plane [41]). We use a one-vs-one approach for multi-class classification, and train on 2–3 short video sequences (1–3 minutes each) per activity using class weight balancing. Predicted activity classes are integrated over time via fixed-lag, mode-based filtering. Trajectory predictions for the optimization horizon are generated using a social force model. All components are integrated via ROS and with RViz for visualization.

IV. Experimental Validation

We validate our proposed approach in three key stages: assessing the robot's behavior (Section IV-B1), conducting statistical analysis (Section IV-B2) and performing real-world experiments (Section IV-B3).

Video of the behavior evaluation and real-world experiments are available in the multimedia material.

A. Setup

1) Baselines: Across several experiments we compare our Context-MPC against the HuMAN-MPC [10] and Timed Elastic Band (TEB) planner [42]. HuMAN-MPC implements a fast-embedded MPC formulation that uses 2D human motion predictions and Euclidean distance constraints.

2) Simulation: The simulation experiments were performed on a laptop with Intel Core i7-10850H 6x2.7 GHz, Nvidia Quadro RTX3000, 32 Gb RAM, and Ubuntu 20.04. The environment and the robot were simulated with Gazebo and the Robotnik XL-Steel platform[2].

Human motion in simulated experiments is realized through mixed reality, using ROS bag files to record the outputs of the 3D perception system running on the real DARKO robot. Trajectory predictions were generated during simulation using this real-world human data: i.e., in all simulation scenarios, people are unable to react to the robot since it is invisible to them, simulating uncooperative behavior.

3) Metrics: Assessing our proposed approach involves combining navigation efficiency and safety metrics. Efficiency is measured as the time the robot takes to reach its goal, denoted as t_g [s]. Safety metrics focus on human-related aspects, such as the minimum distance between the robot and any human during navigation (d_{min} [m]) and the average distance to the closest human (d_{avg} [m]).

In behavior evaluation, we present these metrics alongside the robot's velocity profiles, including both average values and standard deviations (σ) for statistical analysis.

4) Methods' Parameters and Robot Model: In the experiments, we use the same parameters for both Context-Aware and HuMAN-MPCs by setting the following values: prediction horizon 5.0 s, number of shooting nodes 50, stage goal cost weights [1, 1, 0, 250], terminal goal cost weights [40, 40, 20, 0], control cost weights [0, 0], velocity cost weight 400, collision cost: $q = 2.0, k = 5.0, v_{sm} = 0.5$ m/s, objects distance constraint $d_s = 0.5$ m, and human distance constraint $d_h = 0.5$ m. The model is a differential drive robot [43], whose state $\mathbf{x}$ is represented by 2D Cartesian positions, heading, and velocity; controls $\mathbf{u}$ are translational acceleration and angular velocity. For the Timed Elastic Band, we informally set the parameters to achieve the best possible behavior across scenarios, namely minimizing the amount of cost map inflation while still achieving reliable navigation toward the goal.

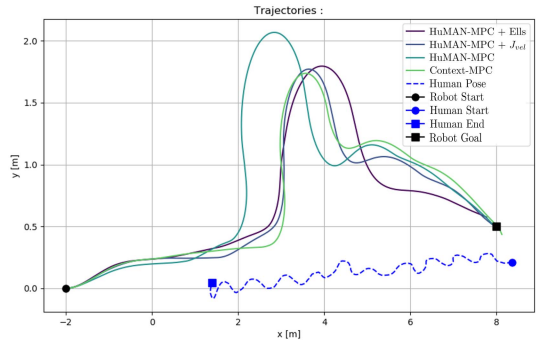
B. Experiments

We perform the following experiments:

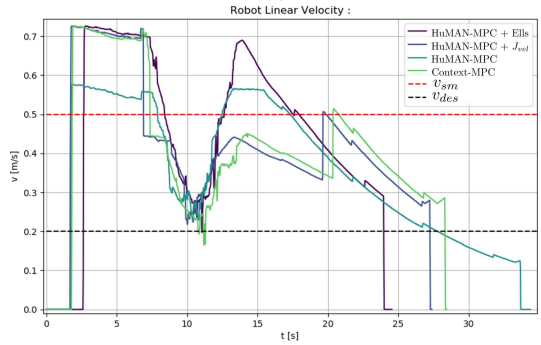
1) Behaviour Evaluation: To evaluate performance in basic encounters, we begin with an ablation study of our approach. We use three scenarios: the robot avoids a single human who is either sitting, standing, or walking, testing four variations of the Context-MPC:

