2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Real-Time Glass Detection and Reprojection using Sensor Fusion Onboard Aerial Robots

Malakhi Hopkins, Varun Murali, Vijay Kumar and Camillo J. Taylor

***Abstract*— Autonomous aerial robots are increasingly being deployed in real-world scenarios, where transparent obstacles present significant challenges to reliable navigation and mapping. These materials pose a unique problem for traditional perception systems because they lack discernible features and can cause conventional depth sensors to fail, leading to inaccurate maps and potential collisions. To ensure safe navigation, robots must be able to accurately detect and map these transparent obstacles. Existing methods often rely on large, expensive sensors or algorithms that impose high computational burdens, making them unsuitable for low Size, Weight, and Power (SWaP) robots. We present a resource-constrained sensing pipeline for detecting and mapping transparent planar obstacles onboard a sub-300g quadrotor. By exploiting Time-of-Flight (ToF) speckle morphology and sonar-gated fusion, our system identifies specular reflections and reprojects their depth into empty space regions in real-time, with safety margins analytically validated for indoor flight speeds. The entire pipeline operates onboard an embedded processor using approximately 20% of a single CPU core at 2 Hz. We validate our system through experiments in controlled and real-world environments, confirming its ability to accurately render transparent obstacles visible. To our knowledge, this is the first CPU-only, real-time demonstration of transparent plane reprojection on a sub-300g quadrotor.**

I. Introduction

Safe and effective robot navigation is contingent on a robot's ability to accurately perceive its environment. While prior work has successfully addressed volumetric mapping, most methods rely on exteroceptive sensors like Time-of-Flight (ToF), RGB-D sensors or LiDAR, which fail in the presence of transparent or reflective surfaces common in indoor spaces. These materials present a unique challenge as they lack discernible surface features and reflect light variably depending on the viewing angle, making it difficult for conventional depth sensors to obtain precise data. This can lead to inaccurate maps and potential collisions that damage the robot or its surroundings.

Researchers have explored various methods to address the challenge of detecting transparent glass obstacles in the environment. On one hand, some studies have focused on detecting glass using a variety of non-contact sensors, including RGB cameras and LiDARs. These methods leverage advancements in machine learning to achieve high accuracy without physical contact. However, their performances are often influenced by factors such as lighting conditions and angular uncertainty, and they can be computationally expensive to run in real-time on small, low Size, Weight, and Power (SWaP) aerial robots. Additionally, high latency from these models can compromise the robot's operational speed for safety.

On the other hand, an emerging class of solutions has

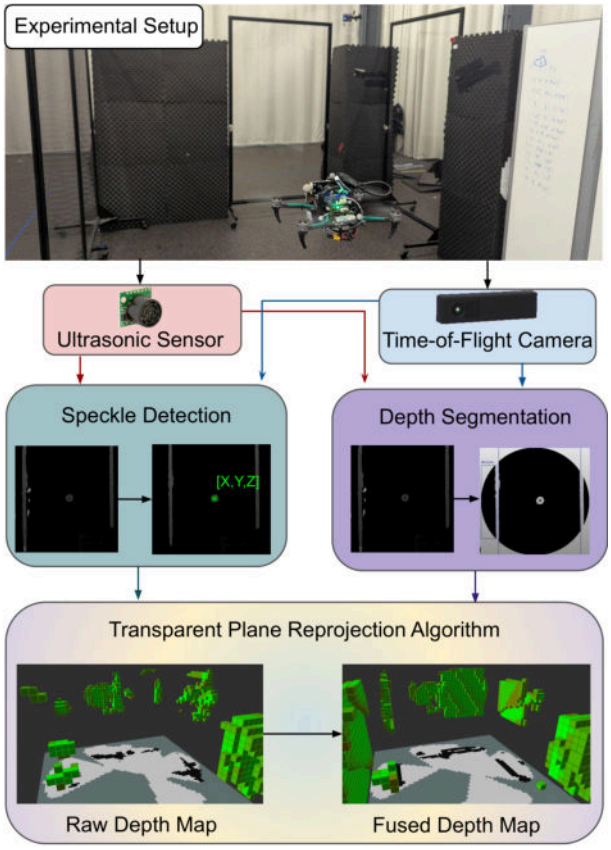


Fig. 1: The experimental setup and algorithmic overview of our framework. Our methodology integrates data from an ultrasonic sensor and a ToF camera to detect specular reflections (speckles) on glass planes and segment the depth image based on empty space. This information is fused by reprojecting the speckle depth into its enclosed segmented region, converting the raw depth map into a fused depth map and effectively making the glass plane visible.

We gratefully acknowledge the support of ARL DCIST CRA W911NF-17-2-0181, NSF Grant CCR-2112665, the NSF Graduate Research Fellowship Grant No. DGE-1845298, and the IoT4Ag Engineering Research Center funded by the National Science Foundation (NSF) under NSF Cooperative Agreement Number EEC-1941529.

explored active, contact-based methodologies. These systems may use visual sensors to identify potential glass surfaces and then actively perform a gentle tactile engagement to confirm or invalidate their presence. While this approach offers robustness by providing physical ground truth, it comes at the cost of increased mechanical complexity, and the need to physically interact with the environment, which can be intrusive and inefficient.

