ResQDrone: Autonomous Indoor Quadcopter for Multi-Layered Mapping in Crisis and Disaster Management Scenarios

Marco Cella*, Jürgen Biber, Pierrick Bournez, Jules Salzinger, Verena Widhalm,
Francesco Vultaggio, Francesco d'Apolito, Felix Bruckmüller, Christoph Sulzbachner, Phillipp Fanta-Jende

Austrian Institute of Technology - Center for Vision, Automation and Control
Email: (marco.cella, juergen.biber, jules.salzinger, verena.widhalm, francesco.vultaggio, francesco.dapolito,
felix.bruckmueller, christoph.sulzbachner, phillipp.fanta-jende)@ait.ac.at, pierrick.bournez@gmail.com

Keywords: Autonomous Exploration, UAV, Crisis and Disaster Management, Mapping

Abstract

This paper introduces a fully autonomous quadcopter developed for the safe and efficient exploration of indoor environments in crisis and disaster management scenarios. These environments are typically partially destroyed, entirely unknown, and suffer from restricted visibility. During exploration, multiple mapping layers are generated and transmitted to human operators in real-time, delivering critical insights into the structure. These layers include geometric, semantic, and thermal maps, each contributing distinct information: the geometric layer provides a precise volumetric layout; the semantic layer enriches this layout with contextual understanding of different structural classes; the thermal layer highlights relevant signatures — e.g., humans — that support timely and informed intervention. Successful tests in controlled and simplified real-world environments demonstrate the system's robustness and its practical value in operational contexts.

1. Introduction

Despite significant progress in rescue robotics, human involvement remains crucial in Crisis and Disaster Management (CDM) scenarios (Murphy et al., 2016), even though it can be hazardous — particularly during chemical or biological incidents. This paper introduces a fully autonomous quadcopter that, although tested only in simplified and controlled environments, is designed to explore complex indoor spaces characterised by the absence of GNSS, limited visibility, and challenging geometries. As illustrated in Figure 1, the system delivers multiple mapping layers in real-time to support first responders. These layers include:

- A **geometric** map, representing the building's internal structure and layout, which helps identify key features such as potential entrances.
- A **semantic** map, computed by processing the point cloud in batches. Ceilings, walls, floors, doors, and windows are assigned distinct class labels, while all other elements are categorised as clutter. This helps clarify ambiguous clusters of points in the geometric map.
- A **thermal** map, in which each point contains thermal information obtained by fusing thermal imagery with LiDAR data. This enables accurate localisation of heat signatures — such as people — within the spatial map, providing valuable context.

Due to potential hazards in the building and the difficulty of manually piloting a quadcopter indoors, complete mission autonomy is critical. To achieve this, the drone requires a suitable sensor suite (Figure 3, Section 3.1) to enable SLAM (Koide et al., 2024), autonomous exploration and safe motion planning (Bolz et al., 2024), as well as point cloud (Bournez et al., 2024) and thermal image semantic understanding.

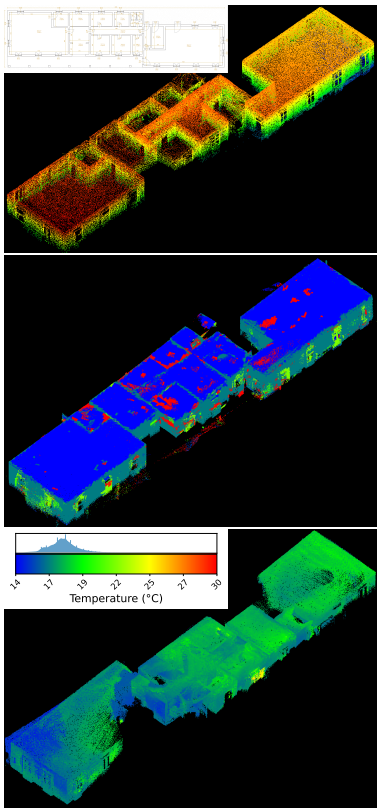

Figure 1. The various mapping layers. From top to bottom: geometric, semantic, thermal. The geometric layer is colour-coded based on the z value, the semantic based on each point's class, and the thermal on its temperature. The maps have been obtained by manually flying the drone.

* Corresponding author

The main contributions to the literature are:

- The implementation and testing on a real drone of a complete algorithmic suite for autonomous exploration, composed of an exploration logic and motion planning module and a multi-layered mapping framework;
- The development of a novel LiDAR/thermal camera fusion module that works in real-time, providing globally referenced thermal signatures;
- The validation of this system in a controlled and simplified real-world environment;
- A framework designed to advance the development of robust Active-SLAM systems for UAVs operating in unstructured, uncooperative indoor environments.

2. Related Work

2.1 Autonomous CDM Quadcopters

Quadcopters have proven to be the ideal platform for collecting data in indoor environments. Their agility and manoeuvrability make them the preferred choice for first responders operating in hazardous indoor settings during rescue missions. This potential was highlighted during the 2023 First Responder UAS 3D Mapping Challenge [1], hosted by the National Institute of Standards and Technology (NIST), where multiple teams demonstrated their drone-based mapping solutions for such scenarios. While the presented approaches were highly capable, they all depended on manual piloting.

Some solutions for complete autonomous exploration of indoor environments are already commercially available, albeit at different readiness levels. Representative examples are Exyn's Nexys [2], Fixar [3], and Skydio [4].

