

MIM: Indoor and Outdoor Navigation in Complex Environments Using Multi-Layer Intensity Maps

Adarsh Jagan Sathyamoorthy^{1*}, Kasun Weerakoon^{1*}, Mohamed Elnoor¹, Mason Russell², Jason Pusey², and Dinesh Manocha¹.

Abstract—We present MIM (Multi-Layer Intensity Map), a novel 3D object representation for robot perception and autonomous navigation. MIMs consist of multiple stacked layers of 2D grid maps each derived from reflected point cloud intensities corresponding to a certain height interval. The different layers of MIMs can be used to simultaneously estimate obstacles’ height, solidity/density, and opacity. We demonstrate that MIMs’ can help accurately differentiate obstacles that are safe to navigate through (e.g. beaded/string curtains, pliable tall grass), from ones that must be avoided (e.g. transparent surfaces such as glass walls, bushes, trees, etc.) in indoor and outdoor environments. Further, to handle narrow passages, and navigate through non-solid obstacles in dense environments, we propose an approach to adaptively inflate or enlarge the obstacles detected on MIMs based on their solidity, and the robot’s preferred velocity direction. We demonstrate these improved navigation capabilities in real-world narrow, dense environments using a real Turtlebot and Boston Dynamics Spot robots. We observe significant increases in success rates to more than 50%, up to a 9.5% decrease in normalized trajectory length, and up to a 22.6% increase in the F-score compared to current navigation methods using other sensor modalities.

I. INTRODUCTION

Mobile robots have been used to navigate in indoor environments (such as households, offices, hospitals, etc. [1], [2], [3]), and outdoor environments such as agricultural fields, forests, etc. [4], [5], [6]. Such complex environments contain obstacles of various sizes, densities/solidities, and opacities that are challenging in terms of the robot’s perception, and navigation. For instance, contemporary indoor environments contain objects such as string/beaded curtains, transparent surfaces such as glass walls [7], [8], etc. Outdoor scenarios, on the other hand, have complex vegetation such as pliable tall grass, bushes, trees, etc. in close proximity to and intertwined with each other. A major challenge, and an important requirement for autonomous navigation, is differentiating *truly solid* and impassable obstacles (furniture, glass surfaces, bushes, trees, etc.) from obstacles that can be passed through (beaded curtains, tall grass, etc.).

To first detect obstacles, mobile robots have predominantly used RGB and depth images [9], 2D lidar scans [10], 3D point clouds [11], etc. The raw data from these sensors has been used to: (1). Estimating the proximity of objects in the robot’s vicinity; (2). Computing a variation of an occupancy grid [12] or a cost map representation that indicates both an obstacle’s size and distance from the robot; or (3). Segmenting the scene to assess obstacles’ size, distance, and

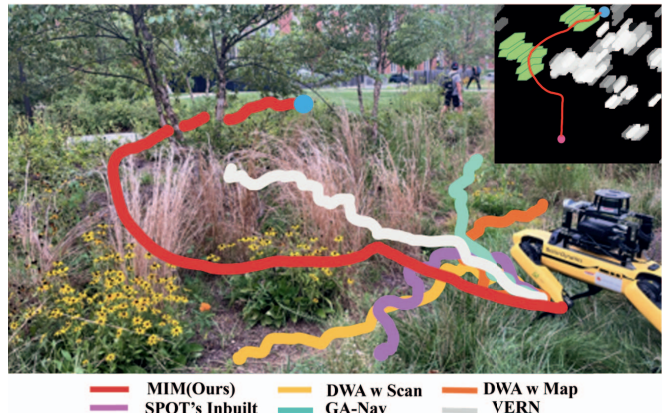


Fig. 1: Comparison of trajectories while navigating using our Multi-layer Intensity Maps (MIM), DWA with laser scan [10], DWA with occupancy map [12], Spot’s Inbuilt Autonomy, and VERN [16] in complex vegetation. The MIM for the scenario is shown in the top right, with the robot in pink, its goal in blue, passable obstacles such as tall grass in green, and the robot’s trajectory overlaid for reference. MIMs help differentiate solid objects such as trees even when they are intertwined with tall grass. Other methods for robot perception in this case either freeze or collide with the solid obstacles.

semantic meaning [13], [14]. Although such approaches have been used to aid navigation, they may not work well in environments composed of thin, pliable/bendable, and transparent obstacles. For instance, time-of-flight sensors such as depth cameras and lidars tend to detect thin, pliable, and passable obstacles (e.g. string curtains, tall grass) as solid obstacles that the robot must avoid [15]. On the other hand, transparent objects such as glass remain undetected since the laser rays mostly pass through them, leading to collisions during navigation. Similarly, perception methods using RGB images may not work well in terms of detecting transparent objects and visually similar but structurally dissimilar vegetation. Moreover, they are severely affected by the environment’s lighting changes.

Main Contributions: In order to address these robot perception challenges, we present a novel obstacle representation called MIM, a *multi-layer intensity map* that can be used to simultaneously estimate the size/height, density/solidity, and opacity of objects in the environment. The MIMs are constructed by stacking individual layers of 2D grid maps, each computed from reflected point cloud intensities corresponding to a certain height interval. It preserves the benefits of existing occupancy grids such as indicating the size and proximity of obstacles around the robot while also accurately detecting passable obstacles as such. The novel components of our approach include:

- A novel obstacle representation called the multi-layer

* Authors contributed equally. Full technical report and video are available at <http://gamma.umd.edu/mim/>.

This research was supported by Army Cooperative Agreement W911NF2120076, ARO grants W911NF2310046 and W911NF2310352. We acknowledge the support of the Maryland Robotics Center.

¹ Authors are with the University of Maryland, College Park.