Motivated by the need for a non-intrusive and computationally-lightweight solution for Size, Weight, and Power (SWaP)-constrained platforms, we propose a novel, non-contact approach. To our knowledge, this is the first system to achieve **CPU-only**, real-time transparent plane reprojection on a **sub-300g quadrotor**, whereas prior fusion methods typically require significant compute or higher-SWaP sensors. In summary, our key contributions are as follows:

1) A novel onboard glass obstacle mapping system for low-SWaP aerial robots that integrates compact Time-of-Flight (ToF) and Ultrasonic sensors to enable robust glass detection.
2) A lightweight, computationally efficient framework that operates in real-time on an embedded processor, utilizing a custom 2D convolution and depth image segmentation algorithm to detect specular reflections and accurately reproject glass planes.
3) An extensive experimental validation of our framework in real-time onboard a SWaP-constrained aerial robot, confirming its ability to accurately and efficiently handle glass planes.

II. Related Work

Accurate perception of transparent surfaces remains a bottleneck for autonomous navigation. Existing methodologies generally bifurcate into passive non-contact strategies and active contact-based engagement. We summarize these approaches below and highlight their operational constraints relative to low-SWaP platforms in Table I.

A. Passive, Non-Contact Glass Detection

A significant body of research focuses on non-contact methods that exploit the unique physical and visual properties of transparent materials.

Methods Exploiting Physical Properties: Early approaches capitalized on geometric and reflective properties, such as mirror symmetry [1] or multi-return LiDAR disparities [2]–[5]. While effective, these methods often operate offline, are limited to 2D mapping, or are restricted to structured environments. Other research explores material-specific light absorption [6].

Machine Learning for Perception: Recent advancements leverage machine learning to estimate depth from monocular contextual clues [7]–[9]. While newer foundation models can predict absolute depth [10], they typically impose high computational latency, making them impractical for real-time operation on resource-constrained aerial robots. This constraint also applies to open-world segmentation models like Grounded SAM [11], which utilize large vision transformer backbones. Dedicated networks for glass segmentation [12]–[16] and depth completion [17] similarly rely on large datasets and powerful GPUs.

Multi-Sensor Fusion Methodologies: Fusing modalities can overcome individual sensor limitations by combining complementary strengths. Common approaches include combining RGB-D cameras with ultrasonic sensors [18] or LiDAR with 3D sonar [19]. Other methods fuse LiDAR with infrared or polarization cameras [20], [21]. However, these often rely on expensive, heavy sensors or make restrictive assumptions about environmental geometry.

B. Active and Contact-Based Methods

An emerging paradigm incorporates physical interaction to provide ground truth regarding obstacle presence.

Contact-Resilient Aerial Vehicles: Research into collision-resilient frames and compliant arms allows vehicles to withstand impacts [22]. However, these mechanical additions increase weight and reduce flight agility.

Autonomous Navigation with Active Contact: Active approaches combine visual detection with a physical "touch action" to resolve ambiguities [23]. While providing absolute certainty, this is inherently slow and intrusive, requiring deceleration and replanning.

C. Comparison and Synthesis

As synthesized in Table I, prior works often fail to satisfy the simultaneous requirements of onboard low-SWaP compatibility, real-time CPU-only execution, and non-contact operation. In contrast, our method leverages a physics-informed sensor fusion framework that provides a robust, computationally efficient, and non-intrusive alternative designed specifically for sub-300g aerial platforms.

TABLE I: Comparison of Glass Detection and Mapping Methodologies

Method	Onboard	Low-SWaP	Real-Time	CPU-only	Non-contact
RGB-D + Sonar [18]	Yes	No	Yes	No	Yes
Acoustic + Stereo [24]	No	No	No	No	Yes
ML-based [12], [16]	Yes	No	Yes	No	Yes
Active Contact [23]	Yes	Yes	No	No	No
Ours	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

III. Problem Statement

Let $D \in \mathbb{R}^{H \times W}$ represent a raw Time-of-Flight depth map where a set of pixels $\Omega \subset D$ contains missing data due to specular divergence on transparent surfaces. Given D and a single-point ultrasonic range measurement $s \in \mathbb{R}$, our objective is to estimate the pose $\pi = \{d, \mathbf{n}\}$ and the spatial support $S \subseteq \Omega$ of a transparent planar obstacle, where d is the orthogonal distance to the plane and $\mathbf{n}$ is the surface normal. Rather than solving a global optimization over π, our system estimates these parameters locally from detected speckles and image geometry.

IV. Methodology

This section details our methodology for transparent plane reprojection. We first describe the system's hardware (§ IV-A), followed by the morphology of a speckle (§ IV-B). Next we detail our speckle detection and glass segmentation algorithm and sensor fusion approach (§ IV-C). Finally, we explain the augmentation of the ToF depth image for glass plane mapping (§ IV-D).

A. System Design

Our base platform is the ModalAI Starling 2, which has a Qualcomm QRB5165 processor on a VOXL2 board. This setup includes 8 cores and 8GB LPDDR5 RAM. The robot's small size (280g) is well-suited for our experiments. We integrated a PMD Flexx 2 VGA Time-of-Flight sensor, which provides a 640x480 resolution depth image and a 56° x 44° FoV. We also added a MaxSonar MB1020 Ultrasonic sensor that provides a single-point depth measurement up to 5 meters. Both sensors are aligned to measure the same general region.

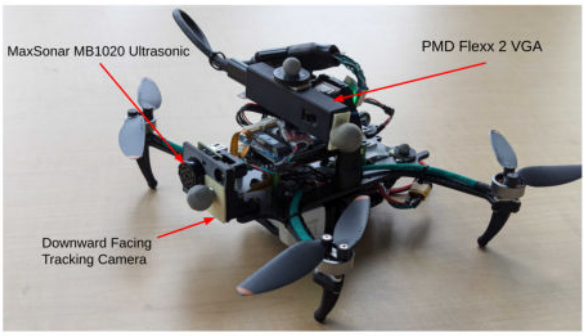


Fig. 2: Our robot is a ModalAI Starling 2 platform with a PMD Flexx 2 VGA Time-of-Flight sensor and a MaxSonar MB1020 Ultrasonic sensor.