Ongoing efforts by various researchers aim to tackle this problem (Zhang et al., 2024, Zhang et al., 2022), however these solutions rely on depth cameras, which are not suitable sensors for CDM scenarios since the environment could be filled with smoke or be completely dark. Furthermore, the use of cameras reduces the efficiency of the autonomous exploration logic, which benefits from having a 360° view of the surroundings - as provided by hemispherical LiDARs. (Yao and Liang, 2024) presents a quadcopter designed for LiDAR-based autonomous forest exploration. Because the UAV operates under the forest's canopy, GNSS signals are unavailable, so the scenario closely mirrors the conditions studied in this paper. The authors employ an Ouster OS1 [5] — a more sophisticated but also far more expensive and heavy LiDAR sensor than the Livox Mid-360 — making their solution unsuitable for missions in which the drone must be lightweight and expendable (*e.g.*, operations in areas contaminated by biological or chemical agents). Moreover, the paper reports no indoor test results. The work most similar to ours is FAEM (Zhang et al., 2025), which likewise uses the Livox Mid-360 for autonomous indoor exploration.

[1] https://www.nist.gov/ctl/pscr/open-innovation-prize-challenges/past-prize-challenges/2023-first-responder-uas-3d-mapping
[2] https://www.exyn.com/products/exyn-nexys
[3] https://fixar.pro/products/fixar-indoor/
[4] https://www.skydio.com/
[5] https://ouster.com/products/hardware/os1-lidar-sensor

Autonomous quadcopters intended for CDM rely not only on robust flight and navigation capabilities but also on the ability to construct accurate, real-time maps of their surroundings. In particular, operating in degraded environments — such as smoke-filled or dark buildings — demands mapping solutions that do not depend on visual cues. This makes LiDAR-based indoor mapping a critical enabling technology for autonomous CDM-focused UAVs, motivating a closer look at the state of the art in this domain.

2.2 Autonomous Indoor Mapping

LiDAR–Inertial SLAM Due to the limited payload capacity of aerial vehicles — especially those operating indoors — most have historically relied on vision-based SLAM techniques (Elmokadem and Savkin, 2021). In contrast, LiDAR-based systems were typically developed for ground use, where weight and power are less constrained (Zhu et al., 2024). Modern lightweight sensors such as the Livox Mid-360, however, now make it feasible to equip indoor aerial robots with LiDAR. In this context, two characteristics become important for robust tracking: incorporating accelerometer data to capture the high-frequency motion typical of indoor flight, and handling the low point density inherent to the Mid-360's scans. In our earlier work (Vultaggio et al., 2023), we analysed the impact of low point cloud density and IMU data on LiDAR SLAM in simulated environments. While some approaches (Xu et al., 2022, Dellenbach et al., 2022, Shan et al., 2020) showed promising results, these methods didn't generalize in our real world tests. For this reason, this work employs GLIM (Koide et al., 2024), a tightly coupled range–inertial SLAM framework that advances the state of the art in three key respects: (i) *Robustness to degeneracy* — a fixed-lag smoother combined with key-frame Generalized-ICP can absorb several seconds of feature-poor data without corrupting the trajectory; (ii) *Throughput* — all scan-matching and global optimisation factors are off-loaded to the GPU, sustaining real-time mapping at > 20 Hz while freeing CPU cycles for flight control (feature which is not used in this paper); (iii) *Sensor-agnostic extensibility* — thanks to its generic range-factor formulation and callback slots, GLIM runs unchanged on spinning, solid-state, or non-repetitive LiDARs and can be augmented with visual or thermal constraints when available. These properties make GLIM an excellent fit for the LiDAR-only, real-time exploration and mapping pipeline required in CDM scenarios.

Semantic Mapping Rapid scene understanding is critical for first responders that must gain knowledge of cluttered indoor spaces while relying only on LiDAR data. As summarised in the recent survey of (Alqobali et al., 2023), most existing indoor semantic-mapping pipelines still depend on RGB cues. (Bournez et al., 2024) addresses this limitation by proposing a LiDAR-only network that maintains quasi-real-time inference (10 Hz), making it suited for time-critical missions while preserving robustness and accuracy.

Only a handful of other works consider RGB-free indoor settings. MapSegNet projects LiDAR scans to 2-D occupancy grids before applying image CNNs, sacrificing geometric fidelity (Foroughi et al., 2021). (Alenzi et al., 2022) relies on handcrafted features and classical classifiers that might not generalise across buildings, as well as closed-source datasets for training. Generic point-centric backbones such as PointNeXt (Qian et al., 2022) and KPConv (Thomas et al., 2019) do not scale well with the size of point clouds, exceeding the

memory or latency budgets of quasi-real-time systems. Consequently, integrating the lightweight LiDAR-only segmentation of (Bournez et al., 2024) into our mapping stack fills a crucial gap, enabling semantics-aware, real-time exploration and mapping for mission operators.

Thermal Mapping Regarding the thermal point cloud generation, some works tackle the topic, but they either do not perform the fusion in real-time (Qiu et al., 2025), or they use LiDARs that are not suited for indoor exploration (Arsene, 2020).

3. Proposed Method

Figure 2 summarizes the flow of information in the proposed approach. Starting from the basic sensors — IMU, LiDAR, and thermal camera — the data is processed by multiple modules and then transformed in the three final products: geometric, semantic, and thermal point clouds. The following section will describe each module in detail.