² Authors are with the DEVCOM Army Research Laboratory, Aberdeen Proving Ground, Maryland, USA.

intensity map constructed from the intensities of the reflected points in a point cloud. In addition to an object’s height and size, the multi-layer intensity map also reflects its true density/solidity. Our multi-layer intensity map can replace occupancy grids and other map representations in existing navigation methods to enable a robot to navigate through passable/navigable indoor and outdoor obstacles that are often misclassified by existing representations.

- A novel method to detect transparent objects using the low-intensity reflected points in the multi-layer intensity map. Our approach accurately extrapolates the transparent surfaces from a small neighborhood of low-intensity points which enables a robot’s motion planner to avoid collisions, significantly improving its rate of safely reaching its goal.
- A novel method using the multi-layer intensity map to accurately identify objects that are safe to navigate through such as thin, passable curtains in indoor scenarios, and pliable tall grass in outdoor environments. Our approach alleviates robot freezing behaviors in the presence of such objects.
- An adaptive inflation strategy that assesses the detected solid obstacles (e.g. concrete and glass walls, furniture indoors, and bushes and trees outdoors) to enlarge them in the multi-layer intensity map for efficient planning. Our strategy handles narrow scenarios such as doors, corridors, and passages between dense vegetation and trees, where existing navigation schemes freeze. We demonstrate significant improvements in navigation using MIMs in indoor scenes using a real Turtlebot, and in complex outdoor environments using a Boston Dynamics Spot robot.

II. RELATED WORK

In this section, we give a brief overview of perception methods that use point cloud intensities. Additionally, we review the existing obstacle detection research in indoor and outdoor settings.

A. Sensing using Point Cloud Intensity

The importance of intensity information in LiDAR point cloud data has been an emerging focus in robotics [17], [18], [19]. Some methods investigate the use of intensity information alongside geometric features to enhance point cloud classification methods in outdoor settings [20]. These methods highlight the potential of using the intensity information to provide a better understanding of obstacles, particularly when scene illumination is not consistent. Other methods include the ISHOT descriptor [21], which combines geometric and intensity data for improved place recognition. Lidar intensity maps have been also used for localization [22], [23], [24]. In [22], the authors present a robust Graph-SLAM framework that improves map accuracy for autonomous vehicles by encoding road surfaces based on LIDAR reflectivity. Moreover, the application of LiDAR intensity in visual navigation tasks has been explored. For instance, [25] introduces a lidar-intensity-image pipeline and demonstrates its performance in visual odometry (VO) and visual teach and repeat (VT&R) tasks. Lidar intensity maps have also been leveraged for various other applications, including orthoimage generation [26] and anisotropic surface detection [27].

B. Detecting Indoor Obstacles

Object detection in indoor settings has been widely studied for numerous applications including robot navigation, mapping, and computer graphics. Popular solutions in the literature include vision-based object detection and semantic segmentation approaches due to the structuredness of indoor environments. Moreover, the generation of necessary image datasets is feasible due to the limited diversity of indoor objects. However, detecting non-opaque objects such as glass remains a formidable challenge for vision-based systems due to the lack of visual clues. The method in [7] proposes GNet, a glass detection network that identifies abundant contextual cues for glass detection using a large-field contextual feature integration (LCFI) module. UCTNet [13] proposes a cross-modal transformer network for indoor RGB-D semantic segmentation to identify different objects such as curtains, doors, etc. However, such methods require large datasets with pixel-level ground truth labeling.

C. Detecting Outdoor Obstacles

Over recent years, there has been significant progress in the development of robotic systems designed for outdoor navigation [14], [28], [29], [30], [31]. One early approach can be found at [32], where the usage of laser measurements enabled navigation capabilities for robots such as detecting short and grass-like vegetation. However, this method is not universally applicable, particularly in complex, unstructured vegetative terrains. Complementary approaches have tackled associated issues in off-road navigation, specifically concerning varying slopes [33], [34] and different terrains [35].

Many studies integrate proprioceptive with exteroceptive sensory data to enhance outdoor navigation [36], [37], [38]. Machine learning techniques have also been incorporated to augment the robot’s capabilities for navigating through pliable vegetative obstacles [39], [40]. To this end, [16] uses a few shot-learning approaches to classify RGB vegetation images based on their traversability. This classifier is then integrated with a 3D LiDAR to construct a vegetation-aware traversability cost map.

III. BACKGROUND

A. Definitions and Assumptions

Our formulation assumes that a sensor capable of generating 3D point clouds (e.g., 3D lidar, depth camera) is mounted on a robot with a 2D linear (v) and angular (ω) velocity space. Rigid coordinate frames are attached to the robot and sensor with the positive x, y, z directions facing forward, left, and upwards respectively, and for simplicity, we assume both frames to coincide. All positions and velocities are measured relative to these frames. At any time t , the robot has a preferred velocity direction aimed at its goal (g_x, g_y) as $\theta = \tan^{-1}(g_y/g_x)$.

Our approach is based on using the reflected intensities of point clouds that could be obtained from sensors such as 3D lidars, depth cameras, etc that have a laser source/transmitter and a receiver. We represent a point \mathbf{p} in a point cloud as $\mathbf{p} = \{x, y, z, int\}$, where x, y, z denote the point’s location relative to the sensor, and $int \in [0, R]$ denotes its intensity. We define a 2D robot-centric grid map as containing $n \times n$ grids. Each grid is denoted by a row r and column c . Each grid represents a $g \times g$ area in the real world, and the value

contained in it indicates the probability of the presence of an obstacle. Finally, we use j and k to denote indices.

B. Point Cloud Intensity

Typically, a point's intensity is high ($int > 0.75R$) when it is reflected from solid, opaque, 3D (length, width, and height dimensions are not infinitesimal) objects since they prevent the sensor's laser rays from passing through, or scattering away from the sensor. In contrast, objects that are low density (e.g. tall grass which is a collection of thin blades of grass that scatter laser rays), and transparent (e.g. glass where laser rays mostly pass through) lead to low intensities ($int < 0.5R$) or in some cases, no intensity ($int = 0$).