B. Specular Reflection Morphology

In a Time-of-Flight (ToF) depth sensing context, a specular reflection (speckle) occurs when infrared light rays reflect off a smooth, transparent surface, such as a glass pane, with minimal scattering. When the ToF sensor's emission cone is oriented nearly perpendicular to the glass, a portion of these reflected rays return to the receiver as high-intensity circular or elliptical artifacts, as illustrated in Fig. 3. The surrounding glass typically appears as empty space or "missing data" because the majority of the divergent light rays are reflected away from the sensor's receiver.

The characteristic circular morphology of these speckles is a systemic artifact of the ToF sensing modality rather than a feature unique to specific hardware. This phenomenon is the result of the sensor's divergent, conical infrared illumination and the flat specular plane of the glass. Since this phenomenon is governed by the fundamental geometry of wide-angle IR emission and specular reflection, speckles are theoretically inherent to Indirect Time-of-Flight (iToF) and Flash LiDAR sensing modalities and were observed consistently on the tested platform.

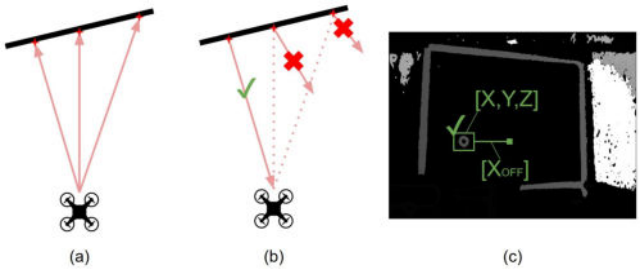
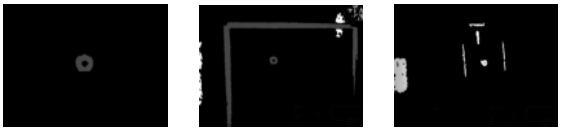


Fig. 3: An illustration of speckle formation. (a) Depicts three light rays from the ToF sensor approaching a tilted glass surface. (b) The light rays reflect, and those perpendicular to the glass surface reflect back to the ToF sensor. (c) These returning rays create the speckle, from which we can extract both the depth and the normal of the glass pane.

We utilize these speckles (Fig. 4) as they provide a verified depth measurement of the transparent surface at the point of incidence. Furthermore, the speckle's image-space coordinates relative to the principal point provide a reliable estimate of the glass pane's surface normal. By fusing the speckle's depth and position, our algorithm can determine the distance and orientation of the glass, enabling the mapping of invisible transparent surfaces as tangible planes within the robot's world model.



(a) Zoomed in (b) Within 2 m (c) Beyond 2 m

Fig. 4: Examples of specular reflections (speckles) in a Time-of-Flight depth image. (a) A close-up view of a single, isolated speckle. (b) The same speckle within the context of a glass pane at a distance of 2 meters. (c) A different speckle and its environment at a distance greater than 2 meters.

C. Speckle Detection and Depth Segmentation Algorithms

To enhance speckle detection, our approach uses the sonar sensor measurement as a filter to remove background clutter from the ToF depth image. Pixels in the ToF image with depth values greater than a small threshold above the sonar's measurement are filtered out, ensuring speckle detection focuses on foreground objects and empty space.

Speckle detection is performed using two custom-designed 2D convolution kernels. These kernels are applied to a Region of Interest (ROI) centered on the image.

Bright Circle Kernel: A uniform circular mask used to detect bright, reflective points from the ToF sensor's light return. The kernel K_B is defined by:

$$K_B(x,y) = \begin{cases} 1 & \text{if } \sqrt{x^2+y^2} \le r_B \\ 0 & \text{otherwise} \end{cases}$$

where r_B is the kernel's radius (*e.g.*, 21 pixels). The kernel is normalized by dividing by the sum of its elements to ensure

a local average.

Dark Ring Kernel: A kernel designed to detect the absence of data (the 'dark hole') immediately surrounded by valid depth data. The kernel K_R is defined by:

$$K_R(x,y) = \begin{cases} 1 & \text{if } r_{in} < \sqrt{x^2+y^2} \leq r_{out} \\ 0 & \text{otherwise} \end{cases}$$

where r_{in} is the inner radius (*e.g.*, 11 pixels) and r_{out} is the outer radius (*e.g.*, 21 pixels). This kernel is also normalized.

The normalized depth image is convolved with both the bright circle and dark ring kernels, producing two response maps. Potential speckle locations are identified as local maxima in both response maps that fall within a score range of 0.3 to 0.9. A successful detection requires a pair of high-scoring peaks from both kernels to be in close proximity, with a Euclidean distance less than the outer kernel's radius.

To validate detected speckles and reduce false positives, we apply a series of filters.

Circularity Check: The raw depth data patch around each detected peak is segmented. The circularity of this contour is calculated as:

$$C = \frac{4\pi \times \text{Area}}{\text{Perimeter}^2}$$

This value is compared against a threshold (*e.g.*, 0.5) to ensure the detected reflection has a circular, blob-like shape, which is a key characteristic of the speckles.