1) HuMAN-MPC: MPC with 2D human motion predictions and Euclidean distance constraints.
2) HuMAN-MPC + Ells: Incorporates ellipses for distance constraints as detailed in Section III-B.
3) HuMAN-MPC + J_{vel}: HuMAN-MPC augmented with the velocity cost term based on human activities as described in Section III-C.

[2]Robotnik xl-steel simulator, https://github.com/RobotnikAutomation/summit_xl_sim



(a)



(b)

Fig. 3. Behavior Evaluation in the Walking Scenario: (a) Trajectories computed by each MPC variant, showing both the human and robot starting and ending poses; (b) Linear Robot Velocity profile for each variant together with v_{sm} and v_{des} for Walking activity.

4) Context-MPC: our approach that extends HuMAN-MPC by including J_{vel}, and Ells formulations.

Each variant undergoes testing in the same simulation setup, starting from the same initial poses and reaching the same goals. Due to space limitations, only the walking scenario is presented here, as similar results were obtained across scenarios. Multimedia materials showcase simulations of each MPC variant in each scenario.

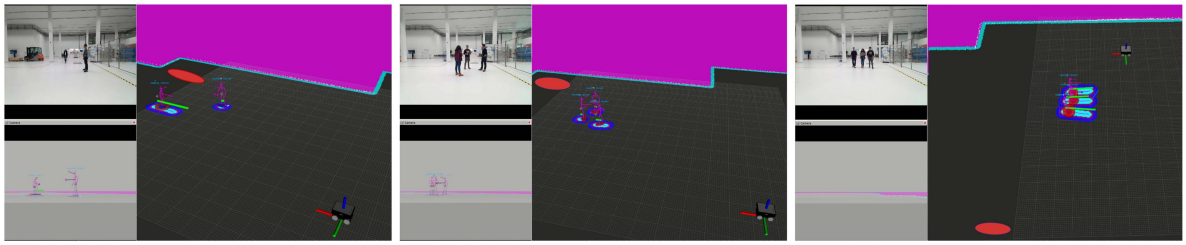
2) Statistical Validation: To validate our proposed planner, we conducted a statistical performance analysis in three complex scenarios with increased difficulty compared to Section IV-B1: two humans engaged in different activities (one walking, the other standing), three people standing, and three people walking. For each scenario and planner (HuMAN-MPC, Context-MPC, and TEB controller), we ran 50 simulations. The robot started from the same initial point in each simulation, but the goal changed. We used the same set of 50 distinct goals for all controllers in each scenario to ensure effective comparison. These goals were designed with a constant Y-coordinate, increasing the X -coordinate by 5 cm in each iteration.

3) Real World Experiments: We conducted real-world experiments on the DARKO robot platform at the research campus ARENA2036in Stuttgart, Germany.

V. Results and Discussion

A. Behaviour Evaluation

Fig. 1 illustrates human avoidance in Context-MPC scenarios testing, featuring relevant human-related information such as a 3D skeleton with activity labels, 2D projections and probability



(a) (b) (c)

Fig. 4. Statistical Analysis Scenarios: (a) Two Different Activities, (b) Three Standing Humans, and (c) Three Walking Humans. Each scenario shows the initial position of the robot, the target position with red circles, the view from the onboard camera, and the RViz view.

TABLE II
BEHAVIOR EVALUATION METRICS RESULTS: COMPARISON OF t_g AND HUMAN DISTANCE METRICS BETWEEN HUMAN-MPC, HUMAN-MPC + ELLS, HUMAN-MPC + J_{vel}, AND CONTEXT-MPC.

