


By Sergi Molina , Anna Mannucci , Martin Magnusson , Daniel Adolfsson ,
Henrik Andreasson , Mazin Hamad , Saeed Abdolshah , Ravi Teja Chadalavada ,
Luigi Palmieri , Timm Linder , Chittaranjan Srinivas Swaminathan , Tomasz Piotr Kucner ,
Marc Hanheide , Manuel Fernandez-Carmona , Grzegorz Cielniak , Tom Duckett ,
Federico Pecora , Simon Bokesand, Kai O. Arras, Sami Haddadin , and Achim J. Lilienthal 

The ILIAD Safety Stack

Human-Aware Infrastructure-Free Navigation of Industrial Mobile Robots

Current intralogistics services require keeping up with e-commerce demands, reducing delivery times and waste, and increasing overall flexibility. As a consequence, the use of automated guided vehicles (AGVs) and, more recently, autonomous mobile robots (AMRs) for logistics operations is steadily increasing.

INTRODUCTION

AGVs are optimized for high performance, predictability, and reliability, while AMRs are more focused on flexibility, agility, and simplicity. AGV technology is based on predefined (virtu-

al) paths, space discretization and, in many cases, supporting infrastructure. AMRs, on the other hand, leverage a smaller mobile base and on-board intelligence to target safe and flexible navigation in infrastructure-free environments. In environments where humans and robots share the workspace, human-robot interaction (HRI) is central and forces AMRs/AGVs to keep up with health and safety regulations at work (e.g., ISO 3691-4:2020) and account for both general safety (i.e., collision avoidance) and perceived safety. While general safety can be ensured by certified safety sensors (e.g., SICK laser scanners), targeting perceived safety (i.e., rendering robot motions more legible to humans) is still an open research problem and is key to increasing human satisfaction with the deployment and operation of mobile robots in mixed human-robot environments.

To this end, several international research projects have boosted research toward more flexible and capable mobile intralogistics robots in the last decade. Among them, ILIAD (Intra-Logistics with Integrated Automatic Deployment: safe and scalable fleets in shared spaces, H2020-EU Project Agreement ID 732737; see <https://iliad-project.eu>) was a European Union (EU) H2020 project that ran from January 2017 to June 2021 with an overarching goal of addressing the limitations in the state of the art impeding the efficient use of a fleet of autonomous forklift trucks in warehouse logistics—most prominently with a focus on safety, efficient deployment, and object manipulation. The manipulation system has been described by Garabini et al. [7].

Here, we focus on the ILIAD’s approach to human safety in navigation, which is based on the synergistic coordination of multiple safety layers from short to long time scales (jointly referred to as the *safety stack* in the following). In short-term interactions, the typical strategy between most current intralogistics robots and human workers is simple; the vehicle stops whenever an obstacle, human or not, is detected in front of it. This is safe yet highly inefficient, and attempts to improve on this are bound by the tradeoff between safety and task efficiency. Predictable motion is important for heavy vehicles in professional environments, and the kinematics of, e.g., forklift trucks do not allow for swift remaneuvering. We argue that robots with semantic understanding, capable of learning models of human behaviors and planning accordingly, can act more safely by exploiting a deeper understanding of social context and how context is likely to unfold during operation. The novelties we propose are summarized as follows:

- a new human-aware safety stack composed of different layers accounting for short- to long-term HRIs
- an implementation and evaluation of the synergy between the safety layers in a real-world warehouse and a simulated environment.

To the best of our knowledge, this is the first safety stack designed for heterogeneous fleets of wheeled robots moving in mixed human–robot infrastructure-free environments.

Research on the individual components of the safety stack has been previously published separately [3], [5], [9], [11], [12], [13], [14], [17], [18], [19], [21], [23], [25], [26], [27]. The main contribution of the present article is the system-level integration of the components that make up the layers of the safety stack, including experimental studies of the performance of the stack and its layers.

BACKGROUND

Layered architectures are a common design approach for robotic navigation systems. One of the earliest is the Three-Layer Architecture [8], a hybrid reactive/deliberative architecture consisting of three main blocks: 1) a reactive feedback control layer, 2) a

reactive plan layer, and 3) a deliberative layer. This three-layer structure still persists in broadly used stacks such as the Robotic Operating System (ROS) [15] and ROS2 [16] navigation, where a local planner runs at higher rates in parallel with a slower global planner. Each planner uses a private costmap to represent the local environment, addressing different semantics and thus using different rates and information sources (from sensors to objects or human trackers). Both the ROS and ROS2 navigation stacks are essentially designed for single robots moving in dynamic unstructured environments. Conversely, the ILIAD architecture is designed for heterogeneous multi-AMR/AGV systems (specifically including robots with nontrivial dimensions and kinematics, such as forklift trucks) moving in human-shared unstructured environments [17].

Thanks to ROS, the gap between AMR/AGV technologies has been widely reduced (2013 to present). A timeline of notable AGV developments [20] includes Kiva Robots (2012: first swarm AGV), Robotnik (2014: first ROS-based AGV), and OTTO 1500 (2015: first collaborative AGV). Other important results advancing AGV navigation capabilities and resulting from collaborations between academia and industries are Magazino and Osnabrück University [22], Elettric80 and PAN-Robots [24], and Kollmorgen Automation and Örebro University (SAUNA/Semantic Robots) [1].

ILIAD extends the modular architecture of the SAUNA project and exploits centralized decision-making modules to achieve globally consistent constraints on robot behaviors, together with decentralized components, to enable flexible decision making at the level of individual AGVs. ILIAD’s extensions toward improving the perceived safety and legibility of motions are detailed in the “[ILIAD Safety Stack: Layer Design](#)” and “[ILIAD Safety Stack: Implementation and Layer Integration](#)” sections.

ILIAD SAFETY STACK: LAYER DESIGN

Our safety stack is based on five layers, each one focused on a different HRI level (see [Table 1](#)). The highest layer tries to

“
HERE, WE FOCUS
ON THE ILIAD’S
APPROACH TO
HUMAN SAFETY
IN NAVIGATION,
WHICH IS BASED
ON THE SYNERGISTIC
COORDINATION OF MULTIPLE
SAFETY LAYERS
FROM SHORT TO
LONG TIME SCALES.
”

TABLE 1. Safety layers in ILIAD.

NO.	LAYER	ROLE
5	Flow-aware global motion planning	Create paths consistent with human patterns
4	Navigation intent communication	Improve the legibility of robot motions
3	Situation-aware constraints	Include local interactions in planning
2	Vehicle safe motion unit	Limit robot speed to account for pain/injury biomechanics
1	Safety stops/slowdowns	Avoid collisions

prevent unnecessary HRIs and potentially harmful maneuvers (layer 5). The middle layers (2 to 4) present relevance to HRIs at local shorter time frames, and finally, the lowest layer ensures basic safety (layer 1). Although mandatory in industrial AMRs/AGVs, an emergency stop button (layer 0) is not considered part of the stack since it needs to be manually triggered. The following sections describe the theoretical ideas of each independent layer, going from layers 1 to 5, all of them developed within the ILIAD project. The implementation of the layers is covered in more detail in the “ILIAD Safety Stack: Implementation and Layer Integration” section, with experimental validation in the “Evaluation” section.

LAYER 1—SAFETY STOP

The lowest safety layer is the industry-standard safety-certified laser scanner hardware, which prevents the robot from getting too close to any obstacle, either human or nonhuman, within a preconfigured safety zone. This halts the robot if any reading coming from the safety laser falls within the defined emergency stop safety area.

