

nvblox: GPU-Accelerated Incremental Signed Distance Field Mapping

Alexander Millane^{*†}, Helen Oleynikova^{*‡}, Emilie Wirbel[†], Remo Steiner[†],
Vikram Ramasamy[†], David Tingdahl[†], Roland Siegwart[‡]

^{*}Equal contribution

[†]NVIDIA, Switzerland, [‡]Autonomous Systems Lab, ETH Zürich, Switzerland

Abstract—Dense, volumetric maps are essential to enable robot navigation and interaction with the environment. To achieve low latency, dense maps are typically computed on-board the robot, often on computationally constrained hardware. Previous works leave a gap between CPU-based systems for robotic mapping which, due to computation constraints, limit map resolution or scale, and GPU-based reconstruction systems which omit features that are critical to robotic path planning, such as computation of the Euclidean Signed Distance Field (ESDF). We introduce a library, *nvblox*, that aims to fill this gap, by GPU-accelerating robotic volumetric mapping. *Nvblox* delivers a significant performance improvement over the state of the art, achieving up to a $177\times$ speed-up in surface reconstruction, and up to a $31\times$ improvement in distance field computation, and is available open-source¹.

I. INTRODUCTION

To navigate and interact with their environment, robots typically build an internal representation of the world. Significant research in the past decades [1] has focused on building maps that are both useful for robotic path-planning, and efficient to construct. However, fulfilling these two requirements simultaneously remains challenging.

Various successful systems have emerged for solving the Simultaneous Localization and Mapping (SLAM) problem efficiently [2, 3]. Typically these systems build sparse representations of the environment in order to reach real-time rates. Sparse maps have proven effective for localization, however, navigation also requires dense obstacle information.

Several systems have emerged for building denser representations of the environment on the CPU [4–6] that are suitable for robotic path-planning. However, the frequency, resolution, and scale at which these systems can operate is limited by the computational burden of 3D reconstruction on a CPU. To address these limitations, systems utilizing GPU programming have emerged [7, 8]. These systems, however, have typically focused on reconstruction alone and have omitted features needed in a robotic path-planning context, such as incremental computation of the signed distance field and its gradients, as well as an explicit representation of free space. We aim to address this gap.

This paper introduces *nvblox*, a library for volumetric mapping on the GPU, specifically targeted at robotic path-planning. *Nvblox* produces high-resolution surface reconstructions at real-time rates, even on embedded GPUs. In addition *nvblox* also produces distance fields, which are a key output for planning collision-free paths. Central to our approach is the use of parallel computation on the

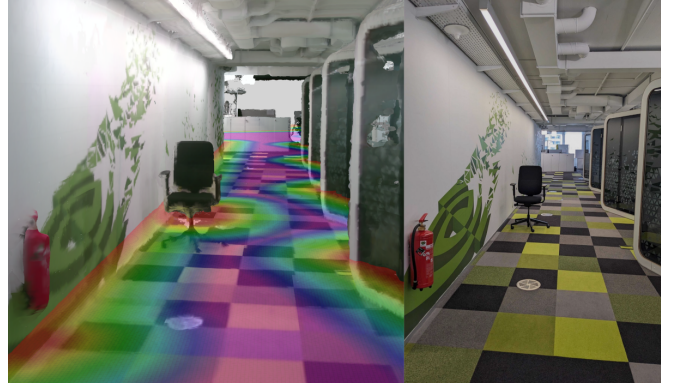


Fig. 1: An example of an *nvblox* reconstruction in an office environment (left) compared to a photo of the mapped scene (right). The reconstruction is built in real-time from a handheld Intel RealSense Depth Camera D455. It shows the mesh and a slice through the distance field. All operations to build this reconstruction are performed in real-time on an embedded GPU.

GPU for all aspects of the pipeline, including queries. We show the efficacy and efficiency of *nvblox* on several public datasets, and applied to several robotic use-cases, such as path planning for robot arms, flying robots, ground robots, and for mapping of dynamic obstacles such as people.

In summary, the contribution of this paper is a GPU-accelerated Signed Distance Field (SDF) library with a convenient and flexible interface. *Nvblox* fuses in data from RGB-D sensors and/or LiDAR, achieving up to $177\times$ faster surface reconstruction (Truncated Signed Distance Field (TSDF)) and $31\times$ faster distance field (ESDF) computation than state-of-the-art CPU-based methods [4, 9]. The library is made available open-source¹ in both ROS1² and ROS2³.

II. RELATED WORK

Mapping is a well-explored problem in robotics [1]. We can categorize robotic mapping approaches into two broad categories: sparse and dense. Sparse methods focus on creating a map representation for pose estimation and localization while dense methods focus on reconstructing the geometry of the environment.

Systems for dense mapping can be organized by the underlying representation of the environment. LSD-SLAM [10] and DVO-SLAM [11] build a map consisting of keyframe depth-maps. Kintinuous [12] and ElasticFusion [8] build reconstructions in the form of a deformable mesh and

¹github.com/nvidia-isaac/nvblox

²github.com/ethz-asl/nvblox_ros1

³github.com/NVIDIA-ISAAC-ROS/isaac_ros_nvblox

collection of surfels, respectively. More recently, Kimera [13] builds a semantic mesh of the environment. These approaches build visually compelling reconstructions, but are difficult to use for robotic path-planning, as they lack information about observed free space and focus only on surfaces.