C. Obstacle Properties

Our formulation leverages the property to accurately detect truly solid objects from the following categories defined based on how objects are sensed by existing perception modalities (2D lidar scan, RGB and depth images, etc) as:

- **True Positives (TP):** Solid, non-traversable objects detected as solid, e.g. walls, wooden furniture, etc.
- **True Negatives (TN):** Non-solid, traversable objects detected as passable or no obstacle, e.g. free space.
- **False Positives (FP):** Non-solid objects detected as solid, e.g. string/beaded curtains, pliable tall grass.
- **False Negatives (FN):** Solid objects detected as non-solid or as free space, e.g. transparent objects.

D. Obstacle Inflation

Once a truly solid object is detected, it must be enlarged or *inflated* for the robot's planner to ensure that it avoids it at a safe distance [41]. Inflation is performed prior to planning to expand obstacles uniformly in all dimensions by a certain amount (typically the robot's radius, or maximum(length, width) to ensure that the planner avoids obstacles by a safe distance. Standard methods for obstacle inflation include performing the Minkowski sum [42] between the robot's radius and the obstacle, cost propagation from the obstacle, dilating obstacles using convolutions, etc. on a grid map. With these preliminaries, we state our problem formulation as follows:

Formulation III.1. *To construct an $n \times n \times m$ grid map representation I_{ML}^t of obstacles from points $\mathbf{p} = \{x, y, z, int\}$ and classify each grid $(r, c) \in I_{ML}^t$ as a true positive (TP), false positive (FP), false negative (FN) obstacle, or true negative (TN) free space and enlarge TP and FN obstacles adaptively based on the robot's preferred velocity direction.*

IV. OUR APPROACH

In this section, we discuss how 3D point cloud intensities can be used to construct the multiple 2D grid map layers of an MIM. The input point clouds could be obtained from any sensor such as a 3D lidar or a depth camera that also measures the intensity of the points reflected from surrounding objects. We show how different layers of the intensity map can be used to accurately differentiate solid obstacles from passable objects, and identify transparent obstacles. Finally, we detail our obstacle inflation strategy, which enables robots to navigate through passable obstacles and narrow passages. Fig. 2 shows our overall proposed architecture.

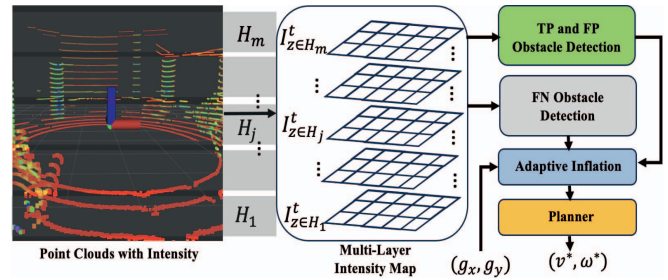


Fig. 2: Our approach's overall system architecture. At time instant t , the intensities of reflected point clouds from a height interval H_j (grey rectangles) are used to construct a 2D grid map layer $I_{z \in H_j}^t$ according to equation 1. Several of such layers are stacked to form an MIM. Certain layers from MIMs can be used to detect TP, FP, and FN obstacles. The adaptive inflation enlarges the truly solid (TP, FN) obstacles that the robot must avoid in the direction opposite to the robot's goal (g_x, g_y) direction, and ignores passable obstacles (FP). The planner finally uses the inflated map to compute collision-free linear and angular velocities v^*, ω^* .

A. Multi-Layer Intensity Map

We construct a single layer of the intensity map corresponding to a height interval H_j at any time instant t as follows,

$$I_{z \in H_j}^t(r, c) = \frac{\sum_x \sum_y int}{g^2} \quad (1)$$

$$\forall x \in [x_{low}, x_{low} + g], y \in [y_{low}, y_{low} + g],$$

$$x_{low} = \lfloor (r - \frac{n}{2}) \cdot g \rfloor \text{ and } y_{low} = \lfloor (c - \frac{n}{2}) \cdot g \rfloor$$

$$H_j = [low_j, high_j], low_j \leq high_j,$$

where $\lfloor \cdot \rfloor$ represents the floor operation, low_j , and $high_j$ are the limits for all the points whose intensities must be considered for the map along the z direction. Extending this definition, we construct a *multi-layer* intensity map as a stacking of m layers as,

$$I_{ML}^t(r, c) = [I_{z \in H_1}^t(r, c) | \dots | I_{z \in H_j}^t(r, c) | \dots | I_{z \in H_m}^t(r, c)], \quad (2)$$

where $H_1, \dots, H_j, \dots, H_m$ are non-overlapping height intervals. We choose to stack multiple non-overlapping layers at various heights instead of combining the points' intensities at all heights into a unified layer. This is due to the flexibility that multiple layers provide in analyzing and modifying them individually. Furthermore, individual layers can be combined after modification and used for planning the robot's trajectories. We highlight these benefits in the following sections.

B. Obstacle Detection

In this section, we describe how challenging obstacles such as tall grass (FP), string/beaded curtains (FP), and transparent objects (FN) can be detected using our multi-layer intensity map.

1) *Differentiating True and False Positive Obstacles:* Existing methods that use various sensor modalities typically detect many thin, pliable obstacles such as tall grass, and passable objects such as string/beaded curtains as solid obstacles that must be avoided. During navigation, such inaccurate detections cause the robot to freeze or oscillate perpetually without reaching its goal. We show how such

false positive obstacles can be detected using the multi-layer intensity map.