LAYER 2—VEHICLE SAFE MOTION UNIT

Layer 2 is in charge of ensuring the biomechanical safety of nearby humans by limiting the robot velocities using a mobile vehicle version of the generalized safe motion unit (vSMU) [9]. Based on the inertial properties of the vehicle’s points of interest (POIs), the Cartesian velocity along possible impact directions, and the most representative experimental impact datasets in an injury database [27], the vSMU pipeline produces instantaneous biomechanically safe velocity limits for the mobile robot, as illustrated in Figure 1. Inside the injury database, the human impact information is usually encoded in a so-called safety curve that eases exchanging the required injury/safety status versus the reflected impact dynamics. Each of these safety curves specifies the maximum biomechanically safe velocity

as a function of the instantaneous robot reflected mass per impact curvature at the contact location. For the vSMU, this safe velocity limit is obtained from live data about possibly colliding human body parts (provided by a perception module [14]), inertial properties, and surface curvature for the mobile robot (usually specified from robot designs or through CAD models). Finally, any desired mobile robot velocity for which at least one POI exceeds the specified

limit, as encoded via the most representative safety curve for the anticipated dynamic impact, is scaled down.



TO THE BEST OF OUR KNOWLEDGE, THIS IS THE FIRST SAFETY STACK DESIGNED FOR HETEROGENEOUS FLEETS OF WHEELED ROBOTS MOVING IN MIXED HUMAN-ROBOT INFRASTRUCTURE-FREE ENVIRONMENTS.



LAYER 3—SITUATION-AWARE PLANNING

Layer 3 incorporates HRI situations by encoding these in the form of costmaps. This is achieved in three steps, corresponding to three submodules.

QUALITATIVE TRAJECTORY CALCULUS VERSION C STATES

First, the spatiotemporal movement of robot and human is encoded using Qualitative Trajectory Calculus version C (QTC_C) [6], where pairs of human and robot trajectories are represented by a four-tuple of state descriptors (h_1, r_1, h_2, r_2). Each descriptor expresses a qualitative spatial relation between the human (h) and the robot (r). With this four-tuple of descriptors composed of three symbols ($-$, 0 , $+$), there are a total of 81 possible QTC_C states [e.g., QTC_C state $(-, -, 0, -)$ means that the human is moving toward the robot and the

robot is approaching, too, and the human is headed directly toward the robot, but the robot is headed toward the left side of the human].

MULTIHIDDEN MARKOV MODEL HUMAN-ROBOT SPATIAL INTERACTION SITUATION CLASSIFIER

The QTC_C encoding is used in a sequence to classify the type of human-robot spatial interaction (HRSI) (Figure 2) using a multihidden Markov model (multi-HMM) [23]. The

six HRSI classes that we consider are defined in Figure 2: passing by on the left (PBL), passing by on the right (PBR), robot overtakes left (ROL), robot overtakes right (ROR), path crossing from the left (PCL), and path crossing from the right (PCR).

SITUATIONAL CONSTRAINTS

Finally, the predicted class is used to generate a costmap around the human so that it can be used by other modules of the navigation architecture [5]. A default constraint costmap is defined to create a safety area around the closest

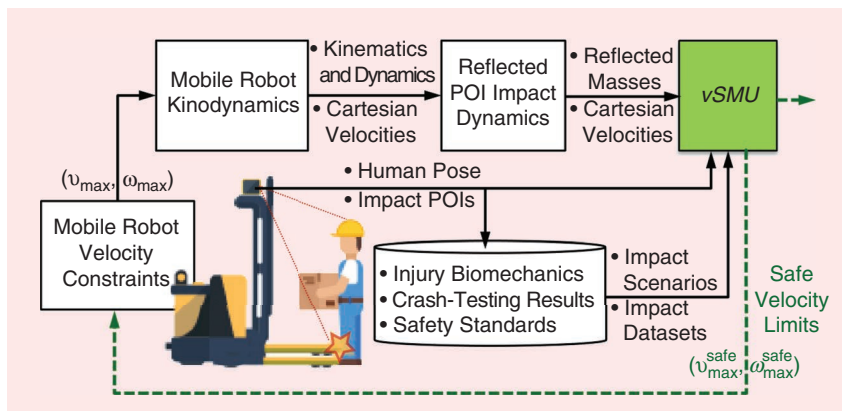


FIGURE 1. The pain/injury-biomechanics-based vSMU pipeline.

human when none of the situations modeled by the multi-HMM situation classifier are detected. Cell costs in this costmap depend on two parameters, the distance to the human (d) and the arc around the human–robot connecting line (a), defined as

$$\begin{aligned} \text{cost}_{\text{cell}} &= e^{-(0.75d)^2} \cdot e^{-(4a)^2}, \\ d &= \sqrt{(x_{\text{cell}} - x_{\text{human}})^2 + (y_{\text{cell}} - y_{\text{human}})^2}, \\ a &= \text{atan2}\left(\frac{y_{\text{cell}} - y_{\text{robot}}}{x_{\text{cell}} - x_{\text{robot}}}\right) - \text{atan2}\left(\frac{y_{\text{human}} - y_{\text{robot}}}{x_{\text{human}} - x_{\text{robot}}}\right). \end{aligned} \quad (1)$$

In case one of the HRSI classes is detected by the multi-HMM classifier, a situation-specific costmap is overlaid on top of the default one. For the situations PBL, ROL, and PCL, higher costs [defined by (2)] are assigned to cells on the left \overrightarrow{hr} line, covering a much wider penalization area (Figure 3). For the situations PBR, ROR, and PCR, the same approach is followed but on the right side of the line \overrightarrow{hr} .

$$\text{high_cost}_{\text{cell}} = e^{-(0.3d)^2} \quad (2)$$

LAYER 4—NAVIGATION INTENT COMMUNICATION

The fourth layer aims at improving perceived safety in HRIs. It comprises the bidirectional communication of navigation intent between robots and humans [3].

In the direction of *robot to human*, we have developed a navigation intent communication system based on spatial augmented reality (SAR): projecting graphics on the floor near the robot to show, e.g., the planned driving direction. This enables the robot to communicate its navigation intent with the aim of making humans intuitively understand the robot's intention and feel safe in the vicinity of robots.

In the direction of *human to robot*, eye gaze can convey information about intentions beyond what can be inferred from the trajectory and head pose of a person. Hence, we propose eye-tracking glasses as safety equipment in industrial environments shared by humans and robots. If the eye gaze is within the area of interest (AOI), then the projected arrow remains static, and if the eye gaze is not within a defined AOI of the robot, then the projected arrow blinks to get the human's attention (see Figure 4). (Videos demonstrating the intent