Recently, reconstruction systems based on neural radiance fields (NERFs) have gained significant attention [14]. The original offline approach has seen dramatic speedups [15], leading to implementations that generate reconstructions in real-time [16, 17]. However, these approaches heavily keyframe the input image data and are therefore unlikely to be reactive enough for robot path-planning in the control loop.

Voxel-based methods build reconstructions that are well-suited to robot path-planning tasks. Voxels capture the reconstructed quantity (for example occupancy probability) over the *volume* of 3D space, and can therefore represent free-space, not only surfaces. The most common approach to volumetric reconstruction is occupancy grid mapping [18] and its efficient implementation in 3D, Octomap [5]. These approaches are widely used in robotic mapping and are the default in common robotics toolkits [19, 20].

Another popular approach to volumetric mapping utilizes a voxelized TSDF. This approach was popularized by Kinect-Fusion [7] which generates surface reconstructions using a consumer-grade depth camera. The original, fixed-grid-based approach was extended to use spatially hashed voxels by Niessner et. al. [21]. Voxelblox [4] follows the approach of voxel-hashing but adds ESDF computation, a feature of particular importance for robotic path-planning. Voxelblox has been used in many follow-up works which have used it for planning [22, 23], as well as extended its mapping capabilities, for example for global mapping [6] and semantic mapping [24]. Despite its success, voxelblox is limited in the resolution of maps it builds due to the computational complexity of updating a high-resolution voxel grid.

Several works have focused on decreasing voxelblox’s ESDF generation error and runtime. Voxfield [25], for example, removes the inaccuracies caused by voxelblox’s quasi-Euclidean distance estimation and improves the ESDF runtime by up to 42%. Similarly, FIESTA speeds up voxelblox’s ESDF computation by $4\times$ while also computing full Euclidean distance. *Nvblox* also uses full Euclidean distance, therefore reducing the ESDF error, but is on average $31\times$ faster than voxelblox.

By improving existing methods through GPU acceleration, we create a library that provides a suitable representation for a large body of path planners and other downstream applications, while reducing the runtime and allowing the creation of higher-resolution maps on the robot.

III. PROBLEM STATEMENT

Given a sequence of measurements from an RGB-D camera and/or a LiDAR, we aim to build a volumetric reconstruction of the scene. In particular, we compute the surface reconstruction (expressed as either occupancy or the TSDF) and the distance field (expressed as the ESDF). Our

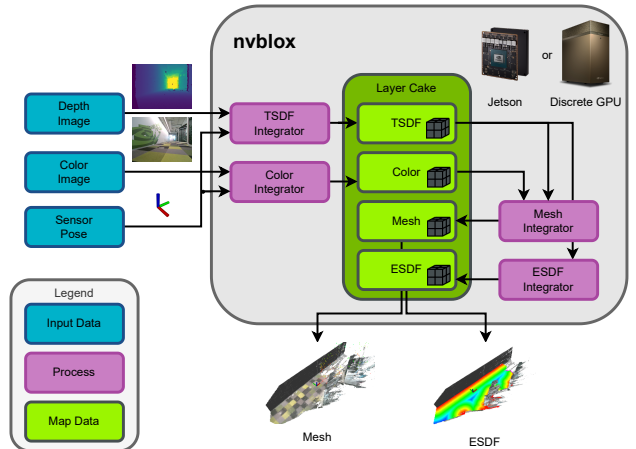


Fig. 2: The system architecture of *nvblox*, configured for TSDF mapping from an RGB-D camera sensor. The reconstructed map (called a *LayerCake*) is composed of co-located and aligned 3D voxel grids. Input depth-maps and color images are integrated into the TSDF and Color voxel layers, from which voxel grids containing the ESDF and a mesh reconstruction are derived. See Sec. IV for details.

reconstructions are functions $\Phi : \mathbb{R}^3 \rightarrow \phi$, which maps a point in 3D space to some mapped quantity ϕ , for example, distance, occupancy, or color. These functions are voxelized, i.e. represented as a sparse set of samples on a regular 3D grid, where samples are allocated in regions of space that have been observed by the sensor. We assume that at time step i the sensor frame C^i is localized in a frame L such that we have access to the sensor pose $\mathbf{T}_{LC^i} \in \text{SE}(3)$. Observations take the form of depth maps and color images. A depth map is $\mathcal{D} : \Omega \rightarrow d$, where $d \in \mathbb{R}$ is a depth value in meters. The domain Ω is the image plane in the case of depth cameras, and the beam angles in the case of rotating LiDARs. Similarly, color images are $\mathcal{C} : \Omega \rightarrow c$ where $c \in \mathbb{R}^3$. In the remainder of this paper, we will refer to observations from both camera and LiDAR as images and treat the two equally.

IV. SYSTEM ARCHITECTURE DESCRIPTION

The architecture of *nvblox* is shown in Fig. 2. The system consists of multiple components: the reconstructed map, which contains several *Layers*, *Integrators* that add sensor data to the map and components that transform one layer type into another, such as mesh and ESDF integrators. We discuss these components in more detail below.

