Human-aware Robot Navigation in Logistics Warehouses

Mourad A. Kenk[1,2] [a], M. Hassaballah[3] [b] and Jean-François Brethé[1] [c]

[1]*Electrotechnics and Automatics Research Group (GREAH), Normandy University, Le Havre, France*
[2]*Department of Mathematics, South Valley University, Qena, Egypt*
[3]*Department of Computer Science, South Valley University, Luxor, Egypt*

Keywords: Human-aware Navigation, Collision Avoidance, Human Detection in RGB-D Camera, 2D-LIDAR Sensor, Logistics Warehouses, Mobile Robot Autonomous Navigation, Robotics.

Abstract: Industrial and mobile robots demand reliable and safe navigation capabilities to operate in human populated environments such as advanced manufacturing industries and logistics warehouses. Currently mobile robot platforms can navigate through their environment avoiding coworkers in the shared workspace, considering them as static or dynamic obstacles. This strategy is efficient for safety, strictly speaking, but is not sufficient to provide humans integrity and comfortable working conditions. To this end, this paper proposes a human-aware navigation framework for comfortable, reliable and safely navigation designed to run in real-time on a mobile robot platform in logistics warehouses. This is accomplished by estimating human localization using RGB-D detector, then generating a virtual circular obstacle enclosing human pose. This virtual obstacle is then fused with the 2D laser range scan and used in ROS navigation stack local costmap for human-aware navigation. This strategy guarantees a different approach distance to obstacles depending on the human or non-human nature of the obstacle. Hence the mobile robot can approach closely to pallet to pick up objects while maintaining an integrity distance to humans. The reliability of the proposed framework is demonstrated in a workbench of experiments using simulated mobile robot navigation in logistics warehouses environment.

1 INTRODUCTION

In logistics warehouses, pick and place mobile robots are designed to navigate in an environment often structured in long and narrow aisles. Robots need to approach very closely the racks located on the side of the aisle (Wahrmann et al., 2017). In the same time, robots are likely to cross human coworkers. Humans need to feel safe in their personal space when sharing workspace with robots. In vision-based autonomous navigation, several typical obstacle avoidance methods have been proposed recently. Krajnik et al. (Krajník et al., 2017) proposed to use heterogeneous visual features such as points, line segments, lines, planes, and vanishing points to process a visual simultaneous localization and mapping (SLAM) or to use depth information from a low-cost sensor (RGB-D) to localize the obstacles. In (Yang et al., 2016), a nonlinear controller is designed to achieve target tracking and obstacle avoidance in complex environ-
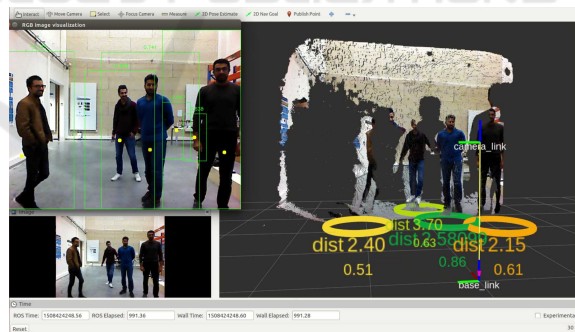

Figure 1: Example of an output frame of ROS based human perception process, showing detector and localizer results in RViz.

ments. In (Malone et al., 2017), stochastic reachable sets are used to generate accurate artificial potential field for dynamic obstacles for online path planning. In (Nardi and Stachniss, 2017), a probabilistic approach is developed for modeling uncertain trajectories of the moving entities that share workspace with robot.

[a] https://orcid.org/0000-0002-2374-4032
[b] https://orcid.org/0000-0001-5655-8511
[c] https://orcid.org/0000-0002-6962-954X

Most of the previous research works consider humans as dynamic obstacles using different techniques and sensing modalities. Nevertheless, considering humans as dynamic obstacles to simply avoid them is certainly sufficient from the safety point of view, but can cause annoyances for humans. Indeed, proxemics theory, one of the most popular principles in Human Robot Spatial Interaction (HRSI), advises to select appropriate interaction distances between robots and humans (Lindner and Eschenbach, 2017). The robot should keep a certain distance from a human while navigating, preferably greater than 1.2m not to violate the Personal Space (PS).