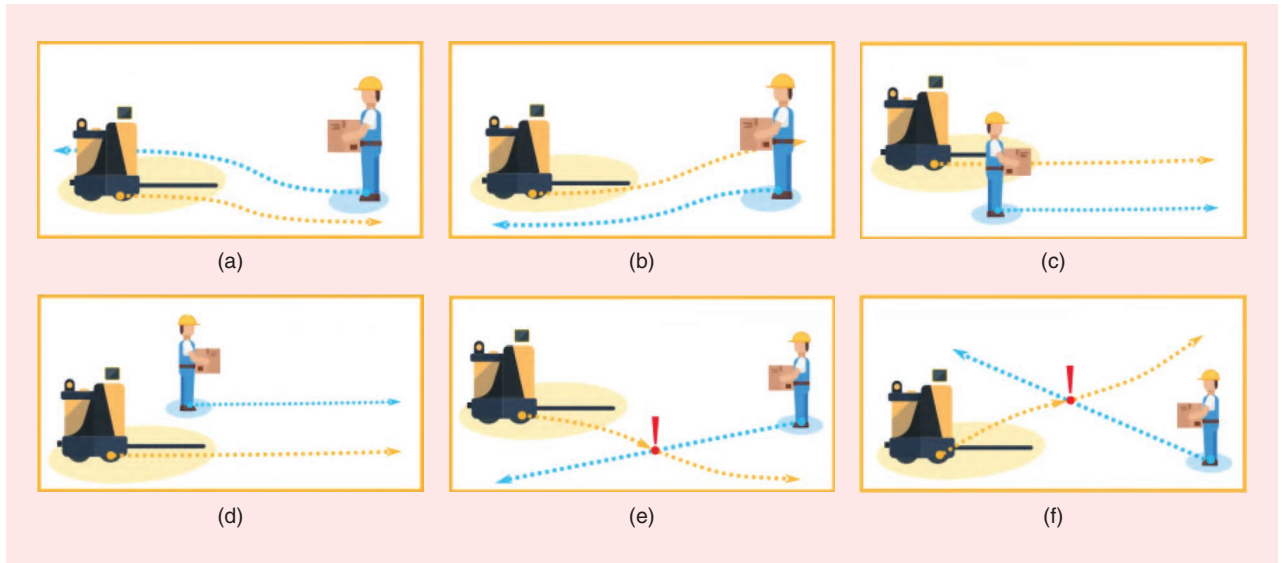


FIGURE 2. The six HRSI classes used in the multi-HMM classifier. (a) Passing by on the left (PBL). (b) Passing by on the right (PBR). (c) Robot overtakes left (ROL). (d) Robot overtakes right (ROR). (e) Path crossing on the left (PCL). (f) Path crossing on the right (PCR).

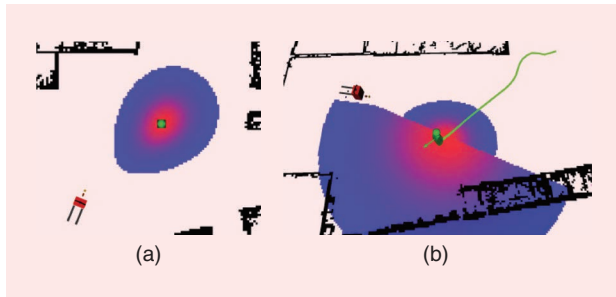


FIGURE 3. (a) and (b) the detected person is shown with the green human-shaped marker, its velocity vector is shown by the green arrow, and its trajectory is shown by a green line. The navigational costmap is marked with a blue (low cost) to red (high cost) gradient. (a) Default costmap. (b) "PCL" situation costmap.



FIGURE 4. (1) The eye-tracker device. (2) The Empatica E4 band used for measuring electrodermal activity. (3) The SAR projection "arrow" projected on the shared floor space to convey the robot's navigation intent. (4) The robot (one of ILIAD's retrofitted AGVs) with the SAR intention communication system. (5) The defined AOI. (6) The eye gaze fixation.

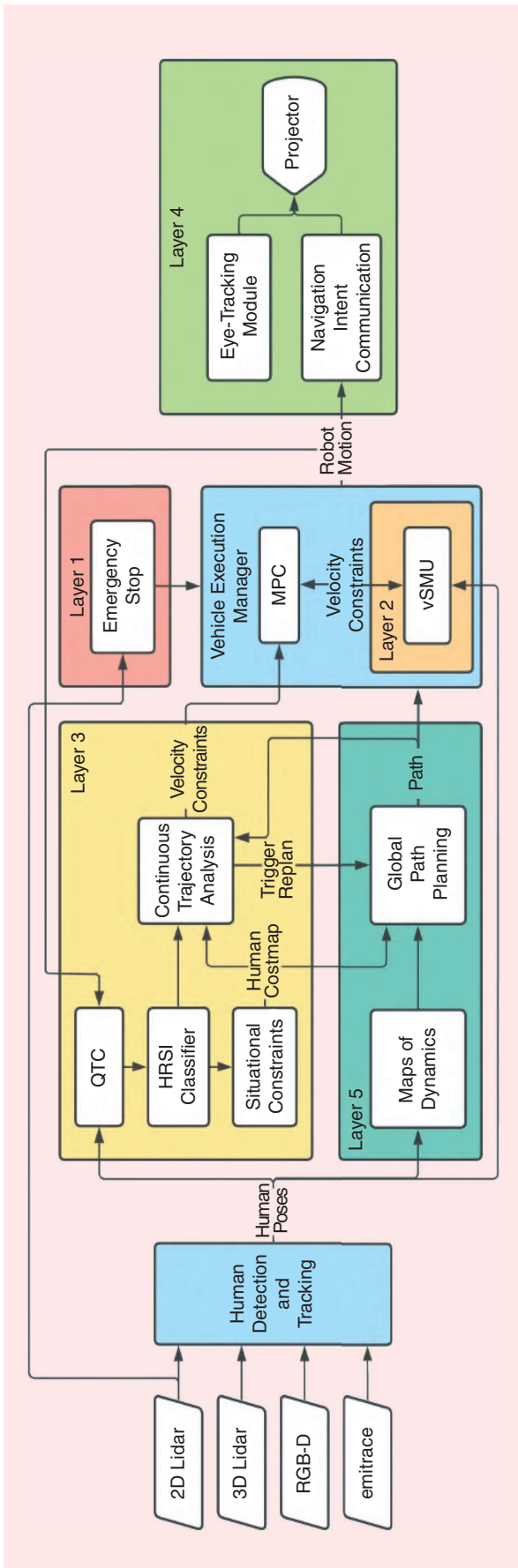


FIGURE 5. The interaction between ILIAD safety layers.

communication can be seen at <https://youtu.be/IMEp6TcjDiw> and https://youtu.be/ov8q_KXB2a4.)

LAYER 5—HUMAN FLOW-AWARE MOTION PLANNING

The top-level safety layer is flow-aware motion planning using maps of dynamics (MoDs) that represent human motion patterns [11], [21], [25], [26]. This layer creates (global) robot motions that aim to reduce the amount of human-robot interference by using learned long-term patterns of human motion. In ILIAD, we have developed two MoDs for modeling site-specific motion patterns [circular-linear flow field map (CLiFF-map) and spatiotemporal flow map (STeF-map)] together with motion planners capable of exploiting such information.

MoDs

The CLiFF-map [12] is a representation for encoding patterns of movement as a field of Gaussian mixtures. The input of the model is a velocity vector \mathbf{V} combining the orientation ($\theta \in [0, 2\pi)$) and speed ($\rho \in \mathbb{R}^+$) for every person detection. Then, for each location in a discrete grid, a “semiwrapped” Gaussian mixture model is employed to model and preserve the multimodality of the pedestrian motion.

The STeF-map [18] is a representation that models the likelihood of motion directions by a set of harmonic functions that capture long-term changes in crowd movements over time. The input used to create the model is a tuple containing the position, orientation, and time stamp (x, y, θ, t) for every person detection. The underlying geometric space is represented by a grid, where each cell contains k temporal models, corresponding to k discretized orientations of the motion of people.

While CLiFF-map stores a continuous representation of multimodal flow at each discrete point in the map, it does not encode dependencies on the time of day, etc. STeF-map, on the other hand, learns to predict periodic (e.g., daily) patterns of motion but does so for a discrete set of directions (eight in our case) in each map cell.

GLOBAL PATH PLANNING

We have presented three flow-aware cost functions that can be used in the rapidly exploring random tree (RRT*) motion planning algorithm; two are based on CLiFF-maps (i.e., the CLiFF Extended Upstream Criterion [21] and Down-The-CLiFF cost [25]), and one is based on STeF-maps (the STeF Extended Upstream Criterion [26]). In CLiFF-RRT*, we generate a robot trajectory by adopting a biasing approach in RRT* [10]; we first generate a discrete path that selects mixtures at relevant locations from the map of dynamics and then use those mixtures to bias the sampling and rewiring procedures in RRT*. In Down-The-CLiFF (DTC)-RRT*, we further extend the CLiFF-RRT* by also considering the mean speed encoded in the distribution and the information about uncertainty. Both strategies have been shown to improve the overall efficiency of robot operation while being flow aware in such human-aware scenarios [26].