Empty Space Verification: This step distinguishes genuine glass speckles from reflections off background objects. It verifies that the space around the detected speckle is empty. To do this efficiently, an integral image of the binarized depth map is computed, which allows for rapid, constant-time queries of the pixel sum within any rectangular region. We check the pixel sum in eight rectangular regions surrounding the speckle's bounding box. If the ratio of filled pixels to total pixels within these regions is below a low threshold (*e.g.*, 0.07), the speckle is considered isolated within a glass plane.

Temporal Consistency: A final filter operates on a **tracking-by-detection** principle to ensure identified features are persistent and not transient sensor noise. A speckle is confirmed and passed to the mapping algorithm only after its `required_count` (*e.g.*, 1-3 detections) is exceeded across multiple consecutive frames. To prevent the accumulation of false positives and old detections, a `max_age` parameter is used to expire and remove tracks that have not been seen for a specified duration.

D. Transparent Plane Reprojection

The final stage of our methodology involves segmenting empty regions in the depth map and reprojecting the confirmed transparent planes. The algorithm first identifies regions of empty space in the depth image and applies a non-maximum suppression (NMS) algorithm to merge redundant segments, ensuring a single, accurate representation of each transparent plane. The algorithm then isolates the specific segment corresponding to a validated speckle location.

To account for the glass orientation relative to the robot, we apply an affine depth gradient across the segmented region. We model the glass as a planar surface defined by an orthogonal distance d and a surface normal $\mathbf{n}$ under three core assumptions: 1) the surface is locally planar, 2) the small-angle approximation holds for typical indoor approach vectors, and 3) the projection is locally affine. Due to the nature of specular reflection, a speckle is only received when the IR emission ray is approximately perpendicular to the plane, a geometric condition illustrated in Fig. 3 (a)-(b). For a plane tilted at an angle θ relative to the sensor's optical axis, the ray satisfying this condition is offset by an angle $\phi \approx 2\theta$. In the image plane, this angular offset manifests as a horizontal displacement from the principal point, $\Delta u_s = u_s - u_{img\_center}$, as demonstrated in Fig.3 (c). Using a small-angle approximation and the camera's focal length f, the horizontal tilt α can be linearized as:

$$\alpha \approx \tan(\theta) \approx \frac{\Delta u_s}{2f} \tag{1}$$

By incorporating an empirical calibration constant k to account for sensor-specific scaling and non-linearities, we model the localized depth $d(u,v)$ as:

$$d(u,v) = \bar{d} - (u - u_c) \cdot \alpha \tag{2}$$

where u_c is the horizontal centroid of the segment and $\bar{d}$ is the verified depth measurement at the point of incidence. To reflect the fact that the same angular tilt causes a larger pixel displacement at closer ranges, the tilt coefficient α is modulated by the distance:

$$\alpha = \frac{k \cdot \Delta u_s}{\bar{d}} \tag{3}$$

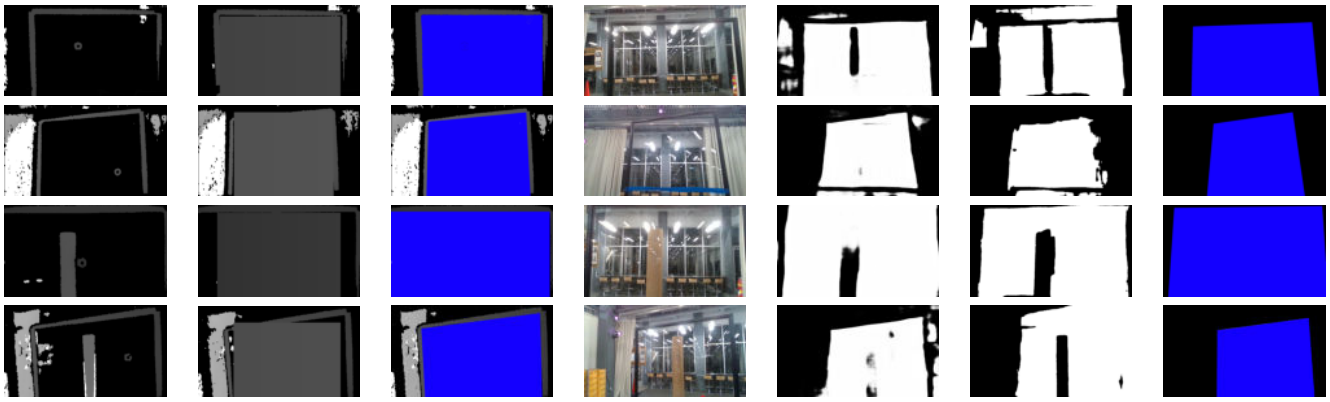
The result is a fused depth map that renders the previously invisible transparent obstacle as a solid, oriented surface. This high-fidelity representation allows the robot's navigation stack to perform precise path planning and obstacle avoidance in environments where traditional ToF or LiDAR sensors would otherwise report traversable free space.

V. Experimental Results

We evaluate the framework through laboratory and real-world experiments. We first quantify segmentation performance against deep learning benchmarks, then characterize mapping completeness and angular accuracy across physical tilts. Finally, we analyze onboard computational efficiency and analytically validate navigation safety margins for real-time operation.