Currently, many researches are devoted to design navigation approaches with safety control in complex environments. For instance, Choi et al. (Choi et al., 2017) proposed Gaussian process motion regression based robot navigation, which predicts future trajectory of human with two Microsoft Kinect sensors in dynamic environments. In (Song et al., 2017), a shared-control scheme based navigation is proposed via combining active obstacle avoidance and passive-compliant motion behavior prediction for human, where walking-assistant robot avoids collision and allows safe guidance for human using leg detector by 2D laser sensor. In (Zimmermann et al., 2018), the robot trajectory planning is done by estimating 3D human pose in RGB-D images while executing tasks in cluttered environments occupied with different features of the obstacles.

Numerous solutions have been presented for human detection and tracking in a human populated environment based on 2D Laser range sensors (Linder et al., 2016; Song et al., 2017). The core idea of human detection is a binary classifier trained on leg appearance features that is reflected by beams of 2D laser sensor. First, the reflected 2D laser data (points) is segmented according to Euclidian distance (jump distance) between continuously neighboring 2D points. These segments are then classified as a human leg if it scores approximation values of 2D geometrical structure features, which are characterized by pre-defined fourteen parameters such as number of reflected points, width, linearity, circularity, radius as described by (Arras et al., 2007). Nevertheless, even if these methods achieve great performances in airport and hospital environment, they are not applicable in logistics warehouses environment due to the high number of false detection as reported in this study. On the other hand, human detectors based on 3D vision show high performances in industrial environment (Munaro et al., 2016) and intralogistics warehouses (Linder et al., 2018).

For these reasons, the main goal of this research is to design a human-aware navigation framework that makes the robot behaviour more sociable (as shown in Figure 1). It implies being able to distinguish between human and non-human obstacles, which is currently the case in the state-of-the-art. We investigate the constraints of human detection algorithm implementation taking into account different point of view, from the computational costs, to the safety and proxemics theory requirements. The proposed framework is designed to be optimal in logistics warehouses environments with their specific properties and constraints. Then, the functionality of the framework is proved in an experiment based on autonomous Summit XL mobile robot in a logistics warehouses simulated environment relying on the well known Robot Operating System (ROS) (Quigley et al., 2009). To illustrate, one of the powerful ROS toolbox is Navigation Stack[1] which detects obstacles through 2D laser sensor in real time. Then, inflation layer is created with around obstacles by inflation radius value to avoid the collision. The idea is to use a small inflation radius value that allow pick and place mobile robot to close enough to its targets that are installed in selective racks. Additionally, coworkers safety area is manipulated as a circular obstacle fused in laser scan data. The final achievement of this research is a global architecture keeping a balance between real-time constraints leading to investment and operational costs and societal issues to make the human working conditions more comfortable.

This paper is structured as follows: Section 2 describes the proposed sensor interaction algorithm for human-aware mobile robot navigation in logistics warehouses. Section 3 presents experimental results demonstrating the advantages of the proposed framework. Finally, conclusions are drawn in Section 4.

2 HUMAN-AWARE NAVIGATION FRAMEWORK

2.1 Mobile Robot Navigation

ROS is a compilation of tools, libraries, and utilities which facilitate the design of complex robot tasks through powerful algorithms implementation such as Navigation Stack which is composed of several algorithms (ROS nodes). These nodes are communicating via publish/subscribe message-oriented middleware. For each ROS node, information is sent/received via a given topic as a structured data message. Consequently, a variety of information patterns can be ex-
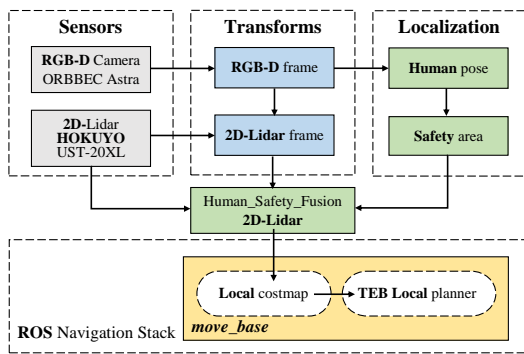
---

[1]ROS navigation, http://wiki.ros.org/navigation.

Figure 2: Conceptual overview of the system architecture based on ROS Navigation Stack. Safety fusion is discussed in Algorithm 1.

Figure 3: Block diagram describing input/output data and the main operations performed by human localization method.

changed between different information sources and manipulated by robust algorithms.