ILIAD SAFETY STACK: IMPLEMENTATION AND LAYER INTEGRATION

In this section, we describe the implementation and interconnection between the different layers in the ILIAD setup. Figure 5 provides a graphical overview of the ILIAD safety architecture. The two modules in blue (the vehicle execution manager and human detection and tracking) do not strictly belong to the safety stack yet are interacting with it.

ROBOT PLATFORMS

Two types of forklifts have been employed in the ILIAD project and in the experiments carried out here: a large forklift and two small pallet trucks (Figure 6). The larger forklift is a Toyota SAE200 BT autopilot truck, which was used during the real-world experiments for the safety stack. The smaller forklifts are Linde CiTi One pallet trucks, whose digital twin is used during the experiments carried out in the simulations. Both types of vehicles have three-wheeled kinematics common to forklifts, with one combined steer and drive wheel and two passive wheels. The motion planning and control pipeline [2] includes kinodynamic-aware motion planners (detailed in the “Layer 5” section) followed by a model predictive control (MPC) controller for accurate trajectory tracking running at 16.7 Hz. The sensor suite and velocity limits for both platforms are defined in Table 2, and the sensor locations are shown in Figure 6.

HUMAN DETECTION AND TRACKING

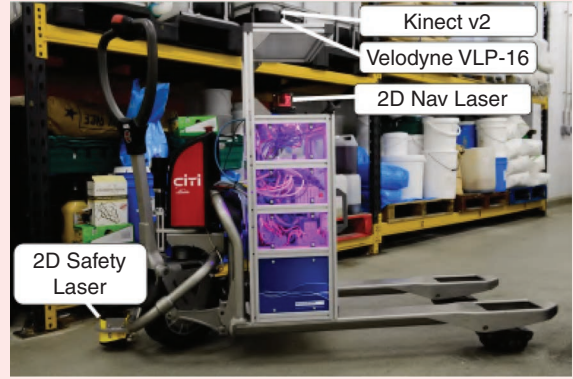
A multimodal human detection and tracking system also developed during the ILIAD project [13] provides input to safety layers 2, 3, and 5. This perception system fuses data from the robots’ heterogeneous on-board sensors (one or more RGB-D cameras, 3D lidars, and 2D safety laser sensors and a novel emitrace [https://www.retenua.com/en/products/emitrace] safety camera [19]). The system outputs human trajectories with an update rate on the human poses of 9 Hz on average and a detection range of up to 13 m.

LAYER 1

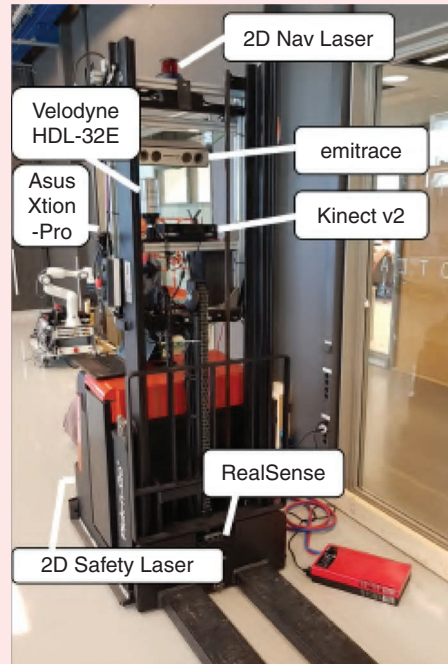
This layer is primarily implemented inside the safety laser scanner, which is interfaced with the vehicle through a set of safety relays. The lowest level e-stop is done by cutting the power to the motors and activating the disk brake as soon as a reading coming from the 2D safety laser is ≤ 0.4 m. To have the flexibility of an online reconfigurable slowdown and soft-stop functionality, a dynamic slowdown and stop area is also implemented. If any reading from the safety laser occurs within the area the robot will occupy in a reconfigurable lookahead time of 2 s, the vehicle execution module through the MPC will either slow down or stop the robot. The response time of this layer depends on the mounted safety laser.

LAYER 2

Since every desired task velocity must be checked by the vSMU, this component is integrated inside the vehicle



(a)



(b)

FIGURE 6. (a) and (b) The sensor locations of both platforms. (a) A pallet truck. (b) A forklift truck. Nav: navigation

TABLE 2. Robot parameters.

ROBOT		TOYOTA SAE200 BT	LINDE CiTi ONE
Sensors	2D laser	SICK S3000	SICK S300 Expert
	RGB-D	Microsoft Kinect v2	Asus Xtion-Pro
	3D lidar	Velodyne HDL-32E	Velodyne VLP-16
	Other	emitrace	—
Velocity	v_{\min}, v_{\max}	0.01 m/s, 0.3 m/s	0.01 m/s, 0.3 m/s
	$\omega_{\min}, \omega_{\max}$	0.01 rad/s, 0.3 rad/s	0.01 rad/s, 0.3 rad/s

execution manager to act on the system as quickly as possible. By continuously monitoring the relative distance d_{rel} between the vehicle POI (body and forks) and the closest human, the vSMU can be employed to calculate biomechanically safe speed limits (v_{max}^{safe} and ω_{max}^{safe}) as soon as $d_{rel} \leq d_{vSMU} = 4$ m. Those safe limits are then sent back to replace the initial task limits for the MPC when needed ($v_{max}^{safe} < v_{max}$ and $\omega_{max}^{safe} < \omega_{max}$). The response time of this layer is ~ 110 ms, which is based on the update rate of the human detection system (~ 9 Hz).

LAYER 3

The costmaps coming from the situational constraints module have two uses; they serve as an input to shape valid paths generated by the path planner, and they are used by the continuous trajectory assessment (CTA) module to continuously monitor the potential human risk on the remaining path.

CONTINUOUS TRAJECTORY ANALYSIS

Using the human costmap provided by the situational constraints module and the path envelope given by the global path planner [see Figure 7(a)], the CTA calculates a cost c_{path} , defined as the highest cell cost found in the remainder of the computed path, which describes the human risk level at a rate of 2 Hz. Depending on the risk level, the CTA module can trigger different actions.

- *No risk* ($0 \leq c_{path} < 35$): Nothing happens.
- *Low risk* ($35 \leq c_{path} < 75$): This allows one to stick to the original navigation plan, which is preferred, but it sends velocity constraints to the vehicle execution manager based on the HRSI situation detected.
 - *PBL or ROL*: ($v = 0.5 \cdot v_{max}$, $v^{rev} = -v_{min}$, $\omega = 0.5 \cdot \omega_{max}$, $\omega^{rev} = -\omega_{max}$).
 - *PBR or ROR*: ($v = 0.5 \cdot v_{max}$, $v^{rev} = -v_{max}$, $\omega = 0.5 \cdot \omega_{max}$, $\omega^{rev} = -\omega_{min}$).

- *PCL or PCR*: All velocities are at the minimum.
- *No situation detected*: All velocities are at 50%.
- *High risk* ($c_{path} \geq 75$): All velocities are at the minimum, and in case the high risk is maintained for longer than 15 s, a recalculation of the current global path is requested.