Walking Scenario	$t_g[s]$	d_{min} [m]	d_{avg} [m]
HuMAN-MPC	32.03	1.26	4.56
HuMAN-MPC + Ells	21.45	1.43	4.60
HuMAN-MPC + J_{vel}	25.60	1.37	4.44
Context-MPC	26.57	1.39	4.34

cost maps. The two red ellipses are used to obtain the local plan (red line) concerning the global path (green line). Fig. 3 displays the results of the MPC variants in the human walking scenario. Fig. 3(a) shows the robot's trajectories, starting from the black circle and ending at the black square, to avoid a human who starts walking from the blue circle and ends at the blue square, in the walking scenario for each variant. All trajectories show significant robot deviation due to integrating human motion predictions, allowing earlier avoidance for safer, more efficient navigation. Context-MPC by incorporating context information reduces the excessive deviation seen in HuMAN-MPC, resulting in less conservative trajectories. Fig. 3(b) presents the robot's linear velocity profiles for each scenario, indicating the maximum socially accepted velocity $v_{sm} = 0.5$ m/s (red dot line) and the desired velocity v_{des} (black dot line) based on human activity (Table I).

Table II provides the ablation results revealing improvements in task efficiency (lower goal-reaching time) with each added contextual component. Considering 3D human body poses (via ellipses) increases the minimum human-robot distance, yet it does not affect the robot velocity during human avoidance, as illustrated in Fig. 3(b). Therefore, the introduction of the new velocity cost term is justified, achieving the desired and socially acceptable velocity levels (i.e., the HuMAN-MPC + J_{vel} variant). The latter approach treats humans as 2D entities with uniform distance constraints thus penalizing collision avoidance metrics. The mean distance to the closest human decreases with the Context-MPC, reinforcing efficiency with its less conservative behavior. Overall, the Context-MPC strikes a balance, enhancing task efficiency, maintaining greater distance from humans, and ensuring safety by slowing down during human encounters.

B. Statistical Validation

Fig. 4 depicts these experiments, with the initial position of the robot and the achieved goals marked by a red circle. Table III presents the results, including the average time to reach the goal, average minimum distance from humans, average mean human distances, and their standard deviations. Finally, Figs. 5, 6, and 7 show qualitative performance comparison of Context-MPC against TEB.

TABLE III
STATISTICAL ANALYSIS OF t_g AND HUMAN DISTANCE METRICS BETWEEN TEB, HUMAN-MPC, AND CONTEXT-MPC.

Average	$t_g \pm \sigma[s]$	$d_{min} \pm \sigma$ [m]	$d_{avg} \pm \sigma$ [m]
Two Activities			
TEB	26.57 ± 2.62	0.45 ± 0.24	5.78 ± 0.20
HuMAN-MPC	49.25 ± 4.40	0.96 ± 0.18	4.85 ± 0.52
Context-MPC	42.45 ± 3.04	1.19 ± 0.12	4.32 ± 0.27
Standing Hum.			
TEB	57.86 ± 31.20	1.03 ± 0.30	4.47 ± 1.38
HuMAN-MPC	49.34 ± 3.19	1.14 ± 0.06	5.51 ± 0.10
Context-MPC	42.87 ± 2.45	1.23 ± 0.02	4.56 ± 0.13
Walking Hum.			
TEB	35.96 ± 8.61	0.79 ± 0.50	3.99 ± 0.84
HuMAN-MPC	33.17 ± 1.87	1.01 ± 0.28	4.66 ± 0.19
Context-MPC	30.25 ± 0.5	1.36 ± 0.10	4.26 ± 0.2