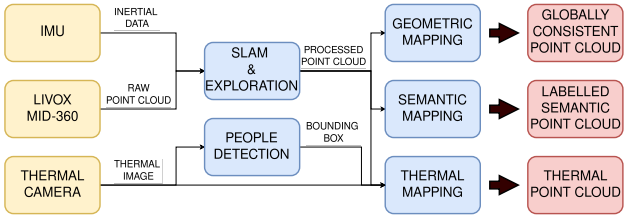

Figure 2. The overall scheme of the proposed solution.

3.1 Platform

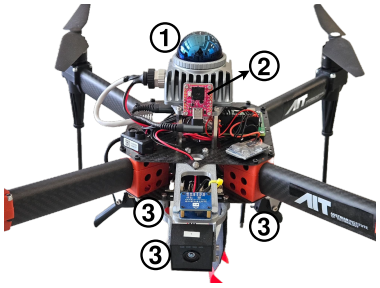

Figure 3. The drone and its sensor suite. (1) Livox Mid-360 LiDAR; (2) Thermal Camera PureThermal 3 with FLIR Lepton 3; (3) Cygbot LiDARs.

Figure 3 showcases the drone that was developed to execute the autonomous exploration and mapping missions. As previously stated, CDM scenarios may be characterized by low visibility. Thus, the sensor suite must be carefully chosen. This limitation excludes the use of any conventional RGB-D cameras and optical flow sensors. Instead, LiDARs can be used.

The main sensor is the Livox MID-360[6], a hemispherical LiDAR characterized by a 360° by 59° Field of View (FOV), a range of 0.1-40m, and by a non-repeating helicoidal acquisition pattern which saturates the environment if left stationary. Then, three 3D solid-state Time-of-Flight (ToF) LiDARs[7] with a 120° by 65° FOV and a range of 0.05m to 2m are mounted on the sides of the drone for local collision avoidance. Finally, a PureThermal3 with a FLIR Lepton 3 thermal camera[8] is mounted next to the main LiDAR. Its purposes are to detect elements of interest such as people, and to assign temperature values to each point of the point cloud that intersects with its FOV.

The total sensor suite has a weight of 361g and a power consumption of 18W.

Given the constrained on-board processing power, all the algorithms that will be introduced — also shown in Figure 2 — are set to run on a Ground Control Station (GCS). The GCS then communicates via Wi-Fi with the drone, ensuring a robust connection in small environments.

3.2 Autonomous Exploration

SLAM Given the lack of GNSS signals, no absolute positioning system is available. For this reason, a SLAM framework was implemented to simultaneously track the drone's position and build a globally consistent map of the environment. The key element for high-quality measurements is the fusion of multiple sensors — an IMU for high-frequency motion prediction and a LiDAR for lower-frequency correction — using a factor graph-based algorithm. As stated in Section 2.2, GLIM (Koide et al., 2024) was selected as the base for SLAM due to its strong performance in long-term position tracking and mapping accuracy (see Section 4 for quantitative results). This global map forms the **geometric** layer referenced in Section 1.

The collected point cloud is further used to construct a voxel map, which can be used by the motion planner for computing collision-free paths. In this map, each voxel can be either unknown, observed free, or observed occupied.

Exploration Logic and Motion Planning To guide the drone toward unexplored areas, our previous work (Bolz et al., 2024) was implemented. Briefly explained, a frontier-based approach is used, where a frontier is any *unknown* voxel adjacent to at least one *observed free* cell. These frontiers are clustered into n centroids. The optimal goal is selected by minimizing a custom cost function, which balances three factors: the distance to the frontier, the number of unexplored voxels visible from that location, and the angle between the drone's motion direction and the frontier. The exploration behaviour can be tuned by modifying the cost function's weights, encouraging efficient and consistent exploration without unnecessary deviations.

Once a goal is selected, a smooth B-Spline trajectory is planned based on the method from (Zhou et al., 2021). A receding-horizon strategy executes only part of the trajectory before replanning, allowing the drone to adapt to new obstacles.

3.3 Semantic Point Cloud Understanding

SuperPoint Transformer for Online segmentation The Superpoint Transformer (Robert et al., 2023) is expressly designed for scenarios where both the spatial extent of the scan and the required prediction horizon can vary at run time. It first constructs lightweight, hand-crafted geometric descriptors on each points' neighbourhoods, then partitions the cloud into geometrically coherent *superpoints* before giving those as input to a neural network.

Building on this backbone, (Bournez et al., 2024) extends the model along three axes that are pivotal for quasi-real time deployment. First, they redesign the handcrafted descriptor set to

[6] https://www.livoxtech.com/mid-360
[7] https://www.cygbot.com/
[8] https://groupgets.com/products/purethermal-3

improve their robustness on sparse point clouds, better discriminate thin structures such as doors and window frames. Second, the expensive partitioning algorithm is replaced with a lightweight voxel-grid clustering, and the k-NN radius is enlarged to stabilise feature statistics under noisy observations. Third, to compensate for the lack of contextual information in small point clouds, the authors devise a two-stage curriculum learning strategy that (i) simulates realistic sensor trajectories inside S3DIS meshes and (ii) adds visibility-aware Hidden-Point Removal augmentation.

These adaptations translate into a real-time capable model, not only because of its latency but also in its capacity to understand low-quality and incomplete point clouds. Crucially, the new pipeline proves robust to SLAM noise and achieves reliable detection of walls, ceilings/floors, as well as competitive results on harder classes such as doors or windows, supplying our semantic layer for mission operators.