Let us consider three layers $I_{z=0}^t, I_{z \in (0,h]}^t, I_{z \in [-h,0)}^t$ of the intensity map. If a grid location (r, c) belonging to all three layers satisfies the following condition \mathcal{C} , we classify that grid as a false positive obstacle:

$$\mathcal{C}(r, c) : I_{z=0}^t(r, c), I_{z \in (0,h]}^t(r, c), I_{z \in [-h,0)}^t(r, c) \leq \Gamma. \quad (3)$$

Here, Γ is an intensity threshold. Using this condition, we construct TP and FP intensity maps for planning as,

$$I_{TP}^t(r, c) = \max(I_{z=0}^t(), I_{z \in (0,h]}^t(), I_{z \in [-h,0)}^t()) \quad \forall (r, c) \mid \mathcal{C}(r, c) \text{ is False} \quad (4)$$

$$I_{FP}^t(r, c) = \max(I_{z=0}^t(), I_{z \in (0,h]}^t(), I_{z \in [-h,0)}^t()) \quad \forall (r, c) \mid \mathcal{C}(r, c) \text{ is True} \quad (5)$$

2) *Detecting False Negative Obstacles:* False negative obstacles are typically transparent objects that allow most of the laser rays from a sensor to pass through. However, a small neighborhood of points with low intensities is detected for laser rays that are incident $\sim 0^\circ$ on a transparent surface only along the $z = 0$ plane [43]. Our approach extrapolates this small neighborhood to detect solid transparent obstacles.

Consider two layers $I_{z=0}^t$ and $I_{z=-\epsilon}^t$ in the multi-layer intensity map at time instant t . Here, ϵ is a small positive value. To isolate the low-intensity neighborhood of points reflected from the transparent object, we first calculate the element-wise difference between the two layers $I_{z=0}^t \ominus I_{z=-\epsilon}^t$. This removes all the points reflected from the same obstacles in both layers and retains only the points corresponding to the small glass neighborhood. To indicate the presence of transparent obstacles for subsequent time steps, we transform $I_{z=0}^t \ominus I_{z=-\epsilon}^t$ based on the robot's motion as,

$$I_{FN}^t = T \cdot (I_{z=0}^t \ominus I_{z=-\epsilon}^t), \quad (6)$$

where T is a 4×4 transformation matrix whose rotational component is based on the robot's yaw, and the translational component is based on its motion from time t to $t+1$. Finally, we augment subsequent time's I_{FN}^{t+1} using I_{FN}^t as,

$$I_{FN}^{t+1} = I_{FN}^{t+1} \cup I_{FN}^t. \quad (7)$$

C. Adaptive Obstacle Inflation

In narrow and dense scenarios, uniformly inflating obstacles (as explained in section III-D) could close the available free space (see fig. 3) leading to the robot freezing problem [44]. If obstacles are not inflated, the robot could move close to the obstacles and even collide with them. Additionally, false positive obstacles that can be traversed need not be inflated. Therefore, our approach adaptively inflates the obstacles based on the robot's goal direction. Obstacles are majorly inflated in the direction opposite to the robot's goal/preferred direction and minorly inflated in all other directions. This ensures that the robot does not navigate too close to a solid obstacle, while also never closing free space near the narrow passages.

Our inflation is performed using the convolution operation on obstacles by computing the appropriate kernel matrices



Fig. 3: The robot's position and its goal are denoted by the pink and blue circles respectively. The robot's heading and goal direction coincide in this case. [Left]: Uniformly inflating obstacles using an $e \times e$ kernel with all ones near a narrow passage closing up the available free space (green rectangle). [Right]: Adaptively inflating the obstacles based on the robot's goal direction preserves the free space while inflating the obstacles in the direction opposite to the robot's goal direction.

K of size $e \times e$ as follows. Let (g_x, g_y) be the goal location relative to the robot's coordinate frame. The goal direction can be defined by the slope $\tan \frac{g_y}{g_x}$. To design a kernel matrix K to inflate obstacles in the opposite direction, we first find the line along the goal direction relative to the kernel, passing through its center $(\frac{e}{2}, \frac{e}{2})$ as,

$$f(r^K, c^K) : c^K - \tan(g_y/g_x) \cdot r^K + \text{const} = 0 \quad (8)$$

$$\text{const} = \frac{e}{2} (\tan(g_y/g_x) - 1).$$

Here, (r^K, c^K) represent the row and column on the kernel. Next, the kernel can be constructed as,

$$K(r^K, c^K) = \begin{cases} 1 & \forall \{r^K, c^K \mid f(r^K + j, c^K + j) = 0\} \\ 0 & \text{Otherwise.} \end{cases} \quad (9)$$

Here, $j \in [-\text{padding}, \text{padding}]$ is added to the row and column indices to control the level of thickness to inflate an obstacle. Finally, using K , we convolve our multi-layer intensity map as,

$$I_{in\,flate}^t = (I_{TP}^t \cup I_{FN}^t) \otimes K. \quad (10)$$

$I_{in\,flate}^t$ contains inflated True Positive and False Negative obstacles. To add information about false positive obstacles prior to planning, we perform,

$$I_{plan}^t = I_{in\,flate}^t \cup I_{FP}^t, \quad (11)$$

where I_{plan}^t is a 2D grid map containing TP, FP, FN, and free space represented by various grids, which can be used as a cost map for motion planning.