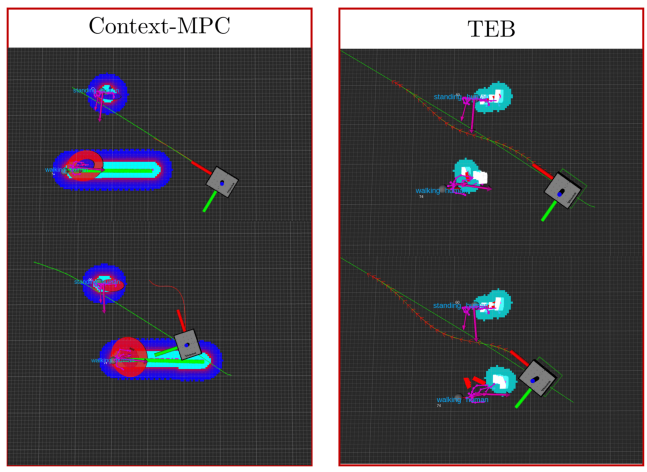


Fig. 5. Two Activities Scenario: Context-MPC proactively avoids collisions using motion prediction of the incoming person. Instead, TEB slows down and stops shortly before the imminent collision, forcing the person to change path.

Results in Table III show the advantages of Context-MPC, with a minimum increase in task efficiency of 8.78% and a 35.49% improvement in safety metrics in the most dynamic situations. Moreover, even if the TEB planner achieved the goal faster in the Two Activities Scenario (Table III, $t_g = 26.57 \pm 2.62$ s), it is not sufficient to ensure safety in human avoidance.

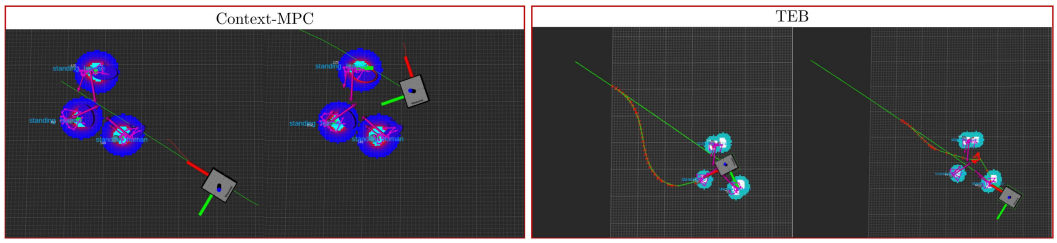


Fig. 6. Standing Humans Scenario: Using Context-MPC the robot plans, in advance, an evasive maneuver before approaching the group, which prevents getting stuck into a standing person or in the middle of the group, as it happens with the Timed Elastic Band.

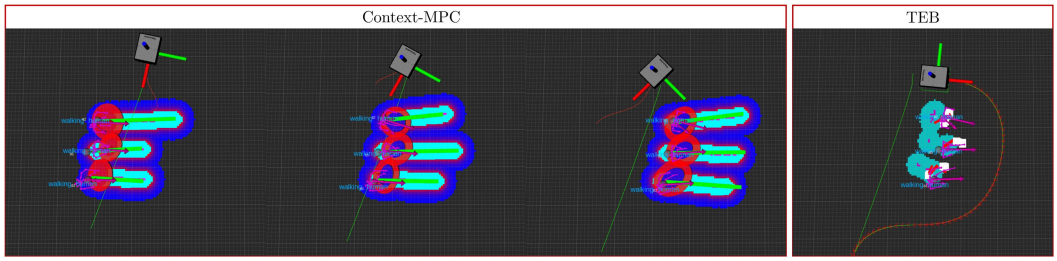
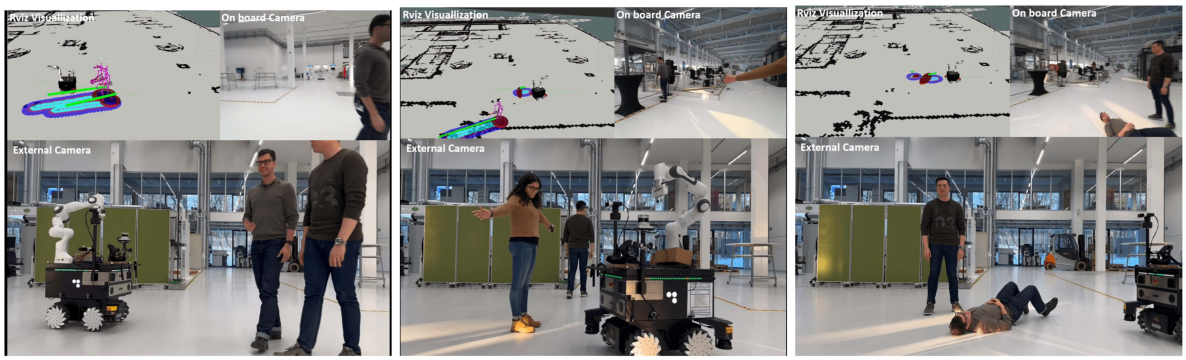