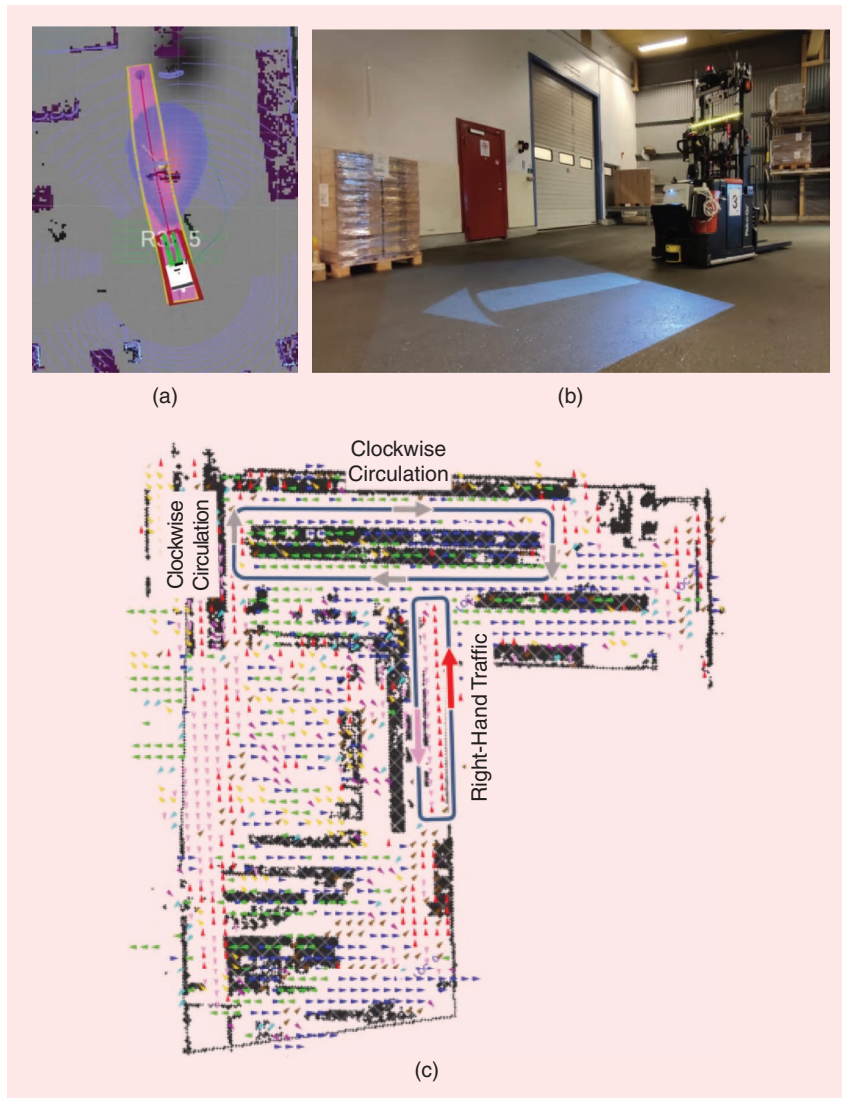


FIGURE 7. (a) Layer 3: a situation where a velocity constraint is triggered. The path envelope is in yellow, and the human costmap is within the purple gradient. (b) Layer 4: the communication of navigational intents by means of a projected arrow on the floor. A horizontal LED bar indicates whether the robot has the safety stack on (green) or off (yellow). (c) Layer 5: an STeF-map showing motion patterns learned from several days of observations.

LAYER 4

The eye-tracking module was not active during the experimental evaluation in the “Evaluation” section because we could not equip all of the staff with eye trackers. Therefore, the implementation of layer 4 that was used in the on-site experimental validation in the “Evaluation” section is a uni-directional robot-to-human system based on a projector displaying the current steering angle with an arrow on the floor [Figure 7(b)].

LAYER 5

MoDs

To generate the MoDs, the human detection and tracking system was deployed on the big forklift for a duration of four days while an operator was using it in a normal routine, obtaining approximately 725,000 human detections corresponding to around 5,000 trajectories. An example of the STeF-map prediction at midday using these data for training is shown in Figure 7(d). It can be seen that the model is able to automatically learn

unwritten site-specific “traffic rules,” such as a clockwise circulation around one of the shelving units and right-hand traffic in a narrow corridor.

GLOBAL PATH PLANNING

This module can take into account both global human patterns from flow maps and local human information from layer 3. Planning in intralogistics scenarios like the ones depicted in Figures 7 and 8, which include narrow and cluttered environments and forklift trucks whose kinematics require a large area when turning, can be particularly demanding from a computational point of view. To this end, a system that runs three different planning algorithms in parallel is deployed, namely a flow-aware variant of bidirectional transition-based RRT (BiTRRT) [4] using the DTC cost [25], a flow-aware DTC variant of probabilistic roadmap [10], and a state-lattice planner [2]. The path generated by the state-lattice planner is used if and only if the flow-aware planners do not find a solution. Otherwise, the shortest path between the two flow-aware solutions is returned.

EVALUATION

In this section, we present the experimental setup, the methodology, relevant performance metrics, and the results of the experiments. Real-world experiments allow us to analyze the performance of the proposed stack while accounting for real HRIs with factory workers. We leverage the reproducibility of tests in the simulated environment to include a more detailed ablation study.

ENVIRONMENT

The experimental evaluation was carried out in both a real food distribution warehouse (Orkla Foods in Örebro, Sweden) and a digital twin of the same warehouse.

REAL WORLD

Data collection was performed in one of the warehouses of food producer Orkla Foods [a picture of the area where the goods are stored can be seen in Figure 8(a)]. The area allowed for the robot deployment and experimentation is around 700 m² [marked with a tan color in Figure 8(b)]. This particular warehouse is rather old and comes with a number of challenges compared to newer warehouses designed with automation in mind. The L shape of the building reduces visibility and leads to longer paths for driving; the floor is uneven and hence more difficult to drive while assuming a 2D planar navigation; and there is little open space for maneuvering in general, which is challenging for sampling-based motion planners. The workers in the real-world trials gave their informed consent to participate. While this research involves human subjects, addi-

“

THE PREDICTED CLASS IS USED TO GENERATE A COST-MAP AROUND THE HUMAN SO THAT IT CAN BE USED BY OTHER MODULES OF THE NAVIGATION ARCHITECTURE.

”

tional ethics approval was not required under Swedish law since we did not collect sensitive personal data, nor make any physical intervention.

SIMULATION

The simulated environment was developed using ROS Gazebo, with data from the real Orkla warehouse. To make the digital twin as close as possible to the real one, a 3D mapping procedure to obtain a point cloud of the whole warehouse was performed. After removing the points belonging to the ceiling structures and floor, the point cloud was converted into a 3D mesh; the final 3D model is shown in Figure 8(b).

To emulate the warehouse workers, four simulated moving pedestrians were added in four different regions of the deployment area [red rectangles



FIGURE 8. Real and simulated warehouse environments used during the experiments. (a) A picture from the real Orkla Foods warehouse. (b) A 3D model of the Orkla warehouse used in the simulation and a digital model of the small pallet truck. The robot deployment area is in orange, the goal locations are marked with blue circles, the pedestrian navigation areas are marked with red rectangles, and the camera location for the picture seen in (a) is marked with the green camera icon.

in Figure 8(b)]. For each pedestrian, a set of trajectories was generated by exploiting the STeF-map model trained with four days of real human motion data recorded in the warehouse. Each pedestrian was provided with 20 different trajectories to be completed sequentially. All trajectories were kept the same for all experiments.

To enable the simulated pedestrians to follow the trajectories, each pedestrian was spawned running its own instance of the standard ROS navigation stack. The maximum linear and angular speeds for all pedestrians were set to 0.6 m/s and 1 rad/s, which are close to the average speeds recorded during the four-day recording period previously mentioned in the “Layer 5” section.

METHODOLOGY

We required the robot to navigate among common pick-and-place locations during normal daily activities involving human workers. A total of six different locations (denoted with the letters A–F) were defined as shown in Figure 8(b).