Actually, ROS Navigation Stack provides generic autonomous navigation via path planning and SLAM system. This allows robot to localize obstacles around it, plan a path to avoid collision, and navigate in unknown or partially known environments. Therefore, online planning is needed to create a collision free trajectory over the global trajectory toward the goal w.r.t trajectory optimization. So that, (Rösmann et al., 2017) policies are used for on-line trajectory planning which is based on Timed-Elastic-Band (TEB) approach. That is called teb-local planner[2] which combines temporal information to reach the goal. This method based on odometry and laser scan data requires only low computational resources. Conceptual overview of the system architecture is shown in Figure 2. based on *move_base*[3] node in ROS Navigation stack.

2.2 Human Detection and Localization

To localize humans, we implement depth-based human detection method using the Point Cloud Library (PCL) and RGB-D data (Munaro et al., 2016) as illustrated in Figure 3. Initially, to obtain a real time performance, the 3D points cloud data is reduced by voxel grid filter, where each voxel is scaled down by threshold size close to its centroid coordinates. This downsize operation reduces voxel size to 0.06 m of its real size that is sufficient to perform the detection process.

The ground plane (GP) is estimated and removed from the obtained voxel grid. The RANSAC-based least square method is used at each frame to compute the GP equation coefficients and update it with
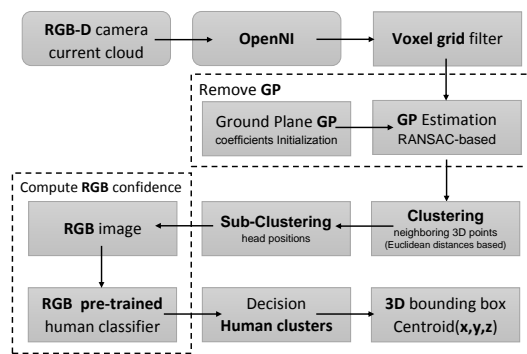
respect of the previous frame estimation. The remaining 3D points are clustered in 3D boxes based on their Euclidean distances. Then, these 3D boxes are segmented into sub-clusters to center them on peaks of a height map which contains threshold distance of the points from GP. These peaks are representing the human's head positions and its 2D RGB image is representing region of interest (ROIs). Instantly, the RGB image exhibiting ROIs is used by RGB-based human detector to compute RGB confidence index for the obtained sub-clusters. From the confidence index, we can select true positive human detection and eliminate false detection. At this stage, two methods are investigated:

- A usual machine learning method: HOG-based human detector with the recommended training configurations by (Dalal and Triggs, 2005) based on support vector machine (SVM).

- A recent deep neural network method based on reinforcement computation implemented in CPU or GPU: YOLO-based human detector by (Redmon and Farhadi, 2018) in its open source Darknet[4] framework with 53 convolutional layers for features extraction.

In experiments, a pre-trained model (yolov3-tiny based on COCO dataset) is used, which is characterised by small model. Then, each human-detector method is tested individually and the obtained results are discussed in more details in results section. An example for the extended code to output markers for visualization of ROIs and their RGB confidence score is given in Figure 4. Finally, the detected humans poses are published as ROS topic to be used further with human safety area generation.

---

[2]teb-local-planner, http://wiki.ros.org/teb_local_planner.
[3]ROS move_base, http://wiki.ros.org/move_base.

---

[4]Darknet, https://pjreddie.com/darknet/

(a) Depth image　　(b) 3D boxes　　(c) Sub clusters

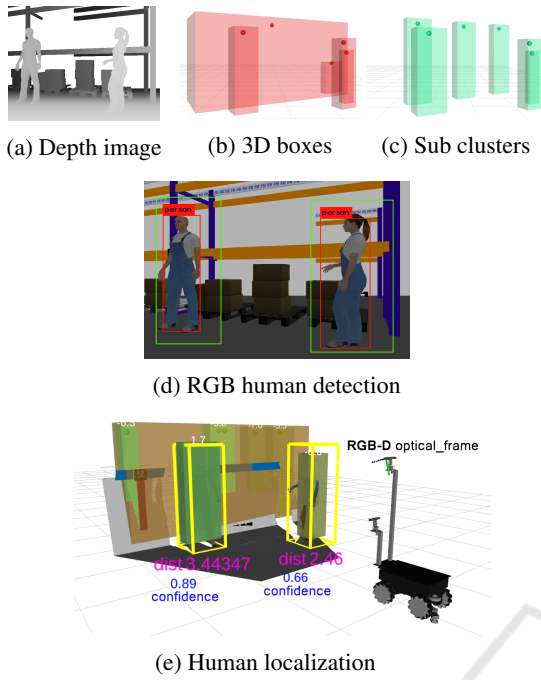(d) RGB human detection

(e) Human localization

Figure 4: Example of output markers for visualization of ROIs in RViz: First row depth image raw, 3D boxes for clusters and sub-clusters, second row in RGB detection red color rectangle for YOLO detector, green color rectangle for HOG detector. Third row, localized human (yellow 3D boxes).