A. Speckle Detection and Glass Segmentation

We evaluated the performance of our speckle detection and transparent plane segmentation algorithms through a series of controlled experiments. These tests were designed to quantify the system's ability to isolate a primary glass pane from background clutter under varying geometric conditions. Table II & III provide the quantitative metrics for pixel-wise and frame-wise precision, recall, and mean intersection over



(a) Baseline (b) Ours (c) GT (d) Baseline (e) GDNet (f) GlassSemNet (g) GT

Fig. 5: Qualitative results for speckle detection and glass segmentation. The figure illustrates a performance comparison between ToF and RGB baselines (a & d), our algorithm (b), ToF and RGB Ground Truth (c & g), GDNet [12] (e), and GlassSemNet [16] (f) in the segmentation of transparent planes with both clear and cluttered backgrounds. Rows 1 and 2 show results from head-on and angled approaches to a clear glass pane, while rows 3 and 4 show results from these approaches repeated with background objects behind the glass.

union (mIoU), while Figure 5 illustrates the corresponding qualitative segmentation results.

TABLE II: Comparative Pixel-wise Segmentation Performance

Experiment Type	Precision	Recall	mIoU
Head-on (Clear) (**Ours**)	**95.7**%	71.2%	**67.2**%
Head-on (Clear) (GSN)	71.8%	54.1%	42.3%
Head-on (Clear) (GD)	62.7%	**89.7**%	58.2%
Angled (Clear) (**Ours**)	**76.8%**	96.4%	**74.7**%
Angled (Clear) (GSN)	76.4%	64.4%	52.9%
Angled (Clear) (GD)	68.2%	**97.2**%	66.7%

TABLE III: Frame-wise Detection Stability and Sonar Impact

Experiment Type	Precision	Recall	mIoU
Head-on (Background)	**92.1**%	**79.6**%	**74.7**%
Head-on (w/o Sonar)	80.2%	65.7%	40.4%
Angled (Background)	100.0%	41.3%	**41.8**%
Angled (w/o Sonar)	100.0%	41.3%	36.6%

Geometric Reliability: The algorithm demonstrated high reliability in clear-background head-on approaches, achieving a precision of 95.7% and a recall of 71.2%. While performance is sensitive to the angle of incidence, the system maintained a precision of 76.8% at a 5-degree angle. The high precision across clear-background experiments suggests that the speckle detection algorithm is highly confident in its identified features, effectively minimizing false-positive glass detections.

Sonar Utility Against Background Clutter: A significant performance delta in frame-wise detection rate was observed with background clutter when the sonar-based depth filter was deactivated. For head-on approaches with background objects, mIoU dropped from 74.7% to 40.4% without sonar filtering. This highlights the ultrasonic sensor's critical role as a "depth-gate," successfully removing background interference that would otherwise prevent the proper segmentation of the transparent plane.

In angled experiments with background clutter, we observed a consistent drop in recall (41.3%) regardless of sonar activation. This phenomenon is likely due to the horizontal speckle offset. In these specific experiments, the speckle has shifted toward the algorithm's predefined detection boundaries, leading to valid features being conservatively rejected. However, the 100.0% precision in these cases indicates that even without the sonar filter, the speckle remained spatially distinct from background occlusions, allowing the algorithm to maintain detection integrity, albeit with a smaller operational window.

Comparative Benchmarking: We compare our framework against two state-of-the-art methods: GDNet [12] and GlassSemNet [16]. To ensure an objective baseline, both models were evaluated using their official parameter configurations and publicly available pre-trained weights. All comparisons were conducted without modifications to the underlying architectures to preserve the integrity of the original methods. Compared to state-of-the-art RGB methods, our approach achieved superior precision and mIoU in both head-on and angled scenarios. While GDNet achieved a higher recall in both experiments, 89.7% head-on and 97.2% angled, qualitative analysis in Figure 5 confirms that our algorithm was the only method capable of consistently excluding background objects from the segmentation mask. This distinction is critical for autonomous navigation, where incorrectly segmenting background clutter as not a part of a glass plane would result in an inaccurate environmental map.

Physical Limitations: We note that the mIoU is consistently impacted by a physical constraint: the robot's sonar mount obstructs the lower-right quadrant of the camera's field of view. This creates a persistent "blind spot" that impedes the segmentation algorithm's ability to close figures with corners in that region, often resulting in inflated segmentation areas. This suggests that while the algorithm is robust, the total system precision is partially bounded by the mechanical integration of the sensors.

B. Mapping Experiments

Our experimental results validate the efficacy of our algorithm. We conducted a series of experiments, progressing from controlled, isolated scenarios to complex, real-world environments, with computation performed onboard the robot. Figure 6 demonstrates our algorithm's ability to successfully map glass panes at angles from 0 to 15 degrees. Table IV displays our quantitative results for our algorithm's frame-wise speckle detection and glass segmentation. The qualitative results in Figure 7 visually demonstrate the mapping completeness of our algorithm.

TABLE IV: Frame-wise Glass Detection and Reprojection Performance Across Experimental Environments

Experiment Type	Environment	Precision	Recall	mIoU
Controlled Experiments	Vicon Room (Handheld)	87.6%	85.7%	66.8%
	Vicon Room (Autonomous)	92.9%	72.3%	64.4%
Real-World Experiments	Half-Glass Room	96.8%	78.9%	69.6%
	Glass Hallway	82.6%	62%	55.7%
	Glass Perimeter	91.2%	50.4%	48.4%