REAL WORLD

The test scenario consisted of driving the big forklift truck between locations F–B–C–D–B–F (~ 100 m) five times. The following three safety layer configurations were evaluated under the same testing conditions:

- *R1*: only layer 1
- *R2*: active layers 1, 3, and 5
- *R3*: active layers 1, 3, 4, and 5

For each configuration, the path F–B–C–D–B–F was repeated five times.

Despite being an important part of the architecture, layer 2 was disabled for the workers’ safety since it is not yet safety certified; similarly to layer 1, layer 2 also operates at the lowest level. Thus, simultaneously activating both layer 1 (a safety-certified hardware solution) and layer 2 (an external software solution) requires a rigorous definition of the involved triggering distance thresholds. This is mainly to ensure that the safety-certified part (i.e., layer 1 in our case) is always active such that ethical approvals and safety standards can be followed. Regarding layer 5, we used CLiFF-map as the single instantiation of MoDs.

SIMULATION

In simulations, two test scenarios were used, both using the digital model of the small pallet truck [see Figure 8(b)]. The first replicates a similar scenario to the real-world experiments and requires driving the robot between goal points A–F–B–C–D–B–A (~ 140 m), and the second requires driving the robot between C–B–E–D–C (~ 100 m).

We evaluated four different layer configurations.

- *S1*: only layer 1
- *S2*: active layers 1 and 2

- *S3*: active layers 1, 2, and 3
- *S4*: active layers 1, 2, 3, and 5.

For each scenario and configuration, 10 runs were performed. Regarding layer 1, only the soft-stop area was used as the slowdown behavior is handled by the constraints from layer 2 and the motor power cut e-break is not reproducible. For layers 2 and 3, all the human detections used as the input were provided from the ground truth positions in the Gazebo simulator at a rate of 25 Hz to reduce the computational load. Layer 4 was not included in the

simulation since people’s reaction to this communication of navigational intent is subjective and hence difficult to evaluate in simulation. The MoD used for layer 5 was the CLiFF-map as the STeF-map was used to generate the trajectories for the pedestrians.

METRICS

To measure the usefulness of the safety stack and the impact of the different layers in the architecture, the following metrics are proposed:

- *Task efficiency*: This is defined as the total distance driven by the robot divided by the interval of time needed to complete its mission. It is expressed in meters per second. The time to compute the paths is omitted.
- *Layer activations*: We propose to measure the number of triggers of each layer (per minute of operation) to measure the impact of each one on the others.
- *layer 0*: in case the emergency button had to be pushed (only in the real-world experiments)
- *layer 1*: the number of times the emergency slowdown/stop was activated
- *layer 2*: the number of times the vSMU had to intervene (only in the simulations)
- *layer 3*: the number of times this layer sent velocity constraints to the robot controller (path replans are omitted since only one was observed during a single test in the simulations).
- *Human safety*: This is done by means of the robot linear speed (in absolute values), based on two different conditions: 1) checking the speed when layer 1 is activated and 2) checking the speed every time a person is detected within the vSMU range in layer 2 (4 m).
- *Robot legibility and perceived safety*: This is done using a questionnaire to measure how understandable/predictable the robot behavior is from the human perspective (only in the real-world experiments). The questionnaire can be summarized with two questions.
- How mentally demanded, rushed, stressed, insecure, and comfortable did you feel when walking around the robot?

“ THE WORKERS GENERALLY RATED THE ROBOT MORE HIGHLY IN THE ASPECTS OF COMMUNICATION, SITUATION AWARENESS, AND RELIABILITY WHEN THE SAR MODULE WAS TURNED ON. ”

- How would you rate the communication from the robot side regarding easiness to understand, clearness, and helpfulness?

RESULTS

REAL-WORLD DEPLOYMENT

The results of the real-world experiments are summarized in Table 3 and Figure 9. In terms of task efficiency, we see that configuration R2 gets the lowest score, probably due to the increased number of layer 1 and layer 3 activations. However, R1 is the only configuration in which the emergency button had to be pressed for safety reasons.

Four workers who were active in the vicinity of the robot during the days of the experiments were asked to fill out the questionnaires while the SAR module was both OFF and ON (R2 and R3 respectively). While we did not notice any significant differences with respect to perceived safety, we did notice a tendency of the workers to favor the robot when the SAR module was ON. The workers generally rated the robot more highly in the aspects of communication, situation awareness, and reliability when the SAR module was turned ON. This increased robot-to-human communication effectiveness in R3 could explain both the decrease of layer 1 and layer 3 activations to their minimum and also the slightly increased median speed of the robot when humans are nearby, as seen in Figure 9.

SIMULATION

The results of both tested scenarios are shown in Figure 10. Regarding task efficiency, an extra configuration called the *baseline* is added, referring to an environment with no pedestrians. For both scenarios, we observed that while configuration S1 keeps a very similar task efficiency as the baseline, as soon as layer 2 is introduced in configuration S2, the efficiency decreases by roughly 20%. However, the inclusion of layers 3 and 5 (configurations S3 and S4) has no impact.

Looking at the number of layer activations, we do not see any clear evidence that adding more layers affects the trigger counts of lower ones. However, adding more layers does affect the robot speed both when layer 1 is triggered and when humans are in the vicinity of the robot. While in configuration S1, the median speed is at its maximum of 0.3 m/s with both conditions, for the rest of the configurations, the robot median speed is reduced by at least 50%.

DISCUSSION

While layers 2 and 4 seem to have a higher impact on the metrics proposed, the effect

of layers 3 and 5 is not clear with the current setup. Layer 3 could probably benefit from the introduction of additional higher frequency local replanning capabilities. This would take into account the HRSI costmaps without being limited to just global replans and velocity constraints. Regarding layer 5, the limitations seem to come from the environment where the experiments were carried out. The combination of narrow aisles with complex robot kinodynamics restricts the paths the robot can take. Unfortunately, layer 2 could

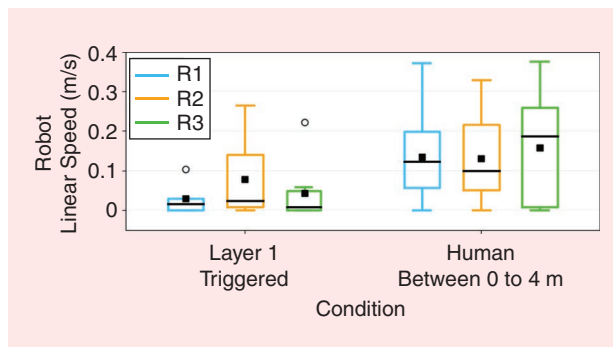


FIGURE 9. Human safety metrics in the real-world experiments. Data were accumulated for all the runs in the respective configuration. The black line represents the median, and the black square represents the mean.

TABLE 3. Task efficiency and layer activation results for the real-world experiments.

CONFIG.	RUN	TASK EFFICIENCY (m/s)	LAYER 1 (TRIGGERS/MIN)	LAYER 3 (TRIGGERS/MIN)
R1	1	0.25	0.15	—
	2	0.25	0.29	—
	3	0.24	0.27	—
	4	Manual e-brake (layer 0) pressed		
	5	0.27	0	—
	Median	0.25	0.21	—
R2	1	Aborted run: robot interrupting warehouse activities		
	2	0.26	0	0.65
	3	0.22	0.5	0.88
	4	0.2	0.27	0.8
	5	Aborted run: robot got stuck		
	Median	0.22	0.27	0.8
R3	1	0.27	0.45	0.6
	2	0.21	0.6	1.64
	3	0.25	0	0
	4	0.25	0	0.57
	5	0.23	0.18	0.72
	Median	0.25	0.18	0.6

The bold values refer to the median of all runs corresponding to the same configuration. Config.: configuration.