D. Integration with Planning Methods

The final planning intensity map I_{plan}^t can be used with any motion planner as a cost map to evaluate a candidate linear and angular velocity's (v, ω) obstacle or collision cost. This can be computed by extrapolating the trajectory produced by (v, ω) relative to I_{plan}^t as in [35], [16] as,

$$\text{traj}^{I_{plan}^t} = [(r_1, c_1), \dots, (r_j, c_j), \dots, (r_p, c_p)]$$

$$\text{Obstacle Cost} = \sum_{j=1}^p I_{plan}^t(r_j, c_j). \quad (12)$$

Additionally, I_{plan}^t can be used as an observation with deep reinforcement learning-based navigation methods [33]

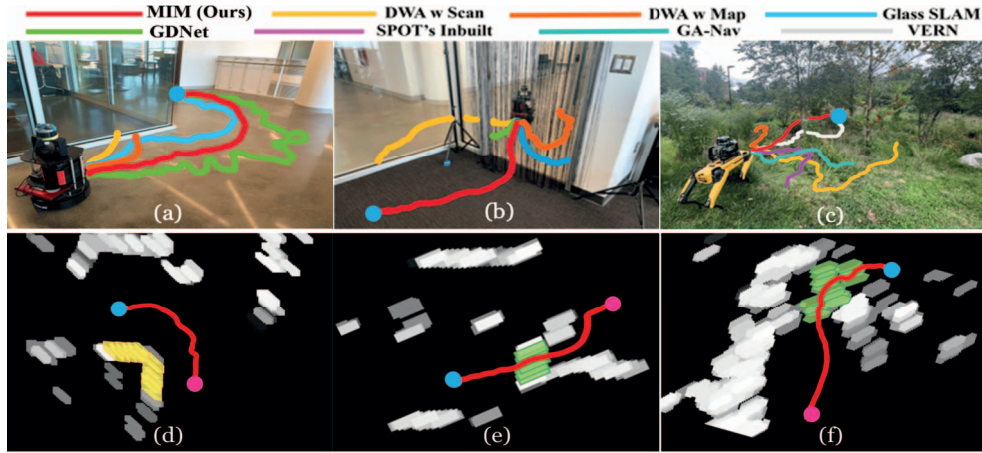


Fig. 4: Robot trajectories when navigating in different complex indoor and outdoor environments using various methods. (a) MIM identifies transparent objects such as glass in real-time and avoids it, (b) MIM identifies string curtains as safe, passable obstacles while other methods detect them as solid, and impassable. (c) MIM accurately identifies pliable tall grass regions (which register lower intensities on I_{plan}^t) to navigate through, avoiding trees. (d,e,f) show the corresponding I_{plan}^t for the scenarios above. White, grey, and black colors represent intensities in a decreasing order of magnitude. The robot’s starting location is in pink, and the goal is represented in blue. The yellow in (d) represents the glass (FN obstacle) extrapolated in real-time by our method in section IV-B.2. Green in (e, f) represents passable, non-solid obstacles such as curtains and tall grass respectively. MIM-based navigation’s trajectory is overlaid for reference. (d, e, f) also depict obstacles inflated based on the robot’s goal direction.

to improve the trained model’s understanding of the solidity of surrounding obstacles.

V. RESULTS AND ANALYSIS

In this section, we summarize our method’s implementation details and evaluation metrics. Then, we present the details of the experiments conducted to highlight the benefits of our approach.

A. Implementation

We use a Boston Dynamics Spot robot for outdoor experiments and a Turtlebot 2 robot for indoor experiments. The Spot robot is equipped with an Intel NUC 11 with Intel i7 CPU and NVIDIA RTX 2060 GPU. The Turtlebot is equipped with a laptop with an Intel i9 CPU and an Nvidia RTX 2080 GPU. Both robots are equipped and a Velodyne VLP16 lidar. We particularly use Velodyne lidar’s dual return mode since it captures low intensities in the point cloud better [43]. We use $n = 200$, and $m = 4$ layers in our implementation (one for $I_{z \in [-h, 0]}^t$, $I_{z \in (0, h]}^t$, $I_{z=0}^t$, $I_{z=-\epsilon}^t$). The height ranges are adjusted based on the robot’s height.

B. Comparison Methods and Evaluation Metrics

We compare the improvements in navigation while using MIMs with various existing methods. We use MIMs with the dynamic window approach (DWA) [10] and calculate candidate velocity’s obstacle costs as described in section IV-D. For a fair comparison, we use DWA as the baseline planner and integrate other methods for obstacle perception with it for navigation. We compare with various indoor and outdoor navigation methods that use a variety of perception inputs such as RGB/depth image, 2D lase scans, or occupancy grids created by any proximity sensor.

DWA is a local planner that can use a 2D LiDAR scan [10] or an occupancy map [12] to perform obstacle avoidance. Spot’s in-built autonomy incorporates a set of stereo cameras around the robot to estimate the obstacles and ground plane to navigate to a goal. VERN is a vegetation-aware navigation approach that uses an RGB image-based

vegetation classifier and a set of 2D occupancy grid maps for perception. In indoor scenarios, Glass-SLAM [45] uses the specular reflection of laser beams from the glass to map environments that include glass. GDNNet [46] is a semantic segmentation method that uses RGB images to segment glass from a scene. We also perform ablations studies between using adaptive and uniform inflation using MIMs.

Success Rate - Number of successful goal-reaching attempts (without collisions with solid objects or freezing behaviors) out of the total number of trials.

Normalized Traj. Length - The ratio between the robot’s trajectory length and the straight-line distance to the goal from the starting location averaged over all runs.

F-Score - A measure of object detection accuracy calculated as a weighted average of the precision and recall. Values are between 0 and 1, where 1 indicates the best accuracy. We use human detection of TP, FP, and FN obstacles to calculate precision and recall.

Inference Time - Time taken from the instant an obstacle is viewed to the instant when it is detected.

C. Testing Scenarios

- **Scenario 1** - Indoor scenario with glass, concrete walls, and pillars (Fig. 4a).
- **Scenario 2** - Indoor scenario with a narrow passage covered with a beaded/string curtain (Fig. 4b).
- **Scenario 3** - Outdoor scenario with tall grass, bushes, and trees separated from each other (Fig. 4c).
- **Scenario 4** - Outdoor scenario with tall grass, bushes, and trees closely intertwined with each other creating narrow passages (Fig. 1).