A. The Map

Our reconstruction is represented as several overlapping 3D voxel grids, called *Layers*, following [4]. Each *Layer* of the map stores a different type of (user-defined) data for overlapping aligned volumes of 3D space. The map is sparse, such that voxels are only allocated in regions of 3D space that are observed during mapping. This sparsity is achieved using a two-level hierarchy, following [21]. The first level is a hash table that maps 3D grid indices to *VoxelBlocks*. In *nvblox* this hash table can be queried in GPU kernels using an interface based on `stdgpu` [26]. In the second level, each *VoxelBlock* contains a $8 \times 8 \times 8$ group of densely allocated voxels which are stored contiguously in GPU memory, leading to coalesced loads in GPU kernels.

The map is designed to be extended with new layers. To create a new *Layer*, a user needs to specify the contents of a single voxel. The library generates the definitions for the corresponding block-hashed voxel grid at compile time, as well as the CPU and GPU interfaces. The *nvblox* library includes commonly used *Layers*: TSDF, ESDF, occupancy, color, and meshes.

B. Frame Integration

Incoming sensor data is added to the reconstruction stored in one of the map layers. This occurs in several steps. We first ray trace through the *VoxelBlock*-grid on the GPU to determine which *VoxelBlocks* are in view using [27] and allocate those not yet in the map. We then project each voxel in view into the depth image:

$$d = \mathcal{D}[\pi_{\text{sensor}}(\mathbf{T}_{CL}\mathbf{p}_L)] \quad (1)$$

where $d \in \mathbb{R}$ is the sampled depth value, \mathbf{T}_{CL} is the camera pose with respect to the *Layer*, and \mathbf{p}_L is the voxel center position in the *Layer* coordinate frame. The sensor projection function for a camera π_{sensor} is defined as

$$\pi_{\text{camera}}(\mathbf{p}_C) = \frac{1}{p_{C,z}} \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \end{bmatrix} \begin{bmatrix} p_{C,x} \\ p_{C,y} \\ 1 \end{bmatrix} \quad (2)$$

where f_u, f_v, c_u, c_v are calibrated pinhole camera intrinsics, and for LiDAR

$$\pi_{\text{lidar}}(\mathbf{p}_C) = \begin{bmatrix} (\tan^{-1}(p_{C,y}/p_{C,x}) - \theta_{\text{start}}) * \alpha_\theta \\ (\cos^{-1}(p_{C,z}/r) - \phi_{\text{start}}) * \beta_\phi \end{bmatrix}, \quad (3)$$

where $\theta_{\text{start}}, \phi_{\text{start}}$ are the minimum polar and azimuth angles, and α_θ and β_ϕ are measured in pixels-per-radian and are calculated by dividing the Field of View (FoV) by the number of steps/beams in the relevant dimension. The function $\mathcal{D}[\mathbf{u}]$ indicates sampling the depth image \mathcal{D} at image-coordinates \mathbf{u} . For a camera, we sample using nearest-neighbor, and for LiDAR-based depth images, which can be very sparse, we use linear interpolation with modifications to avoid interpolating over foreground-background boundaries.

To update voxels, we call a per-voxel update functor on the GPU in parallel over all voxels in view. We will briefly cover two main update functors: one for TSDF maps and one for occupancy maps. TSDF maps store both a truncated, projective distance (d_{tsdf}) and a weight (w) per voxel, while occupancy maps store a single log-odds occupancy value l_o .

The TSDF update functor uses the voxel depth d_v , which is the depth from the voxel center to the sensor, to calculate $d_p = d - d_v$, the projective distance at that voxel, truncates it to within the positive and negative truncation distance ϵ , and computes a weight $w = f_w(d_p)$, where $f_w(x)$ is the weighting function, either a constant or based on the sensor's error model (*nvblox* offers several models). TSDF distance values are combined using a weighted average, and weights are simply added, up to a maximum.

The occupancy update functor, updates a voxel's occupancy probability in log-odds space. A voxel is updated with a constant negative value if $d_p > 0$ (indicating a lower

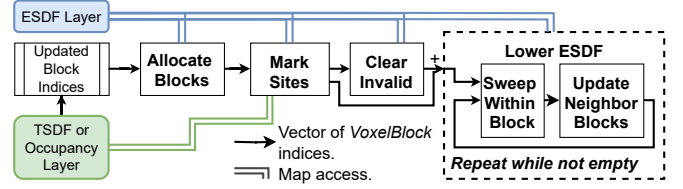


Fig. 3: Overall flow of the incremental ESDF computation on the GPU. The algorithm does as much of the update in parallel as possible, iteratively identifying blocks that need to be updated, and then running kernels on all blocks in parallel.

probability of occupancy) or a constant positive value if $d_p < 0$. Bayesian fusion in log-odds space is implemented as running addition of the update values (see [18] for details).

For TSDF maps, updated blocks are periodically meshed using a parallelized Marching Cubes [28] algorithm.

C. ESDF Computation

The Truncated Signed Distance Field (TSDF) contains projective distances up to a small truncation distance. For path planning applications we require *Euclidean* distances and for greater distances than the truncation band. For a discussion on why a TSDF is insufficient for this, we refer the reader to [29].

Our Euclidean Signed Distance Field (ESDF) computation approach has several requirements. It must be both parallelizable on the GPU and incremental, allowing us to update only parts of the map that have changed to reduce computation time. Finally, we avoid simplifying assumptions (like quasi-Euclidean distance used in *voxbox* [4]) to maintain high accuracy of the resulting distance field.