2.3 Human Safety Area (HSA) Fusion

The navigation stack needs to know the position of sensors, wheels, and joints in respect to robot base. This is done using ROS Transform Frame (TF) package. The located human position coordinate is transformed from RGB-D frame to 2D LIDAR sensor frame for further human safety area calculations.

In warehouse environment, the mobile manipulator needs to come close to the pallets to pick up objects on the pallets. So the parameter for the security distance used to avoid collision between robots and pallets is set low. In the meantime, humans must not be scared by robots and the proxemics theory gives a higher value for the optimal distance between robots and humans (1.2 m). Once the distinction between human and non-human obstacle is done, the fusion process 2D laser plane must take this constraint into account. The HSA is virtually inserted as an obstacle in the warehouse environment and merge with the signal coming from the 2D lidars. This new signal and associated occupancy grid is then used by the robot local planner to navigate safely and human integrity in the warehouse.

The tested 2D-LIDAR is of type Hokuyo UST-

Algorithm 1: Human Safety Fusion.

1　HumanDetector (rgb.img , depth.cloud);
2　2D Laser Scan; //scan.msg;
3　**while** *HumanDetections(x , y , z)* **do**
4　　Calculate HSA $\leftarrow$ (x, y, z);
5　　//Generate 2D laser data message;
6　　(safety.msg) $\leftarrow$ HSA ;
7　**end**
8　//Async Spinner;
9　obstacle.msg $\leftarrow$ (scan.msg , safety.msg);

20LX at a height of 0.12 m with 20 m detection range, and 270° wide detection angle $[-135° : 135°]$. The start laser ray angle $\phi_\circ = -135°$ forward along the X axis in laser frame, alongside high angular resolution $\Delta = 0.25°$ (angle between 2 beams) with measurement discrete steps ($\eta_i, i = [1 : 1081]$) and 25 msec high speed response. To sum up, the laser ray angle $\phi_i \in [\phi_\circ + \eta_i \cdot \Delta]$. As shown in Figure 5a, suppose the HSA is represented as a Virtual Safety Circle (VSC) centroid with detected human coordinates $C = \begin{pmatrix} x_c \\ y_c \end{pmatrix}$ and radius r. The goal is to find all the range μ_i and intensity for each angle θ_i of index i ($i \in 1 : 1081$) which is done using the following calculations:

First, we determine the angle θ_C, the θ_P minimum, and θ_N maximum angles of the $\overparen{PN}$ arc corresponding to intersection points P, N displayed on Figure 5a.

$$\theta_c = \arctan 2(x_c, y_c) \ , \ \eta_c = E\left[\frac{\theta_c - \phi_0}{\Delta}\right] \quad (1)$$

$$\theta_P = \min(\theta_c + \psi, \theta_c - \psi) \ , \ \eta_P = E\left[\frac{\theta_P - \phi_\circ}{\Delta}\right] \quad (2)$$

$$\theta_N = \max(\theta_c + \psi, \theta_c - \psi) \ , \ \eta_N = E\left[\frac{\theta_N - \phi_\circ}{\Delta}\right] \quad (3)$$

where $\psi = \arccos(r/d)$. Let $\omega_i = \theta_C - \theta_i$ and μ_i, θ_i respectively the range and angle associated with Q_i (see Figure 5b). Applying Al Kashi theorem in triangle LQ_iC:

$$r^2 = \mu_i^2 + d^2 - 2d \cdot \mu_i \cdot \cos(\omega_i) \quad (4)$$

The resolution of this quadratic equation leads to two solutions and the minimum is chosen to be μ_i. Thus,

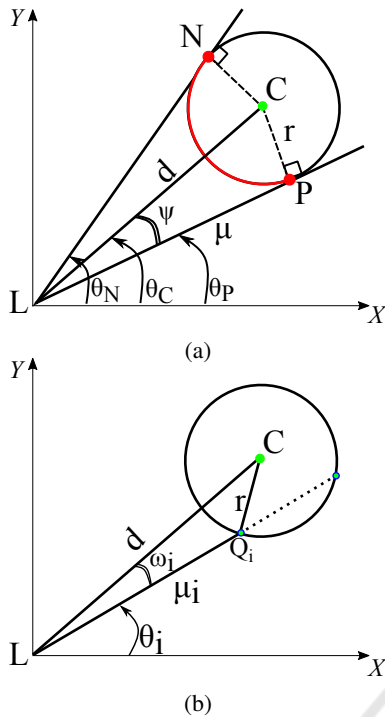$$Q_i = \mu_i \begin{bmatrix} \cos \theta_i \\ \sin \theta_i \end{bmatrix} \quad (5)$$

Figure 5: Calculations of the arc HSA in 2D-LIDAR coordinates system.