D. Analysis and Comparison

We present the qualitative navigation experiment results for the four scenarios in Fig. 4 and the quantitative results in Table I. Scenarios 1 and 2 are complex indoor settings, whereas scenarios 3 and 4 are outdoor settings. We observe that MIM demonstrates the highest success rate compared to the other methods in all four scenarios.

Scenario	Method	Success Rate (%)	Norm. Traj. Length	F-Score
Scn. 1	DWA with laser scan [10]	0	0.342	0.12
	DWA with occupancy map [12]	0	0.415	0.20
	Glass-SLAM [45]	30	0.895	0.76
	GDNet [46]	50	0.633	0.69
	MIM (Uniform)	80	1.256	0.82
	MIM (Adaptive)	80	1.052	0.82
Scn. 2	DWA with laser scan [10]	0	0.765	0.17
	DWA with occupancy map [12]	0	0.355	0.16
	Glass-SLAM [45]	0	0.421	0.46
	GDNet [46]	0	0.318	0.65
	MIM (Uniform)	30	1.134	0.79
	MIM (Adaptive)	70	1.141	0.79
Scn. 3	DWA with laser scan [10]	20	1.571	0.35
	DWA with occupancy map [12]	20	1.482	0.33
	Spot's Inbuilt Autonomy	10	0.311	0.38
	VERN [16]	70	1.154	0.75
	MIM (Uniform)	40	1.292	0.84
	MIM (Adaptive)	80	1.072	0.84
Scn. 4	DWA with laser scan [10]	0	1.543	0.37
	DWA with occupancy map [12]	0	1.412	0.36
	Spot's Inbuilt Autonomy	0	0.298	0.31
	VERN [16]	50	1.267	0.68
	MIM (Uniform)	60	1.248	0.74
	MIM (Adaptive)	70	1.146	0.74

TABLE I: Performance comparison between using MIMs for navigation versus other methods in indoor and outdoor scenarios using various metrics. We observe that MIMs are versatile in detecting a variety of perceptually challenging obstacles and aiding navigation.

Method	Inference Time (sec)	Training Time
MIM	0.020 on CPU	NA
Glass-SLAM	4.5 on CPU	NA
GDNet	0.112 on GPU	12-15 hours
VERN	0.083 on GPU	6-8 hours

TABLE II: The inference rates and training time (where applicable) for several methods that detect glass (Glass-SLAM [45], GDNet [46]), and pliable vegetation (VERN [16]). MIMs are capable of detecting transparent obstacles such as glass, and passable tall grass in real-time, faster than existing methods that use lidars and RGB images.

In scenario 1, 2D laser scan-based and occupancy map-based DWA planners fail to identify the glass region since they do not incorporate lidar intensity for object detection. GDNet-based indoor segmentation fails to identify the glass region from RGB inputs consistently due to lighting changes, and strong lights reflected from the floor. Hence, these methods lead to collisions with the glass by assuming it to be free space. The detecting glass in SLAM can identify the glass region using lidar intensity. However, it does not avoid glass all the time since the glass is constructed slower than the robot's motion. MIM's multi-layer map formulation leads to consistent glass detection which also results in a higher F-score compared to the other methods.

Scenario 2 includes a passable string curtain which is detected as an obstacle from the 2D laser scan and occupancy map-based DWA methods. Hence, these methods attempt to avoid the curtain and collide with the glass during navigation. Further, the glass detection SLAM method identifies both the curtain and the glass as obstacles. Hence, all three methods demonstrate poor success rates in scenario 2. GDNet identified both the open door and the glass near it as glass preventing the robot from entering through the door. In contrast, our MIM demonstrates a significantly high success rate irrespective of the lighting conditions due to the 360° LiDAR-based intensity map formulation.

Scenarios 3 and 4 include traversable tall grass and non-traversable bushes and trees. However, the two DWA methods identify all such vegetation regions as obstacles due to 2D laser scan-based obstacle detection causing freezing behaviors and longer detours during navigation. Similarly, Spot's inbuilt autonomy struggles to estimate the ground and

free space from its vision-based perception in both Scenario 3 and Scenario 4. Hence, the robot demonstrates highly unstable motion in vegetation. In scenario 3, VERN is able to navigate through tall grass while avoiding trees and bushes using its vision-based classifier and the occupancy map formulation. However, VERN's vision-based classifier could not detect trees behind the grass region in scenario 4 since they are closely intertwined. In contrast, MIM's multi-layer intensity map representation identifies such hidden solid objects to avoid during navigation. Hence, MIM demonstrates a relatively higher success rate and F-score in scenario 4. In all scenarios, MIMs (uniform and adaptive) have the highest F-score. The inaccuracies in detecting obstacles in MIMs occur when the robot/lidar is closer than 0.5 meters away from an obstacle, whereas 3D lidars typically have a blindspot.

Benefits of Adaptive Inflation: We use scenarios 2 and 3, which contain narrow passages to highlight the benefits of our adaptive inflation formulation. Laserscan and occupancy map-based DWA, and MIM (Uniform) use uniform inflation around the obstacles to avoid collisions. This closes the narrow passages and represents them as obstacle regions, resulting in freezing or longer trajectories in the presence of narrow passages between the obstacles. However, our adaptive inflation preserves the narrow free spaces in the cost map while inflating the obstacles. Hence, MIM can navigate through such spaces (e.g., through a small door in scenario 2 and between the trees in scenario 3) and reach the goals using shorter trajectories.

Inference Time: We compared the inference times (Table II) of using MIMs to detect glass and pliable vegetation with other methods that either detect glass (Glass-SLAM, GDNet) or vegetation (VERN) on the Intel NUC described in section V-A. We observe that apart from being versatile in detecting obstacles, MIMs are computationally light to be used with a robot's limited onboard computing resources. GDNet and VERN use RGB images passed through deep neural networks and require extensive prior training. While Glass-SLAM does not need training, it requires ~ 4.5 seconds to update a map with obstacles and multiple runs to detect glass.