Model integration and data pre-processing For use during missions, we encapsulate the LiDAR-only Superpoint Transformer variant in a dedicated ROS node. The node listens to the localised messages published by the SLAM back-end and buffers the last 10 *s* of point cloud messages. When the buffer fills, the individual scans are concatenated and passed through a voxel-grid downsampling algorithm (with voxel size 0.03m) to keep a reasonable point density and ensure a processing time under the 10 *s* aggregation time. This also ensures that Superpoint Transformer's neighbourhood-based features remain computed on similar spatial extents as during the model's training.

However, in cases where the UAV finds itself in particularly large locations (with respect to the training data), the number of points kept by the downsampling step might be too large to ensure a timely processing of the buffer. In addition, the features of such large spaces (or conversely small places) might be misinterpreted by the model as they may constitute outliers compared to areas seen during training. To compensate for this effect, we investigate a training-aware rescaling of the buffer before the downsampling operation.

To be precise, we define the **extent** as the cubic root of the mean volume computed over N point cloud messages accumulated over 10 *s*:

$$\text{extent} = \sqrt[3]{\frac{1}{N}\sum_{i=1}^{N} V_i}$$

where V_i denotes the estimated volume of the i-th point cloud message. The volume of an individual point cloud message is estimated using an Oriented Bounding Box (OBB), where the vertical axis (z-axis) is fixed. The OBB is computed via principal component analysis, followed by outlier removal based on the 1st and 99th percentiles along each axis to ensure robustness against noise. We compute the extent of 10 *s* buffers made from the S3DIS training split according to (Bournez et al., 2024), then investigate different ways to use this metric for rescaling the input point clouds (see Section 4).

The resulting down-sampled clouds are then converted into the superpoint graph representation and forwarded to the model. After inference, the node repacks the predicted class labels into a new message and republishes it. These 10 *s* buffers are then aggregated into the final **semantic** map to be used by first responders, providing additional visual cues to interpret the point cloud.

3.4 Human Detection in Thermal Images

The exploration of indoor environments in CDM scenarios introduces specific requirements for the datasets used in model training and evaluation. Traditional RGB imagery is often inadequate in such conditions due to limited visibility, low lighting, and background clutter. In contrast, thermal imaging captures infrared radiation emitted by objects, reflecting their temperature and emissivity properties, and is thus more suitable for these scenarios (Wilson et al., 2023).

In addition, human detection in CDM scenarios must account for uncommon body poses, such as individuals lying on the ground (face-up or face-down), sitting, or squatting — poses frequently encountered in real-world cases. However, most publicly available datasets primarily depict people in upright or moving positions. A limited number of datasets address this gap by including annotated instances of people in atypical postures (Stippel et al., 2023, Cruz Ulloa et al., 2021, Tsai et al., 2022, ThermalObjectDetection, 2024) as well as synthetic generated datasets modelling human behaviour in indoor settings (Pramerdorfer et al., 2020).

For model training, the YOLOv8 (Varghese and Sambath, 2024) network was employed due to its strong performance and real-time inference capabilities, supporting tasks such as object detection and instance segmentation.

3.5 Thermal Point Cloud Generation

The thermal camera can also be leveraged to enhance the collected point clouds with temperature information. The proposed fusion node operates as a real-time middleware layer that unifies LiDAR point clouds with co-registered thermal imagery and per-pixel temperature data. After retrieving the rigid LiDAR-to-camera extrinsics and the camera intrinsics, the node time-synchronises three incoming streams — point cloud, infrared image, and temperature array. Each 3D point that falls within the camera FOV is projected onto the undistorted thermal image, assigned a colour according to its interpolated temperature, and then re-expressed in the global map frame using the vehicle's pose. The algorithm publishes this temperature-augmented point cloud continuously, which constitutes the **thermal** layer mentioned in Section 1.

The same logic is used parallelly to project the detected humans — as previously described in Section 3.4 — on the semantic point cloud. The functioning is straight-forward: the points whose image projections lie inside a "person" bounding box are accumulated and added to the semantic cloud described in Section 3.3.

4. Experiments and Results

In our previous work (Bolz et al., 2024), the autonomous exploration framework was thoroughly tested in complex simulated environments. In this paper, we present real-world tests conducted in two simplified and controlled settings. The first, shown in Figure 4, was designed to give the drone ample freedom to explore while allowing the human pilot enough time to intervene if it began drifting toward an obstacle.

Additional tests took place in a more constrained setting, a 153 cm wide corridor with multiple connected rooms, depicted in Figure 6. For safety, the planner's clearance threshold was set

Figure 4. The setup for the simplified room testing. The top image shows the corrected SLAM point cloud, colour-coded based on the z value of each point, and the trajectory (in white) autonomously executed by the drone to fully map the room.

higher than the width of the doorways, limiting exploration to the corrido itself.

Despite the relative simplicity of these environments, the tests demonstrated the drone's ability to navigate narrow spaces without collisions. Future work will address exploration in more complex scenarios.