In order to create a highly parallelizable ESDF computation algorithm, we base our work in spirit on the Parallel Banding Algorithm (PBA) proposed by Cao *et al.* [30]. The overall flow of the algorithm is shown in Fig. 3 and visualized step-by-step in Fig. 5. The general intuition is to keep a list of blocks that need updating, and update all voxels in all affected blocks in parallel, transmit information between blocks, and repeat until convergence.

ESDF voxels come in two categories: *sites* and regular voxels. Sites are surface boundary voxels, as shown in Fig. 4. Site voxels can be *parents* to *child* voxels, and each *child* stores the location its *parent*. Regular voxels can have three states: unknown, free (positive distance), and occupied (negative distance). Each observed voxel also stores its distance to the nearest site.

First, in *Allocate Blocks*, a TSDF layer and a list of updated *VoxelBlocks* are taken as input. Generally, TSDF blocks that were changed since the last iteration of ESDF integration are considered updated (this allows us to run the ESDF update slower than sensor rate). Any new *VoxelBlocks* are allocated in the ESDF as needed.

The next step is *Mark Sites*. We consider voxels to be sites if their TSDF distance is within a threshold ϵ of the zero crossing (see Fig. 5c); otherwise, an occupancy voxel is a site if it's adjacent to free voxels.

In addition to marking sites, this function ensures consistency between the ESDF and TSDF or occupancy maps.

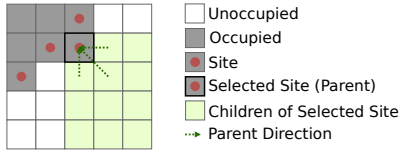


Fig. 4: “Sites” are voxels on the surface boundary, and its children are any voxels which are closer to that voxel than any other site. Each child stores its parent direction - the direction that points to the parent site location (only shown for the closest 3 voxels here for clarity).

There are two general cases we need to handle: (1) newly occupied (free \rightarrow occupied) or newly observed (unknown \rightarrow either) voxels, and (2) newly freed voxels (occupied \rightarrow free). Newly occupied/observed voxels simply take on their TSDF distance values and are added to *Indices to Update*. Newly free voxels have their distances set invalid. To complicate matters, if the cleared voxel was a site, we need to ensure that all of its *children* (voxels whose closest site was this one) are invalidated, adding their blocks to *Indices to Clear*.

Next, we *Clear Invalid*. The idea is to find voxels whose parents are no longer a site, and therefore need distance recomputation (see Fig. 5d). We select a subset of the map that is within the maximum ESDF distance of the blocks in *Indices to Clear*, and then check that each voxel in these blocks still has a valid parent. If the parent is no longer marked as a site, then the voxel distance is set to maximum, and the block index is added to *Cleared Indices*. Checking all blocks in range would seem an expensive operation but in practice, it is very efficient to check them all in parallel.

Finally, the *Lower ESDF* stage (so called since it exclusively *lowers* the ESDF voxel distance) consists of two steps in a loop. The first is *Sweeping Within a Block*. This sweeps once in each axis direction in each *VoxelBlock* in parallel. This is similar to the PBA [30] approach, except confined to the *VoxelBlock* boundaries. Fig. 5e shows the first positive sweep within the block: each neighbor in the $X+$ direction is updated if there is a shorter distance to the site through that direction. We then repeat the process in $X-$, $Y+$, $Y-$, $Z+$ and $Z-$. This is done over all affected *VoxelBlocks* in a single kernel call, and at the end of the 6 sweeps, the distances within a *VoxelBlock* are correct, given the current values on each *VoxelBlock*’s boundaries.

We now reconcile the differences between *VoxelBlocks* by *Updating Neighbor Blocks* by communicating across block boundaries. Values are propagated from the edges of one block to another if there is a shorter distance to a site through the neighboring block (see Fig. 5i). If the last stage communicated over *VoxelBlock* borders, those blocks require another sweep, to propagate the communicated distance *within* the affected block. We repeat the sweep-neighbor update loop until no more blocks can be updated.

V. EXPERIMENTS

In this section, we aim to validate the central claim of our paper: that *nvblox* improves the state-of-the-art in volumetric mapping for robot path planning in terms of run-time performance, without compromising the accuracy of the reconstructed distance field (ESDF). We report timings on

3 different platforms: a desktop computer with an Intel i9 CPU and NVIDIA RTX3090 Ti GPU (Desktop), a laptop computer with an Intel i7 CPU and a RTX3000 Mobile GPU (Laptop), and a Jetson Xavier AGX (Jetson).

A. Whole System Timings

We evaluate the performance of various modules of *nvblox* on the Replica Dataset [31], which provides photorealistic renderings of synthetic rooms, and the Redwood dataset [32] which are real scans of several environments using a consumer depth camera. For Replica, we use the sequences generated in [16].

Table I shows timings for our system’s modules; TSDF fusion, color fusion, incremental ESDF update, and incremental meshing. We perform ESDF generation and meshing every 4 frames. Timings are averaged across 8 Replica Dataset sequences and 5 Redwood Dataset sequences.

When compared to our previous work, *voxblox* [4], which runs on the CPU, we see significant speed-ups in all modules of the system. In the case of TSDF and color this is a well-known result, as the GPU has been used to accelerate TSDF mapping since KinectFusion [7]. We show two additional findings. These speed-ups are also achievable on an embedded GPU. Furthermore, similar speed-ups are available for incremental ESDF calculation, which we describe below.