VI. CONCLUSIONS, LIMITATIONS AND FUTURE WORK

We introduce a novel obstacle representation designed to enhance autonomous robot navigation in complex indoor and outdoor environments. Based on the intensity of reflected points from point clouds, MIM effectively characterizes obstacles by their height, solidity, and opacity. Also, we present an adaptive inflation technique that further refines navigation planning by considering obstacle solidity and available free space. We demonstrate significant improvements in navigation metrics such as success rates, trajectory lengths, and F-scores, validating our proposed approach.

Our method has a few limitations. Since our multi-layer map representation is based on point cloud intensity, it cannot identify passable objects such as cloth curtains and metal fences. This is especially important because the navigability of such objects depends on the context (e.g., window curtains may not be passable but door curtains are generally passable). Hence, a semantic understanding of the environment is required for such cases. Further, our method cannot detect extremely thin objects such as thin poles since a 3D point cloud may not be able to capture sufficient number of samples.

REFERENCES

- [1] L. Zhang, Z. Chen, W. Cui, B. Li, C. Chen, Z. Cao, and K. Gao, "Wifi-based indoor robot positioning using deep fuzzy forests," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10773–10781, 2020.
- [2] N. Ramdani, A. Panayides, M. Karamousadakis, M. Mellado, R. Lopez, C. Christophorou, M. Rebiai, M. Blouin, E. Vellidou, and D. Koutsouris, "A safe, efficient and integrated indoor robotic fleet for logistic applications in healthcare and commercial spaces: the endorse concept," in *2019 20th IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2019, pp. 425–430.
- [3] D. Conte, S. Leamy, and T. Furukawa, "Design and map-based teleoperation of a robot for disinfection of covid-19 in complex indoor environments," in *2020 IEEE international symposium on safety, security, and rescue robotics (SSRR)*. IEEE, 2020, pp. 276–282.
- [4] R. R. Shamshiri, C. Weltzien, I. A. Hameed, I. J. Yule, T. E. Grift, S. K. Balasundram, L. Pitonakova, D. Ahmad, and G. Chowdhary, "Research and development in agricultural robotics: A perspective of digital farming," 2018.
- [5] J. Champ, A. Mora-Fallas, H. Goëau, E. Mata-Montero, P. Bonnet, and A. Joly, "Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots," *Applications in plant sciences*, vol. 8, no. 7, p. e11373, 2020.
- [6] M. S. Couceiro, D. Portugal, J. F. Ferreira, and R. P. Rocha, "Sem-fire: Towards a new generation of forestry maintenance multi-robot systems," in *2019 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2019, pp. 270–276.
- [7] H. Mei, X. Yang, Y. Wang, Y. Liu, S. He, Q. Zhang, X. Wei, and R. W. Lau, "Don't hit me! glass detection in real-world scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3687–3696.
- [8] M. Kuribayashi, S. Kayukawa, J. Vongkulbhisal, C. Asakawa, D. Sato, H. Takagi, and S. Morishima, "Corridor-walker: Mobile indoor walking assistance for blind people to avoid obstacles and recognize intersections," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. MHCI, pp. 1–22, 2022.
- [9] M. Hua, Y. Nan, and S. Lian, "Small obstacle avoidance based on rgb-d semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019, pp. 0–0.
- [10] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.
- [11] Y. Wu, Y. Wang, S. Zhang, and H. Ogai, "Deep 3d object detection networks using lidar data: A review," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1152–1171, 2020.
- [12] B. P. Gerkey and K. Konolige, "Planning and control in unstructured terrain," in *ICRA workshop on path planning on costmaps*. Citeseer, 2008.
- [13] X. Ying and M. C. Chuah, "Ucnet: Uncertainty-aware cross-modal transformer network for indoor rgb-d semantic segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 20–37.
- [14] T. Guan, D. Kothandaraman, R. Chandra, A. J. Sathiamoorthy, K. Weerakoon, and D. Manocha, "Ga-nav: Efficient terrain segmentation for robot navigation in unstructured outdoor environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8138–8145, 2022.
- [15] T. M. Petty, J. D. Fernandez, J. N. Fischell, and L. A. De Jesús-Díaz, "Lidar attenuation through a physical model of grass-like vegetation," *Journal of Autonomous Vehicles and Systems*, vol. 2, no. 2, p. 021003, 2022.
- [16] A. Jagan Sathiamoorthy, K. Weerakoon, T. Guan, M. Russell, D. Conover, J. Pusey, and D. Manocha, "Vern: Vegetation-aware robot navigation in dense unstructured outdoor environments," *arXiv e-prints*, pp. arXiv-2303, 2023.
- [17] L. Di Giammarino, I. Aloise, C. Stachniss, and G. Grisetti, "Visual place recognition using lidar intensity information," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 4382–4389.
- [18] K. Sullivan, W. Lawson, and D. Sofge, "Fusing laser reflectance and image data for terrain classification for small autonomous robots," in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*. IEEE, 2014, pp. 1656–1661.
- [19] S. Laible, Y. N. Khan, K. Bohlmann, and A. Zell, "3d lidar-and camera-based terrain classification under different lighting conditions," in *Autonomous Mobile Systems 2012: 22. Fachgespräch Stuttgart, 26. bis 28. September 2012*. Springer, 2012, pp. 21–29.
- [20] C. Reymann and S. Lacroix, "Improving lidar point cloud classification using intensities and multiple echoes," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 5122–5128.