4.1 Autonomous Mapping

Autonomous Exploration The chosen metrics to evaluate the performance of the exploration logic are the *Observed Volume* $[m^3]$ and the *Distance Traveled* $[m]$. The plots in Figure 5 represent the mean and the standard deviation over various runs — 4 for the simple room and 2 for the corridor — given the non-deterministic nature of the algorithm. Notice that the simple room in which we tested had a volume of roughly 400 m^3, and the small variation towards the end of the plot can be attributed to the mapping of what the LiDAR saw outside the windows. The same consideration is valid for the results in the corridor environment.

Geometric Mapping To evaluate the quality of the geometric map produced by the SLAM module, the standard Cloud to Cloud (C2C) distance metric has been used. For the ground truth, a building plan was used, which was post-processed and converted into a point cloud to make the C2C comparison possible. Figure 7 shows the C2C histogram, highlighting how most of the points are off by 10 cm from the ground truth. This information can be used to adapt the planner's clearance threshold to ensure safe trajectories.

4.2 Semantic Mapping

To evaluate the effect of rescaling the input point clouds on the semantic mapping pipeline, the Intersection over Union (IoU) of each individual class was recorded. The average of the IoUs for all classes, called *mean IoU* (mIoU), is also reported. First,

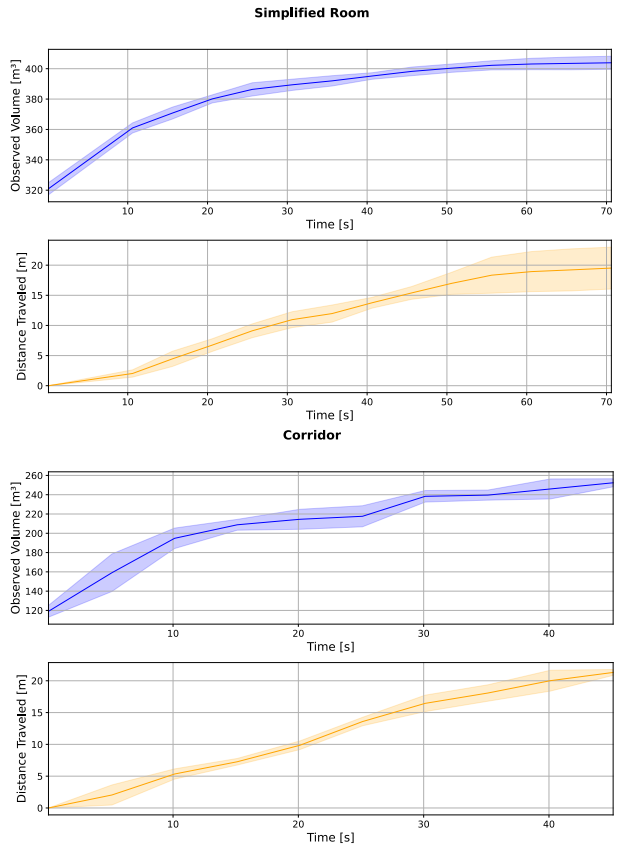

Figure 5. The exploration metrics plot for each scenario, showing the mean and the standard deviation of each metric.

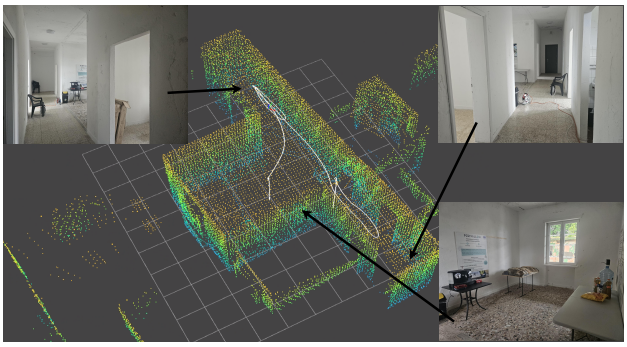

Figure 6. The explored corridor. Notice that the arrows point to the spots on the map where the corresponding photo was taken. The white trajectory represents the path executed by the drone.

the 10 s batches used in (Bournez et al., 2024) at training time are analysed, and the empirical distribution function $\hat{F}(e)$ of its extent is recorded.

In any given experiment, if the extent e of a point cloud is an outlier, defined as $\hat{F}(e) < \alpha$ or $\hat{F}(e) > 1 - \alpha$, it is uniformly rescaled such that its extent becomes an inlier. To be precise, if $\hat{F}(e) < \alpha$, it is rescaled such that its new extent becomes e_{new} with $\hat{F}^{-1}(\beta) = e_{\text{new}}$. Conversely, if $\hat{F}(e) > 1 - \alpha$, it is rescaled such that its new extent becomes e_{new} with $\hat{F}^{-1}(1 - \beta) = e_{\text{new}}$. The parameter α is interpreted as the *correction threshold*, which decides whether to apply a correction or not, and β as the *correction amount*, which, when a correction is

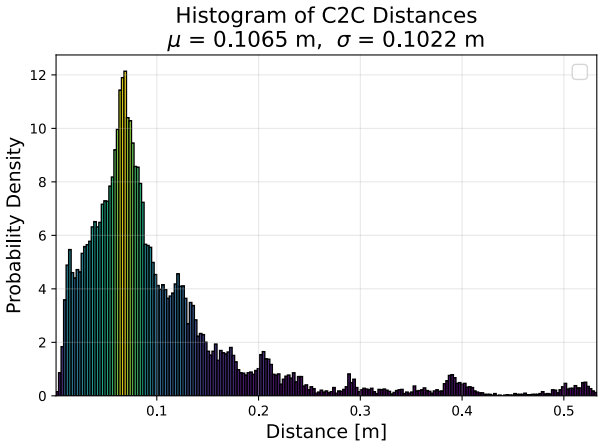

Figure 7. Histogram of the C2C distances between the generated geometric map and the building's plan.

applied, determines how significant this correction should be.