B. ESDF Timings

We aim to validate our claim of improving the state-of-the-art in incremental ESDF calculation. We compare *nvblox* against *voxblox* [4], and Fiesta [9], a recent and more performant algorithm. Table II shows timings and ESDF accuracy for our *Desktop* system. The ESDF error is calculated as the median absolute voxel-wise error between the reconstructed ESDF and a voxelized ESDF ground-truth. The ground-truth is generated by computing the distance between the reconstructed voxel centers, and the dataset-supplied ground-truth surface. Table II shows a significant speedup of $31\times$ with respect to *voxblox* and $7\times$ with respect to Fiesta. Furthermore, the experiments show that this speed-up does not come at the cost of reduced accuracy.

C. Resolution Scaling

The relationship between map resolution and TSDF and ESDF computation time is critical because map resolution is often adjusted to meet performance limitations. We run *nvblox* and *voxblox* [4] on the *office0* Replica sequence for various resolutions. Fig. 6 shows the results of this experiment. Even at high resolution (1 cm for TSDF and 2 cm for ESDF) *nvblox* performs computations faster than *voxblox* running at the lowest tested resolution 10 cm. This speed-up is likely to enable robotic applications requiring higher precision 3D perception.

D. Query Timings

The primary purpose of mapping in a typical robotic system is to provide collision information to path-planning modules. For many optimization or sampling-based planners,

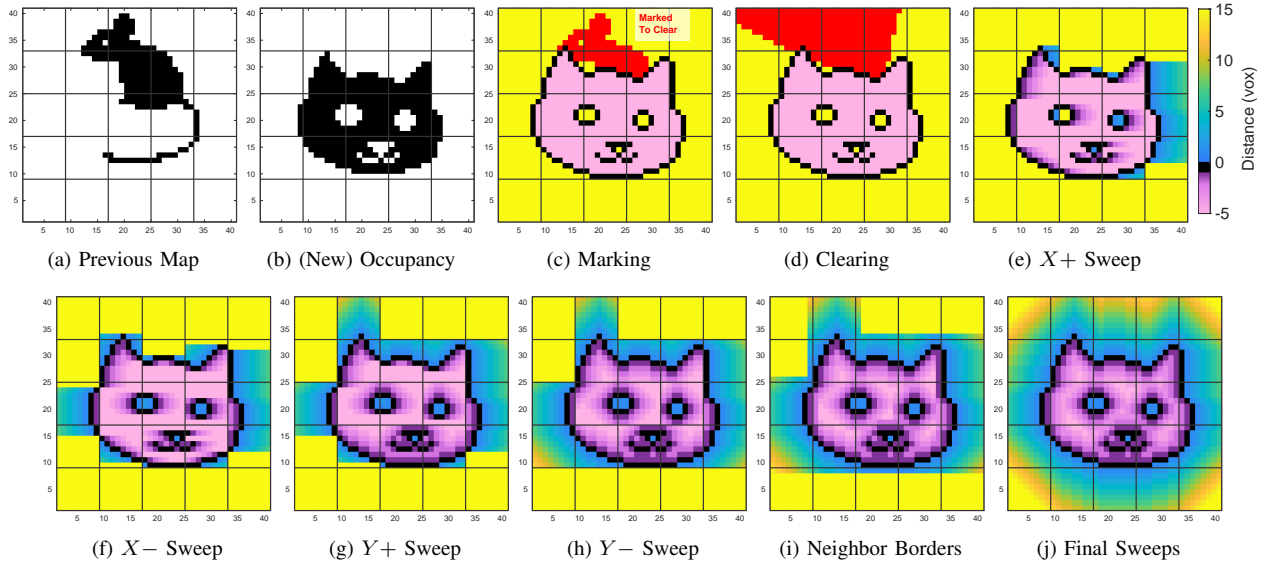


Fig. 5: Marking and lowering, shown step-by-step on an image, for simplicity on 2D occupancy (rather than 3D TSDF). Here we start with a previous map shown in a to demonstrate clearing of previously-occupied space. The bold lines represent *VoxelBlock* boundaries. The general idea of the algorithm is to iteratively compute correct distance values *within* a *VoxelBlock*, and then communicate distance information across *VoxelBlock* boundaries to neighbors.

Component	Replica [31]								Redwood [32]							
	Desktop				Laptop				Desktop				Laptop			
	nvblox	vox.	nvblox	vox.	nvblox	vox.	Speedup		nvblox	vox.	nvblox	vox.	nvblox	vox.	nvblox	Speedup
ESDF	1.9	163.2	3.6	291.5	8.4	231.6	×63		1.5	29.1	2.6	46.5	4.2	38.7		×16
TSDF	0.4	-	0.6	-	1.6	-	×174		0.2	-	0.2	-	0.5	-		×177
Color	1.7	-	2.5	-	4.2	-	-		1.1	-	1.6	-	2.4	-		-
TSDF+Color	2.1	86.7	3.2	106.6	5.8	226.7	×38		1.3	38.4	1.8	33.6	2.9	76.7		×25
Mesh	1.6	6.2	4.0	12.0	12.3	15.4	×3		0.6	12.7	1.5	15.8	2.7	23.0		×13

TABLE I: Timings for various components of *nvblox* and *voxblox* during reconstruction of the Replica [31] and Redwood [32] datasets at 5 cm resolution. Timings are averaged over 8 sequences for Replica, and 5 sequences for Redwood (see Sec. V-A for details). *Speedup* is how many times faster *nvblox* is than *voxblox*. Some values are missing as *voxblox* does not separate TSDF and color integration; because of this, the TSDF speedup is assuming *only* surface integration in *nvblox* vs. both surface and color for *voxblox* (which is relevant in colorless scenarios like LiDAR integration).