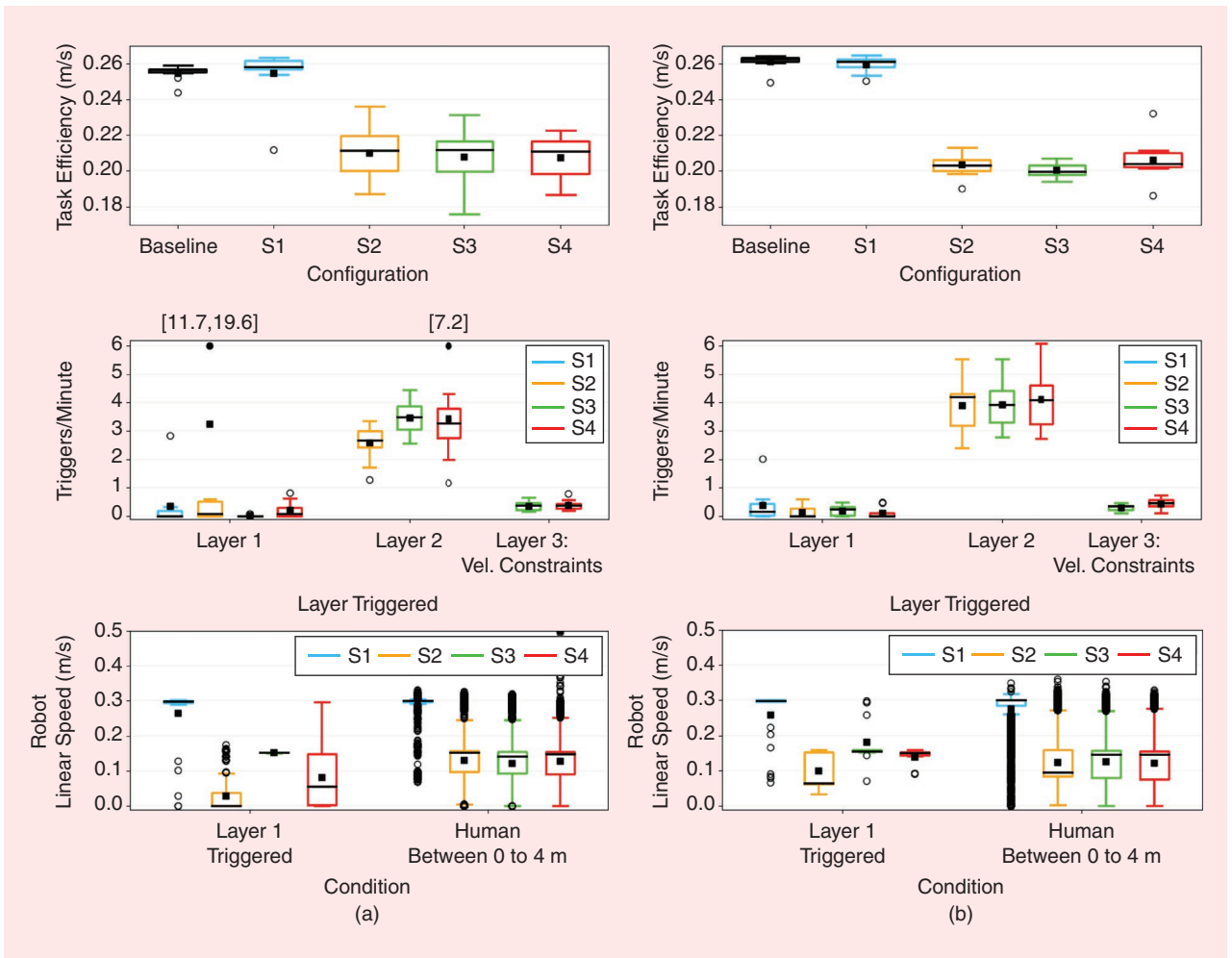


FIGURE 10. Task efficiency, times each layer is triggered, and human safety for each configuration and scenario in simulation. Data were accumulated for all the runs in the respective configuration. The black line represents the median, the black square represents the mean, and the black solid circles represent the outliers outside the y-axis range. (a) Scenario 1: path AFBCDBA. Scenario 2: path CBEDC. Vel.: velocity.

not be evaluated in the real-world setup due to safety protocols, but the results obtained in the simulation seem promising, and an effort on the integration of layer 1 and layer 2 synergies is intended as future work.

SUMMARY AND CONCLUSION

We have proposed, implemented, and evaluated (through experiments on a real robot and in simulation) a new safety stack that leverages learned human motion patterns and increased onboard intelligence toward smarter human-aware navigation. The stack consists of five layers ranging from low to high HRI levels. Specifically, layer 1 ensures basic collision avoidance; layer 2 limits the robot velocity to ensure the biomechanical safety of nearby humans; layer 3 introduces qualitative HRIs; layer 4 enables the improved legibility of robot motions; and layer 5 incorporates long-term human motion patterns in the path-planning step. The results of our ablation study show that when using multiple layers, a decrease in task efficiency is expected in favor of a safer and more legible robot when humans are close. We

believe the results of this article are, however, a good initial step toward including human awareness in the safety stack. The results highlighted the need for global orchestration among layers and a tighter synergistic layer design and implementation to target the Technology Readiness Levels required by current industries.

ACKNOWLEDGMENT

This work has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement 732737 (ILIAD). This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

AUTHORS

Sergi Molina, University of Lincoln, LN6 7TS Lincoln, U.K. E-mail: sergimolina91@gmail.com.

Anna Mannucci, Robert Bosch GmbH, 71272 Renningen, Germany. E-mail: anna.mannucci90@gmail.com.

Martin Magnusson, Örebro University, 70182 Örebro, Sweden. E-mail: martin.magnusson@oru.se.

Daniel Adolfsson, Örebro University, 70361 Örebro, Sweden. E-mail: dla.adolfsson@gmail.com.

Henrik Andreasson, Örebro University, 70182 Örebro, Sweden. E-mail: henrik.andreasson@oru.se.

Mazin Hamad, Technical University of Munich, 80992 Munich, Germany. E-mail: mazinhamd@gmail.com.

Saeed Abdolshah, Technical University of Munich, 80992 Munich, Germany. E-mail: saeed.abdolshah@tum.de.

Ravi Teja Chadalavada, Örebro University, 70182 Örebro, Sweden. E-mail: ravi.chadalavada@oru.se.

Luigi Palmieri, Robert Bosch GmbH, 71272 Renningen, Germany. E-mail: palmieri@informatik.uni-freiburg.de.

Timm Linder, Robert Bosch GmbH, 71272 Renningen, Germany. E-mail: timm.linder@de.bosch.com.

Chittaranjan Srinivas Swaminathan, Örebro University, 70182 Örebro, Sweden. E-mail: ksatyaki@gmail.com.

Tomasz Piotr Kucner, Aalto University, FI-00076 Aalto, Finland. E-mail: tomasz.kucner@aalto.fi.

Marc Hanheide, University of Lincoln, LN6 7TS, Lincoln, U.K. E-mail: marc@hanheide.net.

Manuel Fernandez-Carmona, University of Lincoln, LN6 7TS, Lincoln, U.K. E-mail: mfcarmona@uma.es.

Grzegorz Cielniak, University of Lincoln, LN6 7TS, Lincoln, U.K. E-mail: gcielniak@lincoln.ac.uk.

Tom Duckett, University of Lincoln, LN6 7TS, Lincoln, U.K. E-mail: tom.d.duckett@gmail.com.

Federico Pecora, Örebro University, 70182 Örebro, Sweden. E-mail: fpecora@amazon.com.

Simon Bokesand, Kollmorgen Automation AB, 431 90 Mölndal, Sweden. E-mail: simon.bokesand@kollmorgen.com.

Kai O. Arras, Robert Bosch GmbH, 71272 Renningen, Germany. E-mail: kaioliver.arras@de.bosch.com.