Fig. 7. Three People Walking Scenario: Context-MPC planner slows down when close to people with detected "walking" activity. On the contrary, TEB planner in a similar situation maintains speed, attempting to overtake in front of the walking group, but never manages to execute its maneuver around the continuously moving people.



(a) Two Walking People Scenario (b) Different Activities Scenario (c) Lying Down scenario

Fig. 8. Each figure is an instance in a real-life experiment on the robot, showing the Rviz visualization, the onboard camera image, and an image captured by an external camera.

Indeed, Fig. 5 depicts the situation when the TEB planner has to stop in front of the walking person, ($d_{min} = 0.45 \pm 0.24$ m) before continuing the navigation, while the Context-MPC proactively avoids the human.

Similarly, in the Standing Humans scenario, the TEB is frequently not able to avoid the group of people, while our planner efficiently considers the activities and avoids the standing group with a reasonable decrease in velocity. The same behaviors are visible in the Walking Humans case (Fig. 7), where the TEB planner attempts to maintain speed and overtake the walking people, which may result in collisions and intrusive behavior, while our method benefits from slowing down in front of the group, allowing the robot to pass behind, see Fig. 6.

Finally, adding contextual information does not affect the computational behavior or scalability of Context-MPC compared to HuMAN-MPC. It remains linear with agents, and computation time is under 5 milliseconds per iteration, as detailed in our previous work [10].

C. Real World Experiments

The system was tested in three different dynamic scenarios: two walking people, three humans engaged in various activities (walking and standing), two humans talking, and at one point, one human lying down (an unknown activity). Specific instances for each scenario are illustrated in Fig. 8 with RViz views, the onboard camera images, and external camera images. For a detailed visualization, see the multimedia materials. In the following video we replicated scenarios discussed in Section IV-B1 in a real environment: https://www.dropbox.com/scl/fi/si1ox69k2l7ekdsuk4h8o/Video.mp4?rlkey=3w0xsuypkau7625632mf3jl2m&dl=0

These experiments validate Context-MPC running in real-time on ROS-based robots, demonstrating efficient predictive navigation around humans, as seen in Fig. 8(a) and (b). In addition, precise identification of the space occupied by a person using ellipses is shown in Fig. 8(c).

VI. Conclusion

We propose a novel MPC formulation and system that includes not only geometric information but also context that better explains fine-grained human behaviors. Differently from the prior art, our formulation explicitly considers articulated 3D human poses and semantic activity labels together with 2D motion predictions. Tested extensively in various human-avoidance scenarios and implemented on the DARKO ROS-based platform, the system demonstrated effective real-time operation in dynamic human environments, highlighting the value of 3D human context information for enhanced navigation. The approach is significantly more efficient and safer than the baselines that use only 2D predictions, thus proving our hypothesis that using contextual information can improve the overall robot navigation performance.

Future research will focus on assessing the planner's effectiveness in more diverse human scenarios and gathering data on human comfort during navigation.

References

[1] A. Rudenko et al., "Human motion trajectory prediction: A survey," *Int. J. Robot. Res.*, vol. 39, no. 8, pp. 895–935, 2020.

[2] C. Mavrogiannis et al., "Core challenges of social robot navigation: A survey," *ACM Trans. Hum.-Robot Interact.*, vol. 12, no. 3, 2023, Art. no. 36.

[3] K. J. Singh, D. S. Kapoor, and B. S. Sohi, "Understanding social conventions for socially aware robot navigation," *IEEE Potentials*, vol. 42, no. 3, pp. 37–42, May/Jun. 2023.