C. Controlled Laboratory Experiments

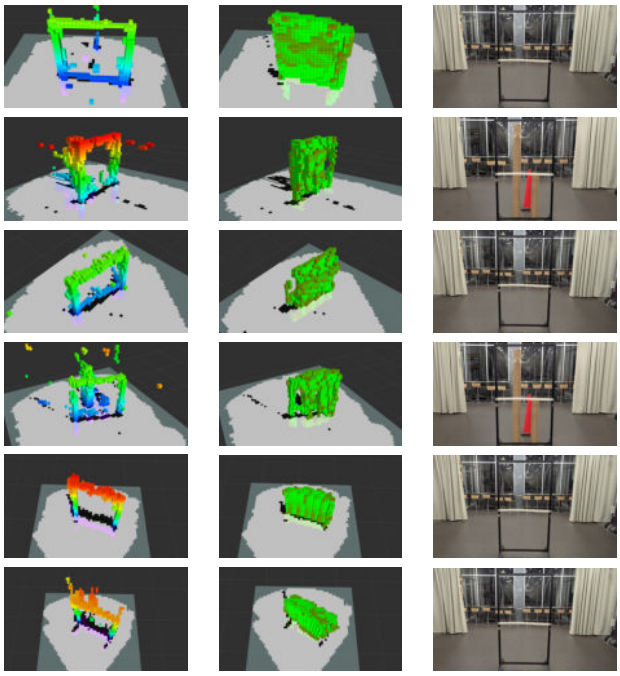
We first established a performance baseline using a series of controlled experiments to isolate and evaluate key components of our algorithm.

Single Glass Pane Experiments: We tested the algorithm's fundamental ability to detect and map during all four of the aforementioned single glass pane experiments. To further display the algorithm's ability to estimate the normal of the glass plane, we conducted two more experiments in which the glass pane was angled at 10 degrees and 15 degrees from either end to center (20 degrees and 30 degrees total). To characterize system stability, 10° and 15° physical tilts were compared against a 0° baseline. The baseline trials established a system noise floor of 1.31° Mean Absolute Error (MAE), while active tilt trials yielded a **combined MAE of $\mathbf{3.12^\circ}$**, confirming the model's accuracy across the operational range.

Vicon Experiments: We conducted mapping experiments within a Vicon motion capture space to provide ground truth for the robot's pose. We built a mock glass room within the Vicon space in a rectangular shape, containing opaque walls as well as 5 glass pane walls. We tested two scenarios: a hand-carried, smooth trajectory and an autonomous flight test using Vicon odometry. Our algorithm was processed onboard the robot in both scenarios. The results are shown in Table IV. The minimal performance degradation in the autonomous flight indicates the algorithm's robustness.

D. Real-World Environment Experiments

Following the controlled tests, we moved to more complex real-world environments to evaluate the system's robustness.



(a) Baseline (b) Ours (c) Scene

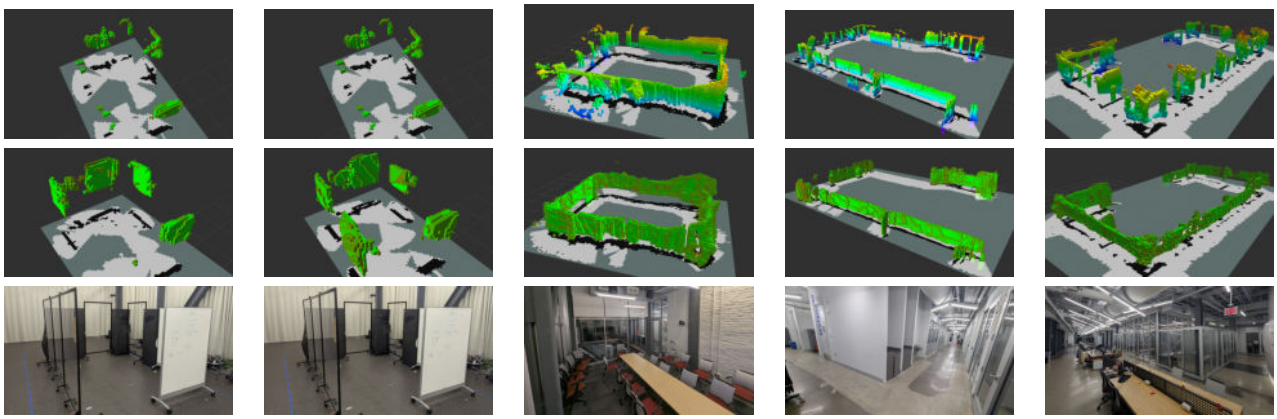
Fig. 6: Qualitative results for mapping Experiments. Each row displays a different experimental condition from top to bottom: a head-on approach to a clear glass pane; a head-on approach with background objects; an approach at a 5-degree angle; an approach at a 5-degree angle with background objects; an approach at a 10-degree angle; and an approach at a 15-degree angle. The columns show the (a) baseline depth map, (b) our algorithm's reprojected map, and the (c) scene view.

During these experiments, all computation was done onboard the robot. The qualitative results in Figure 7 visually demonstrate the high precision and completeness of our mapping algorithm in these environments. The quantitative results are displayed in Table IV.

Half-Glass Room: In this experiment, the robot operated within a room where half of the room is made of several glass windows instead of walls, while the other half is opaque. The algorithm correctly distinguished between the transparent and opaque surfaces, seamlessly integrating the mapped transparent planes with the existing depth data from the solid walls. This scene included stickers on some of the glass panes, preventing speckles from being detected.

Glass Hallway: We performed a mapping experiment in a hallway with large glass panes on one side, which presented challenges from complex reflections and varying ambient light. The algorithm successfully detected and mapped the glass planes along the hallway. In these experiments, multiple glass panes contained stickers that prevented the algorithm from detecting speckles, causing a decrease in mIoU.