Sami Haddadin, Technical University of Munich, 80992 Munich, Germany. E-mail: haddadin@tum.de.

Achim J. Lilienthal, Örebro University, 70182 Örebro, Sweden. E-mail: achim.lilienthal@oru.se.

REFERENCES

[1] H. Andreasson et al., "Autonomous transport vehicles: Where we are and what is missing," *IEEE Robot. Autom. Mag.*, vol. 22, no. 1, pp. 64–75, Mar. 2015, doi: 10.1109/MRA.2014.2381357.

[2] H. Andreasson, J. P. Saarinen, M. Cirillo, T. Stoyanov, and A. J. Lilienthal, "Fast, continuous state path smoothing to improve navigation accuracy," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, 2015, pp. 662–669.

[3] R. Chadalavada, H. Andreasson, M. Schindler, R. Palm, and A. Lilienthal, "Bi-directional navigation intent communication using spatial augmented reality and eye-tracking glasses for improved safety in human-Robot interaction," *Robot. Comput.-Integr. Manuf.*, vol. 61, Feb. 2020, Art. no. 101830, doi: 10.1016/j.rcim.2019.101830.

[4] D. Devaurs, T. Siméon, and J. Cortés, "Enhancing the transition-based RRT to deal with complex cost spaces," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, 2013, pp. 4120–4125, doi: 10.1109/ICRA.2013.6631158.

[5] M. F. Carmona, T. Parekh, and M. Hanheide, "Making the case for human-aware navigation in warehouses," in *Towards Autonomous Robotic Systems*, K. Althoefer, J. Konstantinova, and K. Zhang, Eds. Cham, Switzerland: Springer International Publishing, 2019, pp. 449–453.

[6] N. Van de Weghe et al., "A qualitative trajectory calculus and the composition of its relations," in *GeoSpatial Semantics*, vol. 3799, M. A. Rodríguez, I. Cruz, S. Levashkin, and M. J. Egenhofer, Eds. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 60–76.

[7] M. Garabini et al., "WRAPP-up: A dual-arm robot for intralogistics," *IEEE Robot. Autom. Mag.*, vol. 28, no. 3, pp. 50–66, Sep. 2021, doi: 10.1109/MRA.2020.3015899.

[8] E. Gat, R. P. Bonnaso, and R. Murphy, "On three-layer architectures," in *Artificial Intelligence and Mobile Robots*. Menlo Park, CA, USA: AAAI Press, 1997, pp. 195–210.

[9] M. Hamad, A. Kurdas, N. Mansfeld, S. Abdolshah, and S. Haddadin, "Modularize-and-conquer: A generalized impact dynamics and safe precollision control framework for floating-base tree-like robots," *IEEE Trans. Robot.*, early access, 2022, doi: 10.1109/TRO.2023.3257515.

[10] S. Karaman and E. Frazzoli, "Sampling-based algorithms for optimal motion planning," *Int. J. Robot. Res.*, vol. 30, no. 7, pp. 846–894, 2011, doi: 10.1177/0278364911406761.

[11] T. P. Kucner, A. J. Lilienthal, M. Magnusson, L. Palmieri, and C. S. Swaminathan, *Probabilistic Mapping of Spatial Motion Patterns for Mobile Robots (Cognitive Systems Monographs)*, vol. 40. Cham, Switzerland: Springer International Publishing, 2020.

[12] T. P. Kucner, M. Magnusson, E. Schaffernicht, V. H. Bennetts, and A. J. Lilienthal, "Enabling flow awareness for mobile robots in partially observable environments," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 1093–1100, Apr. 2017, doi: 10.1109/LRA.2017.2660060.

[13] 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 *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, 2020, pp. 1000–1006, doi: 10.1109/ICRA40945.2020.9196899.

[14] T. Linder, N. Vaskevicius, R. Schirmer, and K. O. Arras, "Cross-modal analysis of human detection for robotics: An industrial case study," in *Proc. IEEE Int. Conf. Intell. Robots Syst. (IROS)*, 2021, pp. 971–978, doi: 10.1109/IROS51168.2021.9636158.

[15] D. V. Lu, D. Hershberger, and W. D. Smart, "Layered costmaps for context-sensitive navigation," in *Proc. IEEE Int. Conf. Intell. Robots Syst. (IROS)*, 2014, pp. 709–715, doi: 10.1109/IROS.2014.6942636.

[16] S. Macenski, F. Martín, R. White, and J. G. Clavero, "The marathon 2: A navigation system," in *Proc. IEEE Int. Conf. Intell. Robots Syst. (IROS)*, 2020, pp. 2718–2725, doi: 10.1109/IROS45743.2020.9341207.

[17] A. Mannucci, L. Pallottino, and F. Pecora, "On provably safe and live multi-robot coordination with online goal posting," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1973–1991, Dec. 2021, doi: 10.1109/TRO.2021.3075371.

[18] S. Molina, G. Cielniak, and T. Duckett, "Modelling and predicting rhythmic flow patterns in dynamic environments," in *Towards Autonomous Robotic Systems*, vol. 10965, M. Giuliani, T. Assaf, and M. Giannaccini, Eds. Cham, Switzerland: Springer-Verlag, 2018, pp. 135–146.

[19] R. Mosberger, B. Leibe, H. Andreasson, and A. Lilienthal, "Multi-band Hough Forests for detecting humans with reflective safety clothing from mobile machinery," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, Piscataway, NJ, USA: IEEE Press, May 2015, pp. 697–703, doi: 10.1109/ICRA.2015.7139255.

[20] E. A. Oyekanlu et al., "A review of recent advances in automated guided vehicle technologies: Integration challenges and research areas for 5G-based smart manufacturing applications," *IEEE Access*, vol. 8, pp. 202,312–202,353, Nov. 2020, doi: 10.1109/ACCESS.2020.3035729.

[21] L. Palmieri, T. Kucner, M. Magnusson, A. J. Lilienthal, and K. O. Arras, "Kinodynamic motion planning on Gaussian mixture fields," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2017, pp. 6176–6181, doi: 10.1109/ICRA.2017.7989731.

[22] S. Pütz, J. Santos Simón, and J. Hertzberg, "Move base flex: A highly flexible navigation framework for mobile robots," in *Proc. IEEE Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 3416–3421, doi: 10.1109/IROS.2018.8593829.

[23] L. Roberts-Elliott et al., "Qualitative probabilistic models of HRSI for safe situational human-aware navigation," Ph.D. thesis, Univ. of Lincoln, Lincoln, U.K., 2021.

[24] L. Sabattini et al., "The PAN-robots project: Advanced automated guided vehicle systems for industrial logistics," *IEEE Robot. Autom. Mag.*, vol. 25, no. 1, pp. 55–64, Mar. 2018, doi: 10.1109/MRA.2017.2700325.

[25] C. S. Swaminathan, T. P. Kucner, M. Magnusson, L. Palmieri, and A. Lilienthal, "Down the CLiFF: Flow-aware trajectory planning under motion pattern uncertainty," in *Proc. IEEE Int. Conf. Intell. Robots Syst. (IROS)*, 2018, pp. 7403–7409, doi: 10.1109/IROS.2018.8593905.

[26] C. S. Swaminathan et al., "Benchmarking the utility of maps of dynamics for human-aware motion planning," *Frontiers Robot. AI*, vol. 9, Nov. 2022, Art. no. 916153, doi: 10.3389/frobt.2022.916153.

[27] "ILIAD deliverable D4.4: Report on human injury data for autonomous mobile systems," Technical University of Munich, Munich, Germany, Tech. Rep., 2020. [Online]. Available: http://iliad-project.eu/wp-content/uploads/Deliverables/D4_4.pdf