- [21] J. Guo, P. V. Borges, C. Park, and A. Gawel, "Local descriptor for robust place recognition using lidar intensity," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1470–1477, 2019.
- [22] M. Aldibaja, N. Sukanuma, R. Yanase, and K. Yoneda, "Reliable graph-slam framework to generate 2d lidar intensity maps for autonomous vehicles," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–6.
- [23] C. Wei, T. Wu, and H. Fu, "Estimating initial guess of localization by line matching in lidar intensity maps," in *Information Technology and Intelligent Transportation Systems: Volume 1, Proceedings of the 2015 International Conference on Information Technology and Intelligent Transportation Systems ITITS 2015, held December 12-13, 2015, Xi'an China*. Springer, 2017, pp. 577–588.
- [24] I. A. Barsan, S. Wang, A. Pokrovsky, and R. Urtasun, "Learning to localize using a lidar intensity map," *arXiv preprint arXiv:2012.10902*, 2020.
- [25] T. D. Barfoot, C. McManus, S. Anderson, H. Dong, E. Beerepoort, C. H. Tong, P. Furgale, J. D. Gammell, and J. Enright, "Into darkness: Visual navigation based on a lidar-intensity-image pipeline," in *Robotics Research: The 16th International Symposium ISRR*. Springer, 2016, pp. 487–504.
- [26] Y. H. Shin and D.-C. Lee, "True orthoimage generation using airborne lidar data with generative adversarial network-based deep learning model," *Journal of Sensors*, vol. 2021, pp. 1–25, 2021.
- [27] F. Garestier, P. Bretel, O. Monfort, F. Levoy, and E. Poullain, "Anisotropic surface detection over coastal environment using near-ir lidar intensity maps," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 2, pp. 727–739, 2014.
- [28] J. Dupeyron, S. Viollet, and J. R. Serres, "An ant-inspired celestial compass applied to autonomous outdoor robot navigation," *Robotics and Autonomous Systems*, vol. 117, pp. 40–56, 2019.
- [29] J. Li, H. Qin, J. Wang, and J. Li, "Openstreetmap-based autonomous navigation for the four wheel-legged robot via 3d-lidar and ccd camera," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 3, pp. 2708–2717, 2021.
- [30] G. Huskić, S. Buck, and A. Zell, "Gerona: generic robot navigation: a modular framework for robot navigation and control," *Journal of Intelligent & Robotic Systems*, vol. 95, no. 2, pp. 419–442, 2019.
- [31] K. Weerakoon, A. J. Sathiamoorthy, J. Liang, T. Guan, U. Patel, and D. Manocha, "Graspe: Graph based multimodal fusion for robot navigation in outdoor environments," *IEEE Robotics and Automation Letters*, 2023.
- [32] K. M. Wurm, R. Kümmerle, C. Stachniss, and W. Burgard, "Improving robot navigation in structured outdoor environments by identifying vegetation from laser data," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 1217–1222.
- [33] K. Weerakoon, A. J. Sathiamoorthy, U. Patel, and D. Manocha, "Terp: Reliable planning in uneven outdoor environments using deep reinforcement learning," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 9447–9453.
- [34] K. Weerakoon, S. Chakraborty, N. Karapetyan, A. J. Sathiamoorthy, A. Bedi, and D. Manocha, "Htron: Efficient outdoor navigation with sparse rewards via heavy tailed adaptive reinforce algorithm," in *Conference on Robot Learning*. PMLR, 2023, pp. 1629–1639.
- [35] A. J. Sathiamoorthy, K. Weerakoon, T. Guan, J. Liang, and D. Manocha, "Terrapn: Unstructured terrain navigation using online self-supervised learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 7197–7204.
- [36] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," *Science Robotics*, vol. 7, no. 62, p. eabk2822, 2022. [Online]. Available: <https://www.science.org/doi/abs/10.1126/scirobotics.abk2822>
- [37] T. Homberger, L. Wellhausen, P. Fankhauser, and M. Hutter, "Support surface estimation for legged robots," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8470–8476.
- [38] M. Elnoor, A. J. Sathiamoorthy, K. Weerakoon, and D. Manocha, "Pronav: Proprioceptive traversability estimation for autonomous legged robot navigation in outdoor environments," *arXiv preprint arXiv:2307.09754*, 2023.
- [39] G. Kahn, P. Abbeel, and S. Levine, "Badgr: An autonomous self-supervised learning-based navigation system," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1312–1319, 2021.
- [40] A. Polevoy, C. Knuth, K. M. Popek, and K. D. Katyal, "Complex terrain navigation via model error prediction," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 9411–9417.
- [41] R. Soloperto, J. Köhler, F. Allgöwer, and M. A. Müller, "Collision avoidance for uncertain nonlinear systems with moving obstacles using robust model predictive control," in *2019 18th European Control Conference (ECC)*, 2019, pp. 811–817.
- [42] J. Guthrie, M. Kobilarov, and E. Mallada, "Closed-form minkowski sum approximations for efficient optimization-based collision avoid-

- ance,” in *2022 American Control Conference (ACC)*, 2022, pp. 3857–3864.
- [43] L. Weerakoon, G. S. Herr, J. Blunt, M. Yu, and N. Chopra, “Cartographer_glass: 2d graph slam framework using lidar for glass environments,” *arXiv preprint arXiv:2212.08633*, 2022.
- [44] A. J. Sathyamoorthy, U. Patel, T. Guan, and D. Manocha, “Frozone: Freezing-free, pedestrian-friendly navigation in human crowds,” *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4352–4359, 2020.
- [45] X. Wang and J. Wang, “Detecting glass in simultaneous localisation and mapping,” *Robotics and Autonomous Systems*, vol. 88, pp. 97–103, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921889015302670>
- [46] H. Mei, X. Yang, Y. Wang, Y. Liu, S. He, Q. Zhang, X. Wei, and R. W. Lau, “Don’t hit me! glass detection in real-world scenes,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.