Glass Perimeter: We performed a glass mapping experiment in a building floor with a centroid area with large glass panes on all sides. The algorithm successfully detected and mapped the glass planes within the scene. This scene



(a) Vicon (Handheld) (b) Vicon (Autonomy) (c) Half-Glass Room (d) Glass Hallway (e) Glass Perimeter

Fig. 7: Qualitative results from real-world mapping experiments, showing a comparison between the (TOP) baseline depth map and (MIDDLE) our algorithm's reprojected map. The columns represent different environments (Vicon, Half-Glass Room, Glass Hallway, and Glass Perimeter), while the bottom row (BOTTOM) provides a visual reference of the scene.

included multiple scenarios where an opaque object was directly behind a glass pane and within our defined sonar depth filtering threshold, which prevented the background object from being removed from the image. This is further discussed in our Discussion and Limitations Section VI.

E. Onboard Processing Performance

Our full algorithmic pipeline executes directly on the quadrotor's embedded processor. The system operates at a real-time frequency of **2 Hz**, using approximately **20% of a single CPU core** on the VOXL2's QRB5165 processor. This low computational footprint reserves power for other critical functions like odometry and flight control. We specifically conducted our experiments at this processing rate due to SWaP constraints to allow room for other onboard processes that the robot required for flight. To demonstrate scalability, the system can also run at a maximum rate of **10 Hz**, which aligns with our Time-of-Flight (ToF) sensor's maximum frame rate onboard our robot. This requires the full allocation of a single CPU core, providing a flexible trade-off between resource use and operational frequency. Compared to existing methods like GDNet and GlassSemNet, our approach offers significant advantages in performance and efficiency. Table V summarizes this comparison. Due to the variety of onboard processes our robot requires during flight (eg. odometry), it would be unable to process GDNet or GlassSemNet onboard the CPU.

F. Navigation Safety and Detection Latency

The 2 Hz operational frequency is justified via safety margin analysis. At a typical indoor velocity of $v = 0.5$ m/s, the sensing latency of $T_{lat} = 0.5$ s results in a 0.25 m traversal per update. Given the 5 m reliable range of the PMD Flexx sensor, this provides a look-ahead distance exceeding the required stopping distance for a sub-300g micro aerial vehicle (MAV). For dynamic scenarios, the system is scalable to 10 Hz, reducing latency to $T_{lat} = 0.1$ s and enabling safe navigation at higher velocities.

TABLE V: Comparison of Onboard Processing Performance

Metric	Our Method	GDNet	GlassSemNet
Processing Rate (Hz)	2-10	0.057	0.014
Input Rate (Hz)	2-10	19	19
CPU Usage (%)*	20-100	600	700

**Measured across multi-core CPU equivalent usage and normalized to equivalent single-core utilization.*

VI. Discussion and Limitations

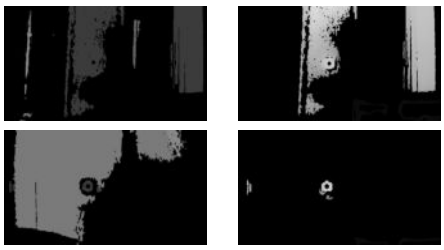
Our algorithm was tested using three distinct presets. Preset 3, the most robust configuration, was employed for the Vicon handheld and autonomous flight experiments, utilizing a circularity threshold of 0.5, an empty threshold of 0.3, and an active sonar depth filter. A more strict Preset 2 was used in real-world experiments, with a circularity threshold of 0.56 and an empty threshold of 0.07. Preset 1, which did not use sonar depth filtering, was applied to the single pane experiments without occlusion. Table VI includes a comparison of all presets, including an extra version of Preset 3 without the sonar filter, validated on the Vicon Room (Autonomous) flight data.

TABLE VI: Ablation Table for Algorithm Preset Comparison

Experiment Type	Preset	Precision	Recall	mIoU
Vicon Room (Autonomous)	1	83%	31%	19.4%
	2	96%	61.7%	56.8%
	3 (w/o sonar)	86.9%	45.9%	32.7%
	3	92.9%	72.3%	64.4%

As depicted in Fig. 8, a weakness in speckle detection is demonstrated by the Glass Perimeter experiment where a background object was so close to the glass that it prevented speckle detection. These objects were within the 0.1 m distance from the glass in which the speckle begins, making them impossible to filter by the sonar. Seven glass panes in the environment were occluded by such obstacles, leading to a significant decrease in performance metrics. In contrast,

the sonar filter in the Vicon Room experiment successfully removed background objects that were further than this distance, allowing for better speckle detection.



(a) Raw Depth (b) Sonar-Filtered

Fig. 8: Comparison of sonar depth filter in (TOP) Glass Perimeter vs (BOTTOM) Vicon Room (Autonomous).

VII. Conclusions

In this work, we presented a novel method for transparent obstacle detection and mapping on a low-SWaP (Size, Weight, and Power) aerial robot. Rather than treating transparent surfaces as perception failures, this work demonstrates that sensing artifacts themselves can be leveraged as reliable geometric cues. By fusing custom 2D convolution for ToF speckle detection with sonar-gated depth filtering, our framework isolates specular reflections from background clutter and reprojects depth into segmented empty regions to reliably map transparent obstacles. The algorithm was validated through a series of experiments, from controlled lab settings to complex real-world environments, demonstrating its ability to accurately and consistently map transparent planes. Our method successfully detected and reprojected glass surfaces during autonomous flight tests, with minimal performance degradation in comparison to handheld experiments. The entire pipeline operates efficiently and in real-time on a resource-constrained embedded processor. This demonstrates our method's viability for onboard deployment and its computational advantage over more intensive, data-driven alternatives that require powerful GPUs. This makes it a practical and effective solution for enabling autonomous navigation in environments with transparent obstacles. To facilitate further research and development in this area, we will open-source our datasets and software. In future work, we plan to enhance our algorithm's robustness to handle more complex glass structures, and extend its capabilities to other transparent and reflective objects.