[4] L. Heuer, L. Palmieri, A. Rudenko, A. Mannucci, M. Magnusson, and K. O. Arras, "Proactive model predictive control with multi-modal human motion prediction in cluttered dynamic environments," in *2023 IEEE/RSJ IEEE Int. Conf. Intell. Robots Syst.*, 2023, pp. 229–236.

[5] A. Tapus et al., "Perceiving the person and their interactions with the others for social robotics–a review," *Pattern Recognit. Lett.*, vol. 118, pp. 3–13, 2019.

[6] T. Schoels, L. Palmieri, K. O. Arras, and M. Diehl, "An NMPC approach using convex inner approximations for online motion planning with guaranteed collision avoidance," in *2020 IEEE Int. Conf. Robot. Autom.*, 2020, pp. 3574–3580.

[7] S. Schaefer, K. Leung, B. Ivanovic, and M. Pavone, "Leveraging neural network gradients within trajectory optimization for proactive human-robot interactions," in *2021 IEEE Int. Conf. Robot. Autom.*, 2021, pp. 9673–9679.

[8] B. Holman, A. Anwar, A. Singh, M. Tec, J. Hart, and P. Stone, "Watch where you're going! gaze and head orientation as predictors for social robot navigation," in *2021 IEEE Int. Conf. Robot. Autom.*, 2021, pp. 3553–3559.

[9] T. Schreiter et al., "The magni human motion dataset: Accurate, complex, multi-modal, natural, semantically-rich and contextualized," 2022, *arXiv:2208.14925*.

[10] T. Hielscher et al., "Towards using fast embedded model predictive control for human-aware predictive robot navigation," in *Long-term Hum. Motion Prediction Workshop - ICRA*, 2024.

[11] R. Verschueren et al., "acados: A modular open-source framework for fast embedded optimal control," *Math. Programm. Comput.*, vol. 14, pp. 147–183, 2022.

[12] V. Vulcano, S. G. Tarantos, P. Ferrari, and G. Oriolo, "Safe robot navigation in a crowd combining NMPC and control barrier functions," in *2022 IEEE 61st Conf. Decis. Control*, 2022, pp. 3321–3328.

[13] J. Chen, X. Chen, and S. Liu, "Trajectory planning of autonomous mobile robot using model predictive control in human-robot shared workspace," in *2023 IEEE 3rd Int. Conf. Electron. Technol., Commun. Inf.*, 2023.

[14] J. Rios-Martinez et al., "From proxemics theory to socially-aware navigation: A survey," *Int. J. Social Robot.*, vol. 7, pp. 137–153, 2015.

[15] T. Akhtyamov et al., "Social robot navigation through constrained optimization: A comparative study of uncertainty-based objectives and constraints," 2023, *arXiv:2305.02859*.

[16] M. Everett, Y. F. Chen, and J. P. How, "Collision avoidance in pedestrian-rich environments with deep reinforcement learning," *IEEE Access*, vol. 9, pp. 10357–10377, 2021.

[17] Y. F. Chen, M. Liu, M. Everett, and J. P. How, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *2017 IEEE Int. Conf. Robot. Autom.*, 2017, pp. 285–292.

[18] Y. Chen, F. Zhao, and Y. Lou, "Interactive model predictive control for robot navigation in dense crowds," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 4, pp. 2289–2301, Apr. 2022.

[19] C. Zheng et al., "Deep learning-based human pose estimation: A survey," *ACM Comput. Surv.*, vol. 56, no. 1, pp. 1–37, Aug. 2023.

[20] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7291–7299.

[21] Y. Xu et al., "ViTPose: Simple vision transformer baselines for human pose estimation," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 38571–38584, 2022.

[22] Y. Yuan and K. Kitani, "Dlow: Diversifying latent flows for diverse human motion prediction," in *Comput. Vis.–ECCV 2020: 16th Eur. Conf.*, Glasgow, U.K., Proceedings, Part IX 16. Springer, 2020, pp. 346–364.

[23] P. Nikdel, M. Mahdavian, and M. Chen, "DMMGAN: Diverse multi motion prediction of 3D human joints using attention-based generative adversarial network," in *2023 IEEE Int. Conf. Robot. Automat.*, 2023, pp. 9938–9944.