Experiments are performed using 3 datasets ($S3DIS_{10s}$, $Office_{10s}$ and $Building_{10s}$ following (Bournez et al., 2024)'s nomenclature). (Armeni et al., 2017)'s meshes are simulated in a Gazebo environment and split every 10 s, resulting in $S3DIS_{10s}$. This dataset was used to train the model, so it is used to compute the statistics described above. $Office_{10s}$ is a real dataset reported for comparability with (Bournez et al., 2024). Finally, we contribute an additional and similarly annotated test area, denoted $Building_{10s}$ (see Figure 1). Those datasets have extents following the distributions shown on Figure 8, where the percentiles of the $S3DIS_{10s}$ extent distribution are also represented for easier reference.

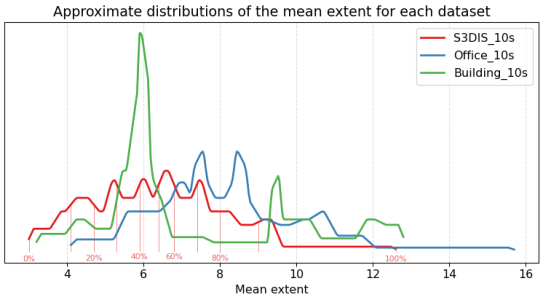

Figure 8. Smoothed approximate distributions of the extent values for each dataset. The $S3DIS_{10s}$ is additionally annotated by vertical bars showing every tenth percentile for easier reference to the experiment parameters used in this study.

As can be seen on Figure 9, the proposed method behaves differently depending on the dataset. The results on $S3DIS_{10s}$ mostly do not change much or degrade sightly, which is to be expected for the model's synthetic training data. For $Office_{10s}$, the total amplitude of the mIoU is below 1% despite its extent distribution being markedly more shifted towards larger point clouds impacted by the method (see Figure 8). The rescaling achieves an average total processing time of 1.8 s (from 2.2 s) for average values of the parameters ($\alpha = 15$ and $\beta = 30$) while our maximum processing time for a 10 s buffer (*i.e.*, the outliers needed to be tackled) goes down to 2.9 s (from 4.5 s) without significant loss of mIoU.

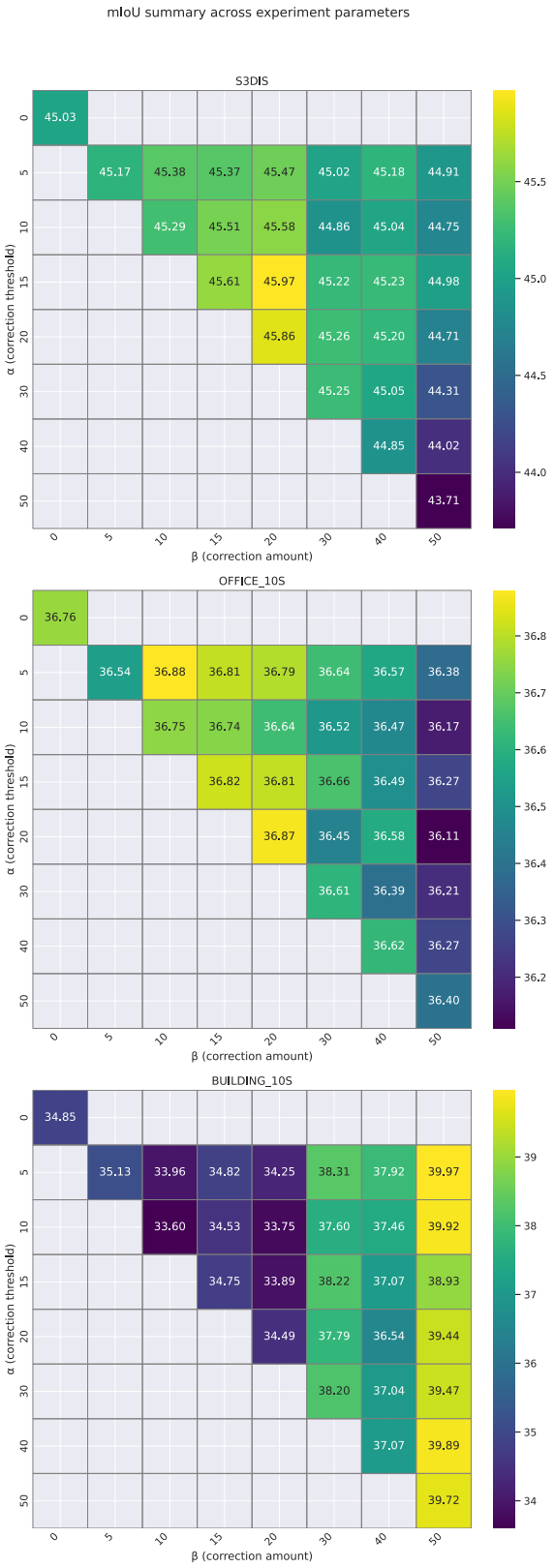

Figure 9. mIoUs for the various experiments on the various datasets, depending on the choice of the Correction threshold α and the Correction amount β. $\alpha = \beta = 0$ represents the baseline, where no scaling is applied.