The Async Spinner[5] is utilized to merge the generated circular obstacles to 2D LIDAR sensor as described in human safety fusion algorithm 1. Whereas, a delayed message is controlled in custom time stamp queue and empty message is eliminated to keep track to the detected obstacles.

3 EXPERIMENTS AND RESULTS

The performance and capabilities of the proposed system are tested using a PC with Intel Core i7-6700HQ CPU 2.60 GHz, 32 GB of RAM and Nvidia Quadro GPU M3000M with 4 GB of memory and 1024 CUDA cores under Ubuntu 16.04 (Xenial) with ROS Kinetic. Several experiments are designed to reinforce the demands invoked in the introduction. First, human detection is investigated and evaluated, then the evaluation of the safety navigation approach is focused on avoiding obstacles. Testing the navigation system is carried out in logistic warehouse simulation environment using Summit XL mobile robot with Orbbec Astra RGB-D camera mounted on top of the robot at 1.63 m and with Hokuyo laser range finder at 0.12 m from ground plane as given in Figure 6.

---

[5]Async Spinner, http://wiki.ros.org/roscpp/Overview/Callbacks and Spinning

3.1 Human Detection Evaluation

Detection-based metrics (Goutte and Gaussier, 2005) are utilized for evaluating human detectors. These metrics include Accuracy, Precision, Recall, F1-Measure and FP-Rate which are depending on the number of correctly detected (True Positive: TP), falsely detected (False Positive: FP), missed detected objects (False Negative: FN) and true rejected object (True Negative: TN). The following performance indices are considered :

- Precision of predictions $= TP/(TP+FP)$

- Recall $= TP/(TP+FN)$

- Accuracy $= ((TP+TN)/(TP+TN+FP+FN))$

- F1-Measure $= 2 \times \frac{precision \times recall}{precision+recall}$

- FP-Rate $= FP/(FP+TN)$

For human detection in 2D laser range data, we compare between different trained classifiers on laser features described in (Linder et al., 2016) using three supervised learning techniques: Adaboost, SVM and the random forest (with 15 trees and maximum depth of 10). Comparatively to the recorded results for human detection which addressed a highly performance in (Linder et al., 2016) due airport environment dataset given in Table 1. Different from the reported result of the same pre-trained human classifiers in logistics environment. Unfortunately, human detection by 2D laser range data is ruled out, where it scores high false alarms and missed detection in respect to the appearance similarity between human leg features and logistics warehouse equipment such as selective racks and pallets at ground plane. 2D-LIDAR features sample for false alarm and missed detection results are shown in Figure 7. In details, false alarms by euro pallet are formatted with a 2D laser features most similarly to the human leg features when coworker is in standing state; where both of them are approximately sharing some features (Circularity, Radius and number of reflected points). Additionally, when the
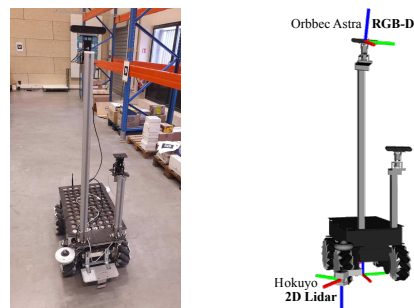

Figure 6: Summit XL robot and its URDF model with TF in RViz.

jump distance value for 2D laser data segmentation is chosen with high value to ignore the false alarms by the selective racks, a highly score of missed detection is found for the coworker in walking state. Under those circumstances, severity rate of human damage is increased. The performance evaluation of 2D laser-based human detector in logistics warehouses simulation is shown in Table 2. Each single experiment is run at least 3 times and detection metrics have been averaged to ensure stable results for each classifier.