[24] I. Sárándi, T. Linder, K. O. Arras, and B. Leibe, "MeTRAbs: Metric-scale truncation-robust heatmaps for absolute 3D human pose estimation," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 3, no. 1, pp. 16–30, Jan. 2021.

[25] T. Linder, K. Y. Pfeiffer, N. Vaskevicius, R. Schirmer, and K. O. Arras, "Accurate detection and 3D localization of humans using a novel YOLO-based RGB-D fusion approach and synthetic training data," in *2020 IEEE Int. Conf. Robot. Automat.*, 2020, pp. 1000–1006.

[26] T. Linder, S. Breuers, B. Leibe, and K. O. Arras, "On multi-modal people tracking from mobile platforms in very crowded and dynamic environments," in *2016 IEEE Int. Conf. Robot. Automat.*, 2016, pp. 5512–5519.

[27] M. Mahdavian, P. Nikdel, M. TaherAhmadi, and M. Chen, "STPOTR: Simultaneous human trajectory and pose prediction using a non-autoregressive transformer for robot follow-ahead," in *2023 IEEE Int. Conf. Robot. Automat.*, 2023, pp. 9959–9965.

[28] A. Ulhaq et al., "Vision transformers for action recognition: A survey," 2022, *arXiv:2209.05700*.

[29] M. Diehl et al., "Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations," *J. Process Control*, vol. 12, no. 4, pp. 577–585, 2002.

[30] G. Frison and M. Diehl, "Hpipm: A high-performance quadratic programming framework for model predictive control," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 6563–6569, 2020.

[31] S. B. Vatan et al., "Social APF-RL: Safe mapless navigation in unknown & human-populated environments," in *2023 Eur. Conf. Mobile Robots*. IEEE, 2023.

[32] H. Zhou et al., "A hybrid obstacle avoidance method for mobile robot navigation in unstructured environment," *Ind. Robot: Int. J. Robot. Res. Appl.*, vol. 50, no. 1, pp. 94–106, 2023.

[33] "The CGAL project, CGAL user and reference manual, 5.6 ed. CGAL editorial board," 2023. [Online]. Available: https://doc.cgal.org/5.6/Manual/packages.html

[34] E. W. Weisstein, "Ellipse," From MathWorld–A Wolfram Web Resource, s.d. [Online]. https://mathworld.wolfram.com/Ellipse.html

[35] R. Bohannon, "Comfortable and maximum walking speed of adults aged 20-79 years: Reference values and determinants," *Age ageing*, vol. 26, pp. 15–19, 1997.

[36] J.-F. Bonnans et al. *Numerical Optimization: Theoretical and Practical Aspects*. Berlin, Germany: Springer Science & Business Media, 2006.

[37] J. T. Butler and A. Agah, "Psychological effects of behavior patterns of a mobile personal robot," *Auton. Robots*, vol. 10, pp. 185–202, 2001.

[38] T. Kruse et al., "Human-aware robot navigation: A survey," *Robot. Auton. Syst.*, vol. 61, no. 12, pp. 1726–1743, 2013.

[39] M. Kamel, J. Alonso-Mora, R. Siegwart, and J. Nieto, "Robust collision avoidance for multiple micro aerial vehicles using nonlinear model predictive control," in *2017 IEEE/RSJ IEEE Int. Conf. Intell. Robots Syst.*, 2017, pp. 236–243.

[40] ISO, *Robotic Devices–Safety Requirements for Personal Care Robots*, International Organization for Standardization, 2014.

[41] DARKO Consortium, "Deliverable D2.2–Initial report on perception in DARKO," 2023.

[42] C. Rösmann, W. Feiten, T. Wösch, F. Hoffmann, and T. Bertram, "Efficient trajectory optimization using a sparse model," in *2013 IEEE Eur. Conf. Mobile Robots*., 2013, pp. 138–143.

[43] L. Palmieri and K. O. Arras, "A novel RRT extend function for efficient and smooth mobile robot motion planning," in *2014 IEEE/RSJ IEEE Int. Conf. Intell. Robots Syst.*, 2014, pp. 205–211.