Finally, on Building_{10s}, the mIoU is either mostly stable or improves significantly. The extent distributions show that this data is most often scaled down by our method. Since this dataset contains very few points on the ground, classifying the ground accurately proves harder for the model with a ground IoU of 0.5. We hypothesise that scaling the point cloud down before performing the voxel-based downsampling partly re-establishes a ground point density closer to that of the training set. This leads to an increase to, for the example of $\alpha = 15$ and $\beta = 30$, 0.78 for the ground IoU. The inference time also goes down using this procedure, from 4 *s* to 3.6 *s* on average and 6.3 *s* to 4.7 *s* for the maximum time.

4.3 Thermal Mapping

People Detection The trained model processes thermal images to detect persons, outputting bounding boxes that indicate their positions. The predictions are forwarded to the thermal layer, where the 2D data is merged with the global 3D point cloud.

Input resolution was set to 640×640 pixels, aligned with the low resolution typical of thermal cameras. On an NVIDIA GeForce RTX 3090 GPU, the model achieved an inference speed of 100 FPS. Training was conducted for each experiment over 50 epochs, with mosaic data augmentation applied at a factor of 0.3 to enhance dataset variability and generalisation.

Due to missing annotations of persons in common poses in the PDWS dataset, it was discarded. For the final trained model, the tristar dataset (Stippel et al., 2023) was combined with parts of the SDT dataset (Pramerdorfer et al., 2020) and the VictimDetectionInDisaster dataset (ThermalObjectDetection, 2024), which reached 97.3% mean Average Precision (mAP). Figure 10 shows examples of predicted persons marked with bounding boxes.

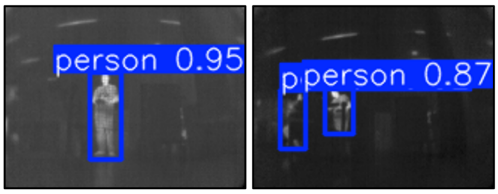

Figure 10. Examples of bounding box predictions from the detection model are shown. The left image depicts a single person standing in the room, while the right image shows two individuals, partially occluded by other objects.

Thermal Map The output of the thermal layer can be seen in Figure 1, where the global point cloud is coloured based on the temperature of each point. The histogram shows that the temperature distribution is in the expected range, with most of the values being close to 18°C.

Furthermore — as described in Section 3.5 — the thermal image can also be overlapped with the global map, allowing for globally localised human detections. An example can be seen in Figure 11, where a person can be seen standing in front of the drone both in the thermal camera and in the generated global point cloud.

5. Conclusions and Future Work

This paper introduced ResQDrone, a quadcopter designed for fully autonomous exploration of indoor environments in CDM scenarios. By integrating a lightweight hemispherical LiDAR with a thermal camera, the system supports a range of algorithms: LiDAR-inertial SLAM for real-time positioning and geometric mapping, frontier-guided exploration, online semantic and thermal mapping, and human detection with global localisation. Real-world tests conducted in both a simple room and a narrow corridor demonstrated the drone's ability to navigate collision-free while delivering multiple map layers to GCS operators in real time.

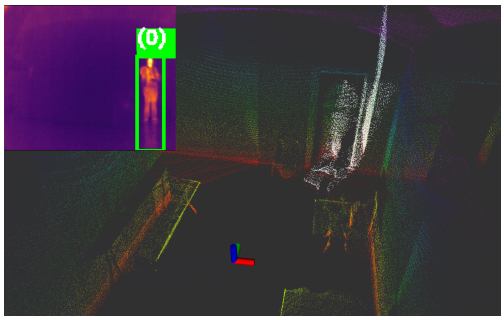

Figure 11. Example of projection of the semantic bounding box on the global point cloud. The top-left image shows what the thermal camera is seeing, highlighting the detected person in the bounding box. The main image shows the global point cloud - colour coded based on the z value - and, in white, the points that overlap the bounding box.

Future work will focus on evaluating the system in larger, more cluttered environments. To improve platform autonomy and safety, both the SLAM and exploration logic will be migrated on-board, reducing reliance on the GCS. Another key objective is to ensure global consistency in the semantic and thermal maps, which currently depend on the low drift of the SLAM pipeline. Finally, for more precise localisation of humans in the global map, bounding box detection will be replaced with a more accurate segmentation-based approach.

6. Acknowledgments

We would like to thank our colleague Vanessa Klugsberger for her help in annotating the point cloud data.

Funded by the European Union

References

Alenzi, Z., Alenzi, E., Alqasir, M., Alruwaili, M., Alhmiedat, T., Alia, O. M., 2022. A semantic classification approach for indoor robot navigation. *Electronics*, 11(13), 2063.

Alqobali, R., Alshmrani, M., Alnasser, R., Rashidi, A., Alhmiedat, T., Alia, O. M., 2023. A Survey on Robot Semantic Navigation Systems for Indoor Environments. *Applied Sciences*, 14(1), 89.

Armeni, I., Sax, A., Zamir, A. R., Savarese, S., 2017. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. *ArXiv e-prints*.

Arsene, C., 2020. Fusion of real time thermal image and 1d/2d/3d depth laser readings for remote thermal sensing in industrial plants by means of uavs and/or robots. *arXiv preprint arXiv:2006.01286*.