Dataset	Sequence	Median ESDF Error (<i>m</i>)			ESDF Runtime (<i>ms</i>) (speedup)		
		nvblox	voxblox	Fiesta	nvblox	voxblox	Fiesta
Redwood [3]	apartment	0.04	0.06	0.05	1.7 (×14)	25 (×1)	5.5 (×5)
Redwood [3]	bedroom	0.02	0.05	0.03	1.4 (×15)	22 (×1)	3.4 (×6)
Redwood [3]	boardroom	0.06	0.08	0.06	1.7 (×17)	30 (×1)	4.0 (×8)
Redwood [3]	lobby	0.10	0.10	0.08	2.1 (×16)	34 (×1)	5.2 (×7)
Redwood [3]	loft	0.04	0.08	0.04	1.8 (×26)	48 (×1)	8.4 (×6)
Cow and lady [4]	-	0.09	0.06	0.07	2.8 (×68)	190 (×1)	52 (×4)
Average		0.06	0.07	0.06	1.9 (×31)	58 (×1)	13 (×4)

TABLE II: Accuracy and runtime performance of incremental ESDF generation of the proposed system compared to baselines: *voxblox* [4], and Fiesta [9]. The systems are compared on the Redwood [32] and Cow and Lady [4] datasets (see Sec. V-A for details). *Speedup* is the runtime performance increase over *voxblox* for both methods. On average, *nvblox* is 7× faster than Fiesta and 31× faster than *voxblox*.

querying for collisions can constitute a significant portion of the total computational cost. Because these queries are often required in batch, performing these queries on the GPU allows us to take advantage of parallelization, and enables GPU-based path planners, like in [33] and [34]. A query in *nvblox* takes a collection of 3D points and returns their distances to the closest surface and optionally the distance

field gradient, by performing an ESDF lookup on the GPU. Table III shows query rates in giga-queries-per-second for an NVIDIA GeForce 3090 Ti as well as a Jetson AGX. The table shows the results for spatially correlated (cor.) and uncorrelated (uncor.) sampling. In correlated sampling, adjacent queries are more likely to fall in the same *VoxelBlock*, leading to coalesced memory access on the GPU and higher query

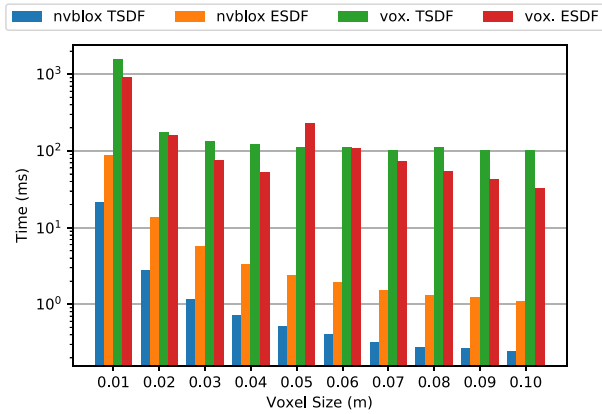


Fig. 6: TSDF and ESDF computation times with for various voxel sizes for voxblox [4] and *nvblox*. Timings are generated using Replica sequences [31] on the *Laptop* compute platform.

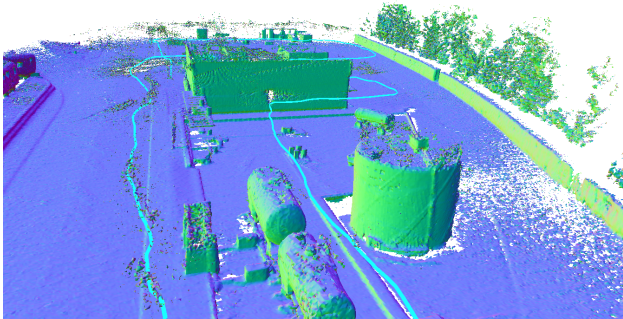


Fig. 7: An example of a reconstructed mesh from a large-scale flying robot dataset. The integration time is less than 7 ms for a 64-beam LiDAR with 25 meter maximum range at 10 cm resolution on a laptop, and less than 20 milliseconds on the Jetson, allowing us to run at at least $2\times$ real time.

rates. This is typically the case for robotic use cases where the queries are spatially correlated, for example, clustered around the robot’s current location.

E. Application examples

To demonstrate the wide utility of *nvblox* at various problem scales we show examples of its use on flying robots, robot arms, and mobile ground robots. Fig. 7 shows the results from a dataset collected with a drone [6] equipped with a 64-beam Ouster OS1 LiDAR. We use Fast-LIO [36] as a pose estimator and integrate the LiDAR up to a range of 25 meters at a resolution of 5 cm, which requires less than 7 ms per LiDAR scan on a Laptop. This is an example of *nvblox*’s suitability for large-scale outdoor scenarios, where the resulting 3D ESDF can be used for both global [37] and local planning [38]. The *nvblox* library has also been used on-board a different flying robot to enable Riemannian Motion Policies which allow reactive navigation at kHz rates [34].