For evaluating the performance of human detector in RGB-D data, the classical supervision machine learning technique are compared with deep learning in two computation scenarios: the first scenario uses CPU computation only, the second scenario uses CPU and GPU computation. Due the classical supervision machine learning technique, a pre-trained HOG descriptor is used for human with svm classifier. On the other hand, a pre-trained model (yolov3-tiny) is used based on COCO dataset. Nevertheless, both HOG-based and YOLO-based detectors achieve high detection rates as reported in Table 3.

The YOLO-based detector scores very slow detection time 1.961 (sec) per frame under CPU calculations only. Conversely, it is run in real time only by GPU computation with 53 Hz, very fast detection time 0.008 (sec) per frame and scores less FP-Rate 0.018. While, the HOG-based human detector is run in real time at 18 Hz on CPU with average detection time 0.039 (sec) per frame and score a few FP-rate 0.154, also the HOG detector based on CUDA implementation is run at 34 Hz and fast detection time 0.021 (sec) per frame. Under these circumstances, non-expensive GPU is useful to obtain high speed human perception, but, when mobile robot platform is not equipped with GPU, the HOG-based human perception is adequate in real time.

3.2 Navigation Simulation

Logistics warehouse environment simulation is designed under Gazebo simulator where standard parallel racks are set up at 3.0 m distances of each other. For mapping and simulation experiments, we have created a map by ROS Hector[6] SLAM (Kohlbrecher et al., 2011) with two costmaps configurations: global costmap and local costmap. In the local costmap, a 8×8 rolling window covers 4 m around mobile robot and 0.05 m inflation radius for obstacles in inflation layer. Besides, local trajectory planning parameters are tuned with optimal parameters allowing the robot to maneuver in narrow aisles and let human workers pass through. Then, the proposed method is tested using 2D navigation goal tool in RViz with avoiding obstacles using laser human safety fusion messages (see Figure 8). Figure 9 shows how the mobile robot avoid standing and walking coworkers using the proposed HSA around their localization in logistics warehouses simulation[7].
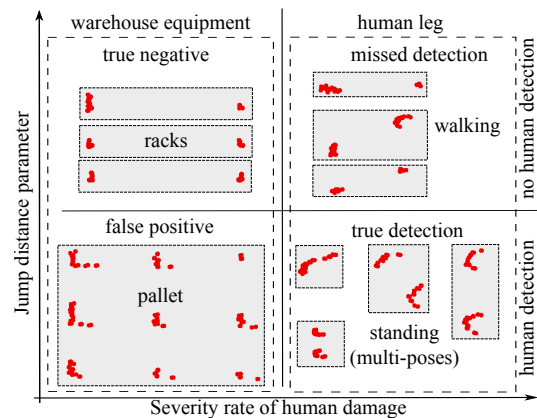

Figure 7: 2D LIDAR features samples for logistics warehouse equipment and human leg with strip rectangle for each object, first column is for logistics warehouse equipment (selective racks, euro pallet), while second column is for human leg (walking and standing states).

4 CONCLUSIONS

In this paper, we study a difficult problem namely human aware navigation in warehouses where research results for human detection do not give the expected good results. To overcome this problem, we proposed a new strategy based on a 2-steps method: the first step is to use the depth signal for clustering and identifying 3D boxes likely to be humans; while, the second step is to compute a human presence confidence index from the RGB signal. For the second step, several methods are compared and found solutions able to be run in real time on CPU with the HOG based method or on GPU with deep learning methods. The choice of the second step method depends on the available resources, but at least in the cheapest configuration, one method can solve the problem. The last contribution consists of fusing the data coming from the human detection algorithm and the 2D laser signal to create a 2D map enabling the robot to navigate not only safely but also human friendly, i.e., taking into account the proxemics theory optimal distances between robots and humans. The experiments in the simulation environment proved that the method can improve greatly the robot behaviors when crossing

---

[6]Hector SLAM, http://wiki.ros.org/hector_slam

[7]Video, https://vimeo.com/325696517

Table 1: Quantitative results for human detection in the 2D laser range data (Linder et al., 2016).

Method	Accuracy	Precision	Recall	F1-Measure	FP-Rate
SVM	0.914	0.741	0.615	0.672	0.031
Adaboost	0.969	0.895	0.846	0.870	0.014
Random Forest	0.962	0.906	0.766	0.830	0.011

Table 2: Quantitative results for human detection in the 2D laser range data in warehouse simulated environment.