Bolz, W., Cella, M., Maikisch, N., Vultaggio, F., d'Apolito, F., Bruckmüller, F., Sulzbachner, C., Fanta-Jende, P., 2024. A robust lidar-based indoor exploration framework for uavs in uncooperative environments. *2024 International Conference on Unmanned Aircraft Systems (ICUAS)*, IEEE, 168–176.

Bournez, P., Salzinger, J., Cella, M., Vultaggio, F., d'Apolito, F., Fanta-Jende, P., 2024. Pushing the limit to near real-time indoor LiDAR-based semantic segmentation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 45–52.

Cruz Ulloa, C., Prieto Sánchez, G., Barrientos, A., Del Cerro, J., 2021. Autonomous thermal vision robotic system for victims recognition in search and rescue missions. *Sensors*, 21(21), 7346.

Dellenbach, P., Deschaud, J.-E., Jacquet, B., Goulette, F., 2022. Ct-icp: Real-time elastic lidar odometry with loop closure. *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 5580–5586.

Elmokadem, T., Savkin, A. V., 2021. Towards fully autonomous UAVs: A survey. *Sensors*, 21(18), 6223.

Foroughi, F., Wang, J., Nemati, A., Chen, Z., Pei, H., 2021. Mapsegnet: A fully automated model based on the encoder-decoder architecture for indoor map segmentation. *IEEE Access*, 9, 101530–101542.

Koide, K., Yokozuka, M., Oishi, S., Banno, A., 2024. Glim: 3d range-inertial localization and mapping with gpu-accelerated scan matching factors. *Robotics and Autonomous Systems*, 179, 104750.

Murphy, R. R., Tadokoro, S., Kleiner, A., 2016. Disaster robotics. *Springer handbook of robotics*.

Pramerdorfer, C., Strohmayer, J., Kampel, M., 2020. Sdt: A synthetic multi-modal dataset for person detection and pose classification. *2020 IEEE International Conference on Image Processing (ICIP)*, IEEE, 1611–1615.

Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H., Elhoseiny, M., Ghanem, B., 2022. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Advances in Neural Information Processing Systems*, 35, 23192–23204.

Qiu, Z., Martínez-Sánchez, J., Arias, P., 2025. Fusion of Thermal Images and Point Clouds for Enhanced Wall Temperature Uniformity Analysis in Building Environments. *Energy and Buildings*, 115781.

Robert, D., Raguet, H., Landrieu, L., 2023. Efficient 3d semantic segmentation with superpoint transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 17195–17204.

Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Rus, D., 2020. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 5135–5142.

Stippel, C., Heitzinger, T., Kampel, M., 2023. A trimodal dataset: Rgb, thermal, and depth for human segmentation and action recognition. *Proceedings of the German Conference on Pattern Recognition (GCPR)*.

ThermalObjectDetection, 2024. Victimsdetectionindisaster dataset. https://universe.roboflow.com/thermalobjectdetection-n8ws5/victimsdetectionindisaster .

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L. J., 2019. Kpconv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE/CVF international conference on computer vision*, 6411–6420.

Tsai, P.-F., Liao, C.-H., Yuan, S.-M., 2022. Using deep learning with thermal imaging for human detection in heavy smoke scenarios. *Sensors*, 22(14), 5351.

Varghese, R., Sambath, M., 2024. Yolov8: A novel object detection algorithm with enhanced performance and robustness. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, IEEE, 1–6.

Vultaggio, F., d'Apolito, F., Sulzbachner, C., Fanta-Jende, P., 2023. Simulation of low-cost mems-lidar and analysis of its effect on the performances of state-of-the-art slams. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 539–545.

Wilson, A. N., Gupta, K. A., Koduru, B. H., Kumar, A., Jha, A., Cenkeramaddi, L. R., 2023. Recent Advances in Thermal Imaging and its Applications Using Machine Learning: A Review. *IEEE Sensors Journal*, 23(4), 3395-3407.

Xu, W., Cai, Y., He, D., Lin, J., Zhang, F., 2022. Fast-lio2: Fast direct lidar-inertial odometry. *IEEE Transactions on Robotics*, 38(4), 2053–2073.

Yao, H., Liang, X., 2024. Autonomous exploration under canopy for forest investigation using lidar and quadrotor. *IEEE Transactions on Geoscience and Remote Sensing*.

Zhang, X., Wang, J., Wang, S., Wang, M., Wang, T., Feng, Z., Zhu, S., Zheng, E., 2025. FAEM: Fast Autonomous Exploration for UAV in Large-Scale Unknown Environments Using LiDAR-Based Mapping. *Drones*, 9(6), 423.

Zhang, Y., Chen, X., Feng, C., Zhou, B., Shen, S., 2024. Falcon: Fast autonomous aerial exploration using coverage path guidance. *IEEE Transactions on Robotics*.

Zhang, Y., Zhou, B., Wang, L., Shen, S., 2022. Exploration with global consistency using real-time re-integration and active loop closure. *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 9682–9688.

Zhou, B., Pan, J., Gao, F., Shen, S., 2021. Raptor: Robust and perception-aware trajectory replanning for quadrotor fast flight. *IEEE Transactions on Robotics*, 37(6).

Zhu, J., Li, H., Zhang, T., 2024. Camera, LiDAR, and IMU Based Multi-Sensor Fusion SLAM: A Survey. *Tsinghua Science and Technology*, 29(2), 415-429.