Nvblox is suitable for small-scale problems as well, as shown for high-rate adaptive planning for robot arms in CuRobo [33], where the authors take advantage of *nvblox*’s fast query speeds directly on GPU to sample more trajectory candidates than previously possible. In general, *nvblox* is useful not only because it is faster than existing methods, but also because all data is already stored on the GPU. This

Dataset	10 ⁹ Queries per Second					
	Desktop		Laptop		Jetson	
	cor.	uncor.	cor.	uncor.	cor.	uncor.
Redwood	6.2	3.3	1.7	1.3	0.8	0.5
Sun3D	7.3	3.3	1.8	1.1	0.7	0.3

TABLE III: The number of distance giga-queries per second delivered by *nvblox*, averaged over several sequences of the Redwood [32] and Sun3D datasets [35]. Even on the Jetson, uncorrelated queries take only 3 nanoseconds per point.

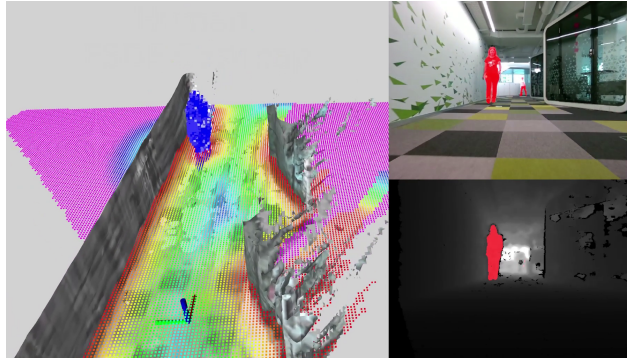


Fig. 8: An example of *nvblox* on a ground robot in an office building. The input color and depth images, as well as the semantic segmentation are shown on the right. On the left is the resulting reconstruction showing the reconstructed mesh, a slice through the distance field, as well as occupancy-probability voxels representing the reconstructed human (in blue).

enables integration with other GPU-accelerated methods, without requiring CPU-GPU memory transfers.

Lastly, we show an image from a robot in an office environment Fig. 8. In this example, we use PeopleSemSegnet⁵ to segment the reconstruction into the static environment and dynamic elements (e.g. humans). Depth image segments belonging to the human class are fed into a 3D occupancy grid using a separate *OccupancyLayer* in *nvblox*. The result is a two-part reconstruction, where dynamic parts of the scene decay over time, but the static parts of the scene are accurately reconstructed using TSDF fusion.

VI. CONCLUSION

In conclusion, we introduce *nvblox*, a library for volumetric mapping on the GPU. The library fills a gap between CPU-based volumetric mapping systems for robots, which are computationally limited, and GPU-based systems that typically omit features that are important for robotics use-cases. As part of the toolbox we include a novel incremental, GPU-accelerated method for computing the ESDF. The system is optimized for operation on both discrete and embedded GPUs. We provide experiments demonstrating that *nvblox* is significantly faster both in mapping and distance field computation, as well as at query time, than other state-of-the-art approaches.

⁵catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/peoplesemsegnet

REFERENCES

- [1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age,” *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “Orb-slam: A versatile and accurate monocular slam system,” *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [3] T. Schneider, M. Dymczyk, M. Fehr, K. Egger, S. Lynen, I. Gilitschenski, and R. Siegwart, “Maplab: An open framework for research in visual-inertial mapping and localization,” *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1418–1425, 2018.
- [4] H. Oleynikova, Z. Taylor, M. Fehr, R. Siegwart, and J. Nieto, “Voxblox: Incremental 3d euclidean signed distance fields for on-board mav planning,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [5] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, “Octomap: An efficient probabilistic 3d mapping framework based on octrees,” *Autonomous robots*, vol. 34, pp. 189–206, 2013.
- [6] V. Reijgwart, A. Millane, H. Oleynikova, R. Siegwart, C. Cadena, and J. Nieto, “Voxgraph: Globally consistent, volumetric mapping using signed distance function submaps,” *IEEE Robotics and Automation Letters*, vol. 5, no. 1, pp. 227–234, 2019.
- [7] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, *et al.*, “Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera,” in *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011, pp. 559–568.
- [8] T. Whelan, S. Leutenegger, R. Salas-Moreno, B. Glocker, and A. Davison, “Elasticfusion: Dense slam without a pose graph,” *Robotics: Science and Systems*, 2015.
- [9] L. Han, F. Gao, B. Zhou, and S. Shen, “Fiesta: Fast incremental euclidean distance fields for online motion planning of aerial robots,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2019, pp. 4423–4430.
- [10] J. Engel, T. Schöps, and D. Cremers, “Lsd-slam: Large-scale direct monocular slam,” in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part II 13*, Springer, 2014, pp. 834–849.
- [11] C. Kerl, J. Sturm, and D. Cremers, “Dense visual slam for rgb-d cameras,” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2013, pp. 2100–2106.
- [12] T. Whelan, M. Kaess, M. Fallon, H. Johannsson, J. Leonard, and J. McDonald, “Kintinuous: Spatially extended kinectfusion,” 2012.
- [13] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, “Kimera: An open-source library for real-time metric-semantic localization and mapping,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 1689–1696.
- [14] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [15] T. Müller, A. Evans, C. Schied, and A. Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM Transactions on Graphics (ToG)*, vol. 41, no. 4, pp. 1–15, 2022.
- [16] E. Sucar, S. Liu, J. Ortiz, and A. J. Davison, “Imap: Implicit mapping and positioning in real-time,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 6229–6238.
- [17] Z. Zhu, S. Peng, V. Larsson, W. Xu, H. Bao, Z. Cui, M. R. Oswald, and M. Pollefeys, “Nice-slam: Neural implicit scalable encoding for slam,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 786–12 796.
- [18] S. Thrun, “Probabilistic robotics,” *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.
- [19] S. Macenski, F. Martín, R. White, and J. G. Clavero, “The marathon 2: A navigation system,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020, pp. 2718–2725.
- [20] S. Macenski, D. Tsai, and M. Feinberg, “Spatio-temporal voxel layer: A view on robot perception for the dynamic world,” *International Journal of Advanced Robotic Systems*, vol. 17, no. 2, p. 1729881420910530, 2020.
- [21] M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger, “Real-time 3d reconstruction at scale using voxel hashing,” *ACM Transactions on Graphics (ToG)*, vol. 32, no. 6, pp. 1–11, 2013.
- [22] M. Tranzatto, T. Miki, M. Dharmadhikari, L. Bernreiter, M. Kulkarni, F. Mascarich, O. Andersson, S. Khattak, M. Hutter, R. Siegwart, *et al.*, “Cerberus in the darpa subterranean challenge,” *Science Robotics*, vol. 7, no. 66, eabp9742, 2022.
- [23] T. Dang, M. Tranzatto, S. Khattak, F. Mascarich, K. Alexis, and M. Hutter, “Graph-based subterranean exploration path planning using aerial and legged robots,” *Journal of Field Robotics*, vol. 37, no. 8, pp. 1363–1388, 2020, Wiley Online Library.
- [24] M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto, “Volumetric instance-aware semantic mapping and 3d object discovery,” *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 3037–3044, 2019.