Method	Accuracy	Precision	Recall	F1-Measure	FP-Rate
SVM	0.310	0.221	0.249	0.235	0.636
Adaboost	0.440	0.285	0.497	0.363	0.588
Random Forest	0.291	0.214	0.331	0.260	0.733

Table 3: Quantitative results for human detection by RGB-D data in warehouse simulated environment.

Method	GPU	detection time (sec)	Hz	Accuracy	Precision	Recall	F1-Measure	FP-Rate
HOG	Off	0.039	18.82	0.718	0.664	0.619	0.640	0.154
HOG	On	0.021	34.57	0.741	0.761	0.672	0.713	0.136
YOLO	Off	1.961	5.29	0.966	0.906	0.813	0.856	0.021
YOLO	On	0.008	53.80	0.974	0.912	0.830	0.869	0.018

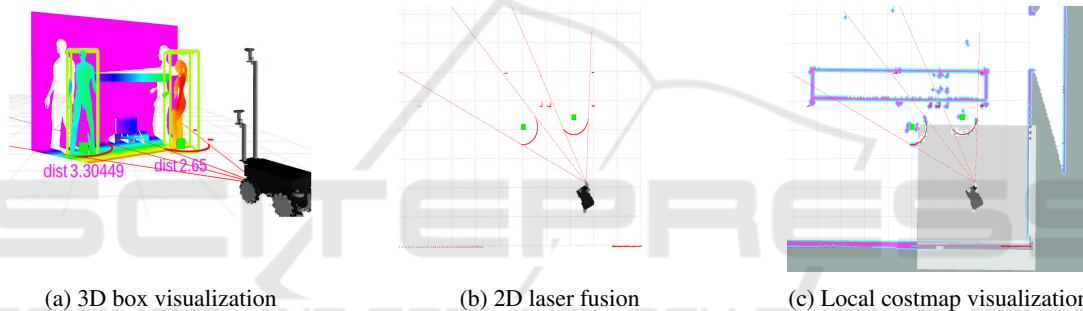(a) 3D box visualization   (b) 2D laser fusion   (c) Local costmap visualization

Figure 8: Visualisations of 2D laser fusion and local costmap outputs using RViz for logistics warehouse simulation.

human coworker in logistics warehouses. In future, the proposed framework will be tested on real robot platform follow through compress human localization using deep neural network models for reducing memory and computational requirements and allowing algorithm implementation on the CPU only.

ACKNOWLEDGEMENTS

REFERENCES

Arras, K. O., Mozos, O. M., and Burgard, W. (2007). Using boosted features for the detection of people in 2d range data. In *IEEE International Conference on Robotics and Automation*, pages 3402–3407. IEEE.

Choi, S., Kim, E., Lee, K., and Oh, S. (2017). Real-time nonparametric reactive navigation of mobile robots in dynamic environments. *Robotics and Autonomous Systems*, 91:11–24.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE International Conference on Computer Vision & Pattern Recognition*, volume 1, pages 886–893. IEEE.

Goutte, C. and Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European Conference on Information Retrieval*, pages 345–359. Springer.

Kohlbrecher, S., Von Stryk, O., Meyer, J., and Klingauf, U. (2011). A flexible and scalable slam system with full 3d motion estimation. In *IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 155–160. IEEE.

Krajník, T., Cristoforis, P., Kusumam, K., Neubert, P., and Duckett, T. (2017). Image features for visual teach and repeat navigation in changing environments. *Robotics and Autonomous Systems*, 88:127–141.

Linder, T., Breuers, S., Leibe, B., and Arras, K. O. (2016). On multi-modal people tracking from mobile platforms in very crowded and dynamic environments. In

(a) Green arrow for target (f/t=0s)   (b) Localize human safety (f/t=5s)   (c) Wait human to pass (f/t=10s)