References

[1] S.-W. Yang and C.-C. Wang, "On solving mirror reflection in lidar sensing," *IEEE/ASME Transactions on Mechatronics*, vol. 16, no. 2, pp. 255–265, 2010.

[2] R. Koch, S. May, P. Koch, M. Kühn, and A. Nüchter, "Detection of specular reflections in range measurements for faultless robotic slam," in *Robot 2015: Second Iberian Robotics Conference: Advances in Robotics, Volume 1*. Springer, 2016, pp. 133–145.

[3] R. Koch, S. May, P. Murmann, and A. Nüchter, "Identification of transparent and specular reflective material in laser scans to discriminate affected measurements for faultless robotic slam," *Robotics and Autonomous Systems*, vol. 87, pp. 296–312, 2017.

[4] R. Koch, S. May, and A. Nüchter, "Effective distinction of transparent and specular reflective objects in point clouds of a multi-echo laser scanner," in *2017 18th International Conference on Advanced Robotics (ICAR)*. IEEE, 2017, pp. 566–571.

[5] ——, "Detection and purging of specular reflective and transparent object influences in 3d range measurements," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 377–384, 2017.

[6] H. Tibebu, J. Roche, V. De Silva, and A. Kondoz, "Lidar-based glass detection for improved occupancy grid mapping," *Sensors*, vol. 21, no. 7, p. 2263, 2021.

[7] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, and H. Zhao, "Depth anything: Unleashing the power of large-scale unlabeled data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 10 371–10 381.

[8] N. Simon and A. Majumdar, "Mononav: Mav navigation via monocular depth estimation and reconstruction," in *International Symposium on Experimental Robotics*. Springer, 2023, pp. 415–426.

[9] R. Birkl, D. Wofk, and M. Müller, "Midas v3. 1–a model zoo for robust monocular relative depth estimation," *arXiv preprint arXiv:2307.14460*, 2023.

[10] S. F. Bhat, R. Birkl, D. Wofk, P. Wonka, and M. Müller, "Zoedepth: Zero-shot transfer by combining relative and metric depth," *arXiv preprint arXiv:2302.12288*, 2023.

[11] T. Ren, S. Liu, A. Zeng, J. Lin, K. Li, H. Cao, J. Chen, X. Huang, Y. Chen, F. Yan, *et al.*, "Grounded sam: Assembling open-world models for diverse visual tasks," *arXiv preprint arXiv:2401.14159*, 2024.

[12] 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.

[13] H. Mei, B. Dong, W. Dong, J. Yang, S.-H. Baek, F. Heide, P. Peers, X. Wei, and X. Yang, "Glass segmentation using intensity and spectral polarization cues," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12 622–12 631.

[14] E. Xie, W. Wang, W. Wang, M. Ding, C. Shen, and P. Luo, "Segmenting transparent objects in the wild," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*. Springer, 2020, pp. 696–711.

[15] H. Lin, Z. Zhu, T. Wang, A. Ioannou, and Y. Huang, "Glass surface segmentation with an rgb-d camera via weighted feature fusion for service robots," 05 2025, pp. 517–524.

[16] J. Lin, Y.-H. Yeung, and R. W. Lau, "Exploiting semantic relations for glass surface detection," *NeurIPS*, 2022.

[17] A. Costanzino, P. Z. Ramirez, M. Poggi, F. Tosi, S. Mattoccia, and L. Di Stefano, "Learning depth estimation for transparent and mirror surfaces," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 9244–9255.

[18] Z. Huang, K. Wang, K. Yang, R. Cheng, and J. Bai, "Glass detection and recognition based on the fusion of ultrasonic sensor and rgb-d sensor for the visually impaired," in *Target and Background Signatures IV*, vol. 10794. SPIE, 2018, pp. 118–125.

[19] T. Q. Tran, A. Becker, and D. Grzechca, "Environment mapping using sensor fusion of 2d laser scanner and 3d ultrasonic sensor for a real mobile robot," *Sensors*, vol. 21, no. 9, p. 3184, 2021.

[20] P. Zhang, G. Fan, L. Rao, S. Cheng, X. Song, N. Chen, and Z. Xu, "A glass detection method based on multi-sensor data fusion in simultaneous localization and mapping," in *International Conference on Autonomous Unmanned Systems*. Springer, 2022, pp. 1386–1400.

[21] E. Yamaguchi, H. Higuchi, A. Yamashita, and H. Asama, "Glass detection using polarization camera and lrf for slam in environment with glass," in *2020 21st International Conference on Research and Education in Mechatronics (REM)*. IEEE, 2020, pp. 1–6.

[22] S. Shen, N. Michael, and V. Kumar, "Autonomous indoor 3d exploration with a micro-aerial vehicle," in *2012 IEEE international conference on robotics and automation*. IEEE, 2012, pp. 9–15.

[23] X. Chen, Y. Zhang, H. Zou, J. Wang, and S. Shen, "Active contact engagement for aerial navigation in unknown environments with glass," *IEEE Robotics and Automation Letters*, 2025.

[24] M. Ye, Y. Zhang, R. Yang, and D. Manocha, "3d reconstruction in the presence of glasses by acoustic and stereo fusion," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4885–4893.