- [25] Y. Pan, Y. Kompis, L. Bartolomei, R. Mascaro, C. Stachniss, and M. Chli, "Voxfield: Non-projective signed distance fields for online planning and 3d reconstruction," in *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2022.
- [26] P. Stotko, "Stdgpu: Efficient stl-like data structures on the gpu," *arXiv preprint arXiv:1908.05936*, 2019.
- [27] J. Amanatides, A. Woo, *et al.*, "A fast voxel traversal algorithm for ray tracing," in *Eurographics*, vol. 87, 1987, pp. 3–10.
- [28] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," in *Seminal graphics: pioneering efforts that shaped the field*, 1998, pp. 347–353.
- [29] H. Oleynikova, A. Millane, Z. Taylor, E. Galceran, J. Nieto, and R. Siegwart, "Signed distance fields: A natural representation for both mapping and planning," in *Workshop on Geometry and Beyond, RSS 2016*, 2016.
- [30] T.-T. Cao, K. Tang, A. Mohamed, and T.-S. Tan, "Parallel banding algorithm to compute exact distance transform with the gpu," in *Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games*, 2010, pp. 83–90.
- [31] J. Straub, T. Whelan, L. Ma, Y. Chen, E. Wijmans, S. Green, J. J. Engel, R. Mur-Artal, C. Ren, S. Verma, A. Clarkson, M. Yan, B. Budge, Y. Yan, X. Pan, J. Yon, Y. Zou, K. Leon, N. Carter, J. Briales, T. Gillingham, E. Mueggler, L. Pesqueira, M. Savva, D. Batra, H. M. Strasdat, R. D. Nardi, M. Goesele, S. Lovegrove, and R. Newcombe, "The Replica dataset: A digital replica of indoor spaces," *arXiv preprint arXiv:1906.05797*, 2019.
- [32] J. Park, Q.-Y. Zhou, and V. Koltun, "Colored point cloud registration revisited," in *ICCV*, 2017.
- [33] B. Sundaralingam, S. K. S. Hari, A. Fishman, C. Garrett, K. Van Wyk, V. Blukis, A. Millane, H. Oleynikova, A. Handa, F. Ramos, *et al.*, "Curobo: Parallelized collision-free robot motion generation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2023, pp. 8112–8119.
- [34] M. Pantic, I. Meijer, R. Bähmann, N. Alatur, O. Andersson, C. Cadena, R. Siegwart, and L. Ott, "Obstacle avoidance using raycasting and riemannian motion policies at khz rates for mavs," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2023, pp. 1666–1672.
- [35] J. Xiao, A. Owens, and A. Torralba, "Sun3d: A database of big spaces reconstructed using sfm and object labels," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1625–1632.
- [36] W. Xu and F. Zhang, "Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3317–3324, 2021.
- [37] H. Oleynikova, Z. Taylor, R. Siegwart, and J. Nieto, "Sparse 3d topological graphs for micro-aerial vehicle planning," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 1–9.
- [38] H. Oleynikova, C. Lanegger, Z. Taylor, M. Pantic, A. Millane, R. Siegwart, and J. Nieto, "An open-source system for vision-based micro-aerial vehicle mapping, planning, and flight in cluttered environments," *Journal of Field Robotics*, vol. 37, no. 4, pp. 642–666, 2020.