(d) Change local plan (f/t=20s)   (e) Keep safety crossing (f/t=25s)   (f) Reach to target (f/t=45s)
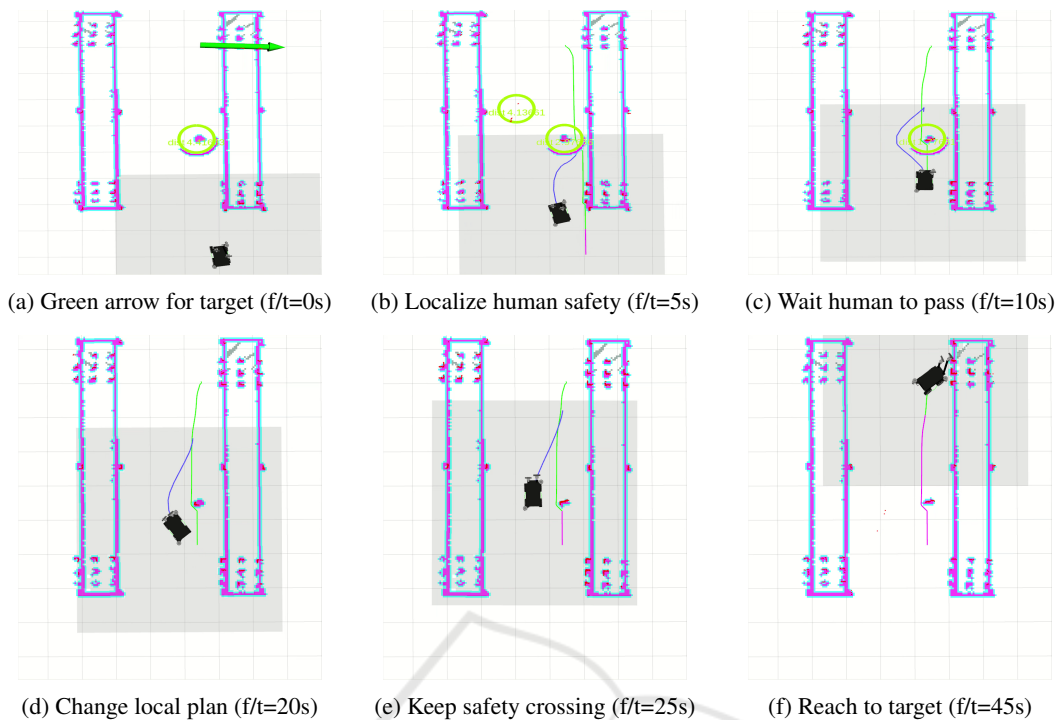
Figure 9: Snap-shots of visualisations for robot navigation in simulation experiment using RViz (local planner: blue line, global planner: green line, human safety: green circle (1.2 m radius), local costmap: light gray).

IEEE International Conference on Robotics and Automation, pages 5512–5519. IEEE.

Linder, T., Griesser, D., Vaskevicius, N., and Arras, K. O. (2018). Towards accurate 3D person detection and localization from RGB-D in cluttered environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Workshop on Robotics for Logistics in Warehouses and Environments Shared with Humans*.

Lindner, F. and Eschenbach, C. (2017). An affordance-based conceptual framework for spatial behavior of social robots. In *Sociality and Normativity for Robots*, pages 137–158. Springer.

Malone, N., Chiang, H.-T., Lesser, K., Oishi, M., and Tapia, L. (2017). Hybrid dynamic moving obstacle avoidance using a stochastic reachable set based potential field. *IEEE Transactions on Robotics*, 33(5):1124–1138.

Munaro, M., Lewis, C., Chambers, D., Hvass, P., and Menegatti, E. (2016). RGB-D human detection and tracking for industrial environments. In *Intelligent Autonomous Systems (IAS-13)*, pages 1655–1668. Springer.

Nardi, L. and Stachniss, C. (2017). User preferred behaviors for robot navigation exploiting previous experiences. *Robotics and Autonomous Systems*, 97:204–216.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., and Ng, A. Y. (2009). ROS: An open-source robot operating system. In *ICRA Workshop on open source software*.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Rösmann, C., Hoffmann, F., and Bertram, T. (2017). Integrated online trajectory planning and optimization in distinctive topologies. *Robotics and Autonomous Systems*, 88:142–153.

Song, K.-T., Jiang, S.-Y., and Wu, S.-Y. (2017). Safe guidance for a walking-assistant robot using gait estimation and obstacle avoidance. *IEEE/ASME Transactions on Mechatronics*, 22(5):2070–2078.

Wahrmann, D., Hildebrandt, A.-C., Schuetz, C., Wittmann, R., and Rixen, D. (2017). An autonomous and flexible robotic framework for logistics applications. *Journal of Intelligent & Robotic Systems*, 93:419–431.

Yang, H., Fan, X., Shi, P., and Hua, C. (2016). Nonlinear control for tracking and obstacle avoidance of a wheeled mobile robot with nonholonomic constraint. *IEEE Transactions on Control Systems Technology*, 24(2):741–746.

Zimmermann, C., Welschehold, T., Dornhege, C., Burgard, W., and Brox, T. (2018). 3D human pose estimation in RGBD images for robotic task learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1986–1992